When Math is Not a Universal Language:
Improving the Use of Evidence-Based Vocabulary Instruction in Inclusive Middle School Mathematics Classrooms

---

A Dissertation

Presented to

The Faculty of the Curry School of Education and Human Development

University of Virginia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Katherine Nolen Peeples

B.A., University of North Carolina at Greensboro

M.Ed., Elon University

May 2019

Abstract

Middle school students with disabilities often struggle with reading and literacy skills including vocabulary, and require support in content-area classes where a student's weak vocabulary knowledge can negatively affect his/her learning, achievement, and deep understanding. Mathematics at the middle school level is one such content area in which students who struggle with literacy and language (including but not limited to students with disabilities) tend to perform poorly on assessments of mathematical knowledge and procedures. In general, middle school mathematics teachers are not provided with training in literacy instruction, much less in explicit vocabulary instruction. Additionally, there are relatively few studies in the current literature that examine content-area vocabulary instruction at the middle school or secondary level.

This study aimed to address these gaps in the research by examining the effect of the performance feedback and coaching component of the Content Acquisition Podcasts for Professional Development (CAP-PD) system on middle school mathematics teachers' use of explicit mathematics vocabulary instruction. Results of the study indicate that performance feedback and coaching had moderately positive effects on teachers' use and quality of explicit mathematics vocabulary instruction. Explicit vocabulary instruction in mathematics has potential to support and improve the mathematical conceptual and procedural understanding of students with disabilities.

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

APPROVAL OF THE DISSERTATION

This dissertation, "When Math is Not a Universal Language: Improving the Use of Evidence-Based Vocabulary Instruction in Inclusive Middle School Mathematics Classrooms," 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.

____

Chair - Michael J. Kennedy, Ph.D.

____

Member – Robert Q. Berry, Ph.D.

____

Member – John W. Lloyd, Ph.D.

____

Member – Lysandra Cook, Ph.D.

____

Date

Dedication

For my father, Ralph, and my mother, Faith. Thank you for your infinite and unconditional love and patience while I figure this whole life thing out.

For my grandparents, Ralph Sr. and Rita, whose presence and pride I feel always.

For my brothers, Sam and Michael, and my sister, Emma.

***

Audaces fortuna iuvat.
-Publius Virgilius Maro
The *Aeneid*, Book X

Acknowledgements

Doing a doctoral dissertation feels, at times, like the most isolating thing in the world. But the truth is that it's not possible to do it alone, and in my case, there are a number of people I wish to thank specifically for their guidance along the way.

My advisor and dissertation committee chair, Dr. Michael Kennedy, has been so willing to share his advice and knowledge, as well as friendship, throughout my time at the University of Virginia. Thank you for taking in a misfit, and for all of the opportunities from which I've learned so much, and thank you especially for your support of this dissertation.

Each member of my dissertation committee has contributed greatly to the success of this study. Dr. John Lloyd shared with me a fraction of his vast expertise in the area of single-subject research designs, and has been exceedingly patient with my many, *many* methodological questions. Dr. Robert Berry has helped me better understand the "art" of teaching mathematics and his support for the study of disciplinary literacy and vocabulary in mathematics made this entire study possible. Last but definitely not least, Dr. Lysandra Cook graciously agreed to serve on this committee even before arriving in Charlottesville from halfway across the world, and I'm very grateful to her for bringing her expertise and experience in teacher learning to this work.

Getting to this point would not have been possible at all without the support of several other important people. I owe many thanks to Dr. Peter Youngs, for every email and every meeting and every bit of feedback, from coursework to comps to the job search. Additionally, I'm very grateful to the teachers who participated in this study, for being open to new ideas and for being so dedicated to supporting their students' learning.

Another group without whom I could not have completed this dissertation is the research team, past and present, that I've been lucky to be a part of. A special thank you to Dr. John Romig for taking all of my calls and texts and being patient with so many questions. Along with John, thanks also to Dr. Hannah Mathews and Dr. Wendy Rodgers for helping me find a place in MKRT and get through it alive; your guidance and support has meant more than you know, even when we're not all siting at a table in John's Bavaro office. Thank you to Shanna Hirsch for checking up on me and being a friend as well as a role model in fierceness and fearlessness. I would also like to thank Vickie VanUitert for all the hours spent observing teachers and helping me improve my skills as a CT Scan trainer.

Finally, I would like to thank my family and friends. My parents' unwavering faith in me was astounding, sustaining me throughout this process. I owe them, as well as my siblings and extended family, so much more than mere thanks. Last but certainly not least, to my dear friends of my cohort (and one or two honorary members), what exquisite timing we all had to be here together. You all have lifted me up in all the desperate and discouraging times. I can't thank you all enough for the love, laughs, and support over these four years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Table 1

Glossary of Acronyms

Acronym	**Full Name**	**Definition Provided on Page:**
NAEP	National Assessment of Educational Progress	1
SOL	Standards of Learning (Virginia)	1
CCSS	Common Core State Standards	3
CAP-PD	Content Acquisition Podcasts for Professional Development	9
CAP-TV	Content Acquisition Podcasts with Teacher Video	10
CAP-TS	Content Acquisition Podcasts Teacher Slides	11
CT Scan	Classroom Teaching Scan	12
QVI	Quality Vocabulary Index	16, 76, 88
NCTM	National Council of Teachers of Mathematics	17
EMVI	Explicit Mathematics Vocabulary Instruction	28
IMs	Implementation Markers (CT Scan)	29

Chapter 1: Introduction

Students with disabilities have consistently underperformed in mathematics, according to the National Assessment of Educational Progress (NAEP). Since the first administration of the NAEP in 1990, eighth grade mathematics scores have steadily increased by approximately 20 points over the course of those nearly thirty years. However, in 2017, the most recent administration of the NAEP, only 9% of eighth grade students with disabilities scored at or above “proficient” in mathematics (National Center for Education Statistics, 2018). That year marked the third administration in a declining trend in eighth grade mathematics scores of students with disabilities (NCES, 2018). Additionally, 69% of eighth grade students with disabilities scored “below basic” on the 2017 assessment, compared with 25% of eighth grade students without disabilities. On the NAEP, a score at or above “proficient” indicates that the student has demonstrated competency on the tested material; a score below this level indicates that the student has weaknesses and gaps in his/her understanding of the tested material. The most recent NAEP results mirror those from the previous assessment given in 2015 as well as 2013, and show an continuing trend in students with disabilities scoring “below basic” (NCES, 2015; NCES, 2018). Compared to the positive trend in scores since 1990 described above, the results from 2013, 2015, and 2017 mathematics assessments show a leveling off of achievement in across all eighth graders’ scores, but the increase in the number of eighth grade students with disabilities scoring “below basic” is a concerning trend.

At the state level, students with disabilities in general have demonstrated low passing rates on standardized mathematics assessments. In Virginia, for example, 47% of sixth grade students with disabilities did not pass the end-of-year mathematics assessment, called the Standards of Learning (SOL), in 2014-2015, 2015-2016, or 2016-2017 (Virginia Department of Education, 2017). Across those same years in Virginia, 61% of seventh grade students with disabilities and 59% of eighth grade students with disabilities also did not pass their end-of-year mathematics assessments (VDOE, 2017). The eighth grade NAEP assessment and the Virginia SOL middle grades exams in mathematics are similar in a number of ways; for example, both include multiple choice and multiple response items (the SOL tests refer to these as “technology-enhanced items”). A key difference between them is that the eighth grade NAEP mathematics assessment includes several constructed response items in which the student must explain or justify an answer; this is not currently a component of SOL exams. In a 2008 report that calculated school-level correlations between Virginia SOL scores and NAEP scores in eighth grade mathematics (based on 2003 testing data), there was a correlation of .63 (*SE* = .028) between the “proficient” levels on each assessment (McLaughlin et al., 2008).

Poor achievement outcomes in mathematics are not limited to students with disabilities. On the 2017 eighth grade NAEP mathematics assessment, students with disabilities’ average scale score remained in the “below basic” range, but the average scale score of students without disabilities was also low, just short of proficient (NCES, 2018). To illustrate, only 38% of eighth grade students without disabilities who took the NAEP in 2017 scored at or above the “proficient” level. Data from previous administrations of the assessment show the same or lower results for eighth grade students both with and without disabilities (NCES, 2015). Looking again at Virginia in particular, the average passing rates for the three end-of-year mathematics assessments given in 2015-2017 was 20% for all sixth grade students, 21% for all seventh grade students, and 21% for all eighth grade students (VDOE, 2017).

Current standards in middle grades mathematics expect students to be able to justify and communicate their reasoning on mathematical problems and tasks, which requires understanding and comprehension of the problem itself, but also requires enough familiarity, fluency, and knowledge of academic language and technical vocabulary to be able to compose and communicate those justifications. For students who struggle with these skills, whether due to a disability or not, one barrier to demonstrating competency and ability in mathematics may, in fact, be language (Pierce & Fontaine, 2009).

Academic Language and Vocabulary in Mathematics

The Common Core State Standards (CCSS) and the Principles and Standards of the National Council of Teachers of Mathematics (NCTM), both place an emphasis on students’ ability to communicate about mathematical concepts, such as measurement, and on their ability to explain their problem-solving process (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010); NCTM, 2000). As an example, one of the competencies listed in the CCSS for sixth grade mathematics is to “construct viable arguments and critique the reasoning of others” (2013, “Mathematical Practices”). Without strong vocabulary knowledge, both technical (i.e., domain-specific) and general, this skill is not attainable.

Assessments like the NAEP are designed to assess students across a wide range of mathematical topics, including measurement, computation, geometry, and statistics. For the majority of test items, however, students must read and understand several words and sentences in a given scenario to be able to answer the question (see sample NAEP and SOL test items in Figures 1.1 and 1.2). Figure 1.1 shows a sample item from the most recent (2017) eighth grade NAEP mathematics assessment. In this item, the student must read the scenario, the results table, and then the actual task directions. The task in this item is to correctly fill in the pie graph according to the numerical results on the table given. In order to understand and do what is being asked of him/her, the student needs to know the meaning of the word “sector” and, more importantly, how its meaning can be applied to constructing a pie graph.

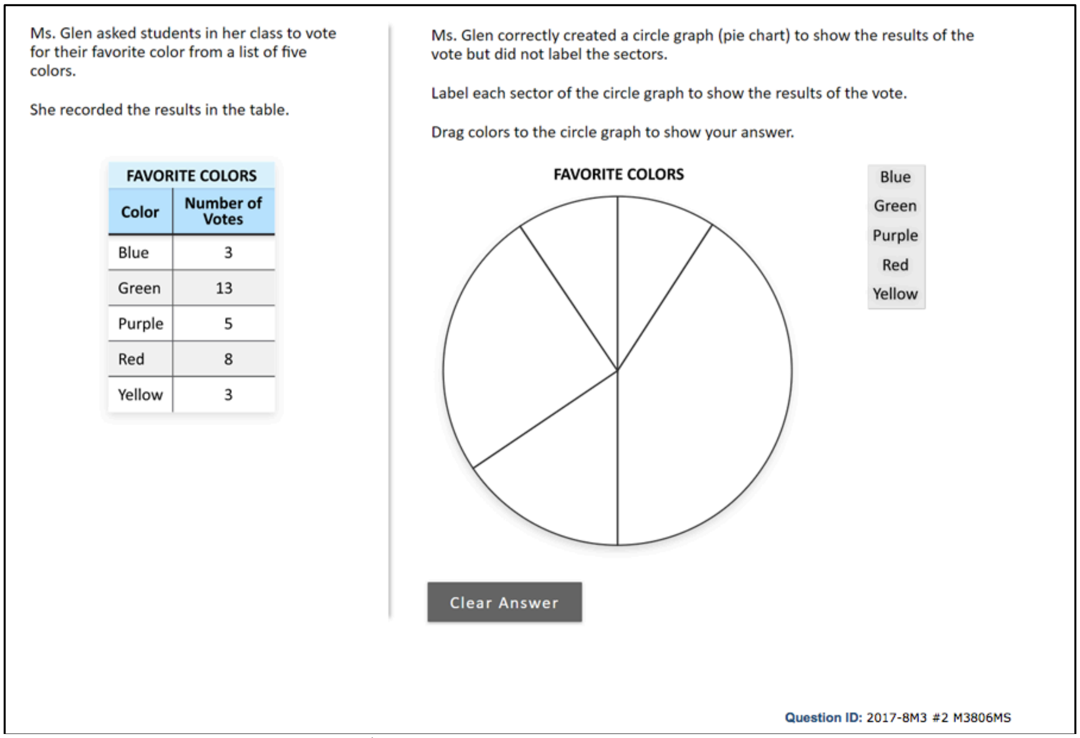


Figure 1.1. Sample NAEP 8th grade mathematics test item, incorporating both a story problem and graphical representations.

Which number in this list is NOT an integer?

$\frac{12}{4}$, -4^2, $\sqrt{25}$, -4.8

- A $\frac{12}{4}$
- B -4^2
- C $\sqrt{25}$
- D -4.8

Figure 1.2. Sample Virginia SOL 8th grade mathematics test item on integers.

Samuel bought 4 rolls of tape to seal boxes. Each roll contains 32.9 meters of tape. He uses 1.2 meters of this tape to seal each box. What is the total number of boxes Samuel can seal with these 4 rolls of tape?

- **A** 109 boxes
- **B** 130 boxes
- **C** 132 boxes
- **D** 157 boxes

Figure 1.3. Sample Virginia SOL 8th grade mathematics story problem test item.

Figure 1.2 shows another sample item from a released SOL exam. In this item, the student must know and apply the definition of an integer, similar to the NAEP example in Figure 1.1. Figure 1.3 shows another sample item from a released SOL exam. In this item, the student must read and comprehend the word problem, perform the correct computation, and then select the correct answer. This item requires less application of vocabulary than the NAEP item in Figure 1.1, but instead requires the student to recognize the words that indicate what operations to perform (and in what sequence) in order to correctly solve the problem. Unlike the term “sector” from the NAEP example, the relevant terms in the SOL item are “each” and “total,” hardly technical vocabulary, but important “clues” about what steps the student should do. Problems like this example can be particularly challenging for some students with disabilities, who tend to struggle with accurately applying solving strategies, such as using context clues (Cook, Collins, Morin, & Riccomini, 2019; Gersten et al., 2009; Riccomini, Smith, Hughes, & Fries, 2015).

From these examples, it is reasonable to infer that students' ability to comprehend and make accurate inferences about key terms is a factor in whether or not they are able to correctly solve problems and pass the assessment (Powell, Driver, Roberts, & Fall, 2017). Mathematics teachers, especially at the middle school level and with struggling readers (including students with disabilities) in their classes, are then challenged to provide effective instruction in vocabulary as well as in computation, procedures, and numeracy-based skills, but most mathematics teachers are not trained in the use of evidence-based practices for vocabulary instruction, nor in the use of evidence-based practices in special education, such as such as explicit instruction (Faggella-Luby, Ware, & Capozzoli, 2009; Wei, Darling-Hammond, & Adamson, 2010).

In middle school mathematics, as topics become increasingly complex, the specialized, domain-specific vocabulary can be particularly challenging for students with disabilities and other students who struggle with literacy and language (Faggella-Luby et al., 2009; Pierce & Fontaine, 2009). In addition to using technical, domain-specific words (e.g., *sector* or *integer*; see Figures 1.2 and 1.3), many words that are used in mathematics are used differently and with different meanings than the same word in another context (e.g., *round* or *volume*).

The Need For Impactful Professional Development

Despite the importance of vocabulary in mathematics and other content areas, there is little evidence that content-area teachers are adequately prepared to implement high-quality, evidence-based instruction in vocabulary (e.g., McKenna, Shin, & Ciullo,

2015) or, indeed, in much literacy instruction at all (Fagella-Luby et al., 2009). Findings from direct, observational studies as well as syntheses of studies of teachers' vocabulary instruction across content areas have largely shown that teachers implement few, if any, evidence-based practices for vocabulary, and rarely implement them with fidelity (e.g., Swanson, 2008; Swanson et al., 2012), especially in secondary grades (Ford-Connors & Paratore, 2015). However, these studies were conducted, for the most part, by observing reading instruction. Very few studies have relied on direct observation of mathematics teachers' vocabulary instruction, instead relying on teacher self-reporting of their instruction or measuring student outcomes instead of observing and recording instruction. As a result, little is known about the presence or quality of vocabulary instruction in mathematics.

Providing professional development and training in vocabulary instruction to teachers may seem like an obvious solution; however, not all professional development is created equal. Despite large amounts of money spent and a large amount of teachers' time devoted to teacher training and development, there is little evidence that traditional professional development is effective at improving teaching quality or student learning outcomes (Hill, Beisiegel, & Jacob, 2013; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). In recent years, however, there have been more attempts by researchers to identify and understand the features of professional development (including but not limited to professional development for mathematics teachers; e.g., Borko, Jacobs, Eiteljorg, & Pittman, 2008; Borko, Koellner, Jacobs, & Seago, 2011) that are effective, and where and when those features should be applied (Borko, 2004; Penuel, Fishman, Yamaguchi, & Gallagher, 2007).

Many education researchers have long called for a practice-based, teachers-as-learners approach to professional development (Ball & Cohen, 1999; Benedict, Holdheide, Brownell, & Foley, 2016; Desimone & Garet, 2015). A key feature of practice-based professional development is that it provides teachers with performance feedback and follow up coaching on the practices they are learning to implement (Desimone & Garet, 2015). Performance feedback is defined as data and other information collected via observation and provided to the teacher as a means of strengthening his/her practice (Scheeler, Ruhl, & McAfee, 2004; Sweigart, Landrum, & Pennington, 2015). Performance feedback provides information to the teacher and coach about behaviors or instructional components that the teacher is executing well, and information about behaviors or instructional components that he/she is not executing as desired or planned. Research has indicated that teachers respond positively to receiving this feedback and coaching, both in terms of actual improvement of teaching practice and social validity (i.e., whether they feel it was something that helped improve their teaching). Performance feedback was the variable of interest in the current study, and as such a more thorough review of its literature base is presented in chapter two.

Conceptual Framework

The aim of the current study was to improve the quality of middle school mathematics teachers' vocabulary instruction through a specific professional development model. The conceptual framework for the professional development used in the current study is largely based on cognitive apprenticeship (Collins, Brown, & Newman, 1989). Originally developed as an instructional model for K12 teaching, the cognitive apprenticeship framework fits particularly well with a practice-based approach to professional development, as its core pedagogical methods are modeling, coaching, and scaffolding. While it is rare that these methods are applied in empirical studies explicitly aligned with the cognitive framework, each has a strong empirical base as an approach to teacher learning. The current literature points to each of these methods as effective approaches for designing and implementing professional development (Darling-Hammond, Hyler, & Gardner, 2017; Desimone & Garet, 2015; Kraft & Blazar, 2017). A full description and review of the cognitive apprenticeship framework is provided in Chapter 2.

Professional Development and Performance Feedback in the Current Study

The professional development intervention that was utilized in this study to train and support middle school mathematics teachers' vocabulary instruction is in fact a professional development system, rather than a one-time training, called Content Acquisition Podcasts for Professional Development, or CAP-PD. CAP-PD was developed by Kennedy and colleagues at the University of Virginia, and has demonstrated success in improving specific aspects of instruction for middle school science teachers (Kennedy, Rodgers, Romig, Lloyd, & Brownell, 2017), classroom management (Kennedy, Hirsch, Rodgers, Bruce, & Lloyd, 2017) and middle school writing instruction (Romig, 2018). The current study aimed to extend this work into supporting middle school mathematics teachers.

The CAP-PD system. CAP-PD is a multi-component learning and practice system rooted in the cognitive apprenticeship framework (Collins et al., 1989; Kennedy, Rodgers et al., 2017; see Figure 1.5). At its core, the cognitive apprenticeship framework focuses on a cycle of modeling, coaching, and scaffolding. The individual components of the CAP-PD system are instructional modeling, materials to support classroom instruction, and performance feedback and coaching. Each of these has demonstrated effectiveness in previous, separate studies, though research on the system as a whole is promising but still quite new.

Instructional modeling. The instructional component of CAP-PD consists of brief multimedia vignettes that first provide instruction on a specific evidence-based practice, including its purpose and the steps necessary to implement it effectively, and then provide a video model of the practice in use (Kennedy, Rodgers, et al. 2017). These vignettes, called Content Acquisition Podcasts with Teacher Video, or CAP-TVs, are designed not only to provide high-quality instruction and modeling, but also to support the viewer's acquisition of declarative, procedural, and conditional knowledge about the topic or practice (Alexander, Schallert, & Hare, 1991). Much of teacher professional development is focused on delivering declarative, or factual knowledge about evidence-based practices (Desimone, 2009), but procedural (i.e., what steps to follow) and conditional (i.e., when and why to use the practice and steps) knowledge is essential to supporting teachers' implementation of evidence-based practices (Klingner, 2004). In the CAP-PD system, the researcher/coach directs the teacher to a specific CAP-TV, or even to a specific section of it, to teach them about evidence-based practice (e.g., explicit vocabulary instruction) that they should implement in their teaching. An example CAP-TV on student-friendly definitions can be viewed at: https://vimeo.com/143387419.

Materials to support classroom instruction. The next component of the CAP-PD system is customizable, ready-to-use, curriculum-aligned instructional materials that have the evidence-based practices for explicit mathematics vocabulary instruction from the CAP-TVs already embedded. In the CAP-PD system, these materials are called CAPs Teacher Slides, or CAP-TS. CAP-TS provide scaffolding for the instruction teachers receive on evidence-based practices. Each set of slides follows the same instructional sequence aligned with explicit instruction (e.g., review background knowledge, provide student-friendly definitions, use examples and non-examples; and embedded opportunities to respond to questions or other prompts; Archer & Hughes, 2011). In the CAP-PD system, the researcher/coach provides the CAP-TS to the teacher as a scaffold to help them implement evidence-based practices.

Performance feedback and coaching. Within the CAP-PD system is a built-in cycle for observation and feedback, designed to provide precise coaching to teachers as they are learning to implement evidence-based practices with fidelity. This component corresponds to the coaching element of the cognitive apprenticeship framework. This component of the CAP-PD system is arguably the most unique in the current research and implementations of teacher professional development, and was the focus of the current study. Unlike the CAP-PD system, most teacher professional development does not provide teachers with specific, individualized performance feedback on their implementation of what they were taught to implement. While it is a defining feature, its role within the CAP-PD system is less well understood than the other two components. The current study sought to provide information about the specific role that this component plays in changing teachers' instructional practice.

In this part of the CAP-PD system, the researcher/coach uses an observational instrument called the Classroom Teaching Scan, or CT Scan. The CT Scan is a low-inference tool that records teachers' moves in real-time, and allows the observer to record descriptive data about various aspects of the lesson, such as duration, frequency, and implementation fidelity (Kennedy, Rodgers, Romig, Mathews, & Peeples, 2018; see also http://www.classroomteachingscan.com/ctscan/). The CT Scan also produces detailed, data-based performance feedback that provides the teacher and coach with rich information about what the teacher did, when he/she did it, and the steps he/she used (or did not use) in implementing specific evidence-based practices (see an example from a middle school science lesson at http://www.classroomteachingscan.com/ctscan/timeline.htm?menus.txt&473). In recent CAP-PD empirical studies, the performance feedback output from CT Scan is accompanied by a written coaching explanation and narrative of the teacher's performance, emailed to the teacher (e.g., Kennedy, Rodgers et al., 2017; Romig, 2018).

The written coaching highlights successes and areas for improvement, and indicates specific resources within the CAP-TV library that will support that improvement. An example of the written coaching can be seen in Figure 2.6. This observation-feedback cycle continues for the duration of the CAP-PD intervention.

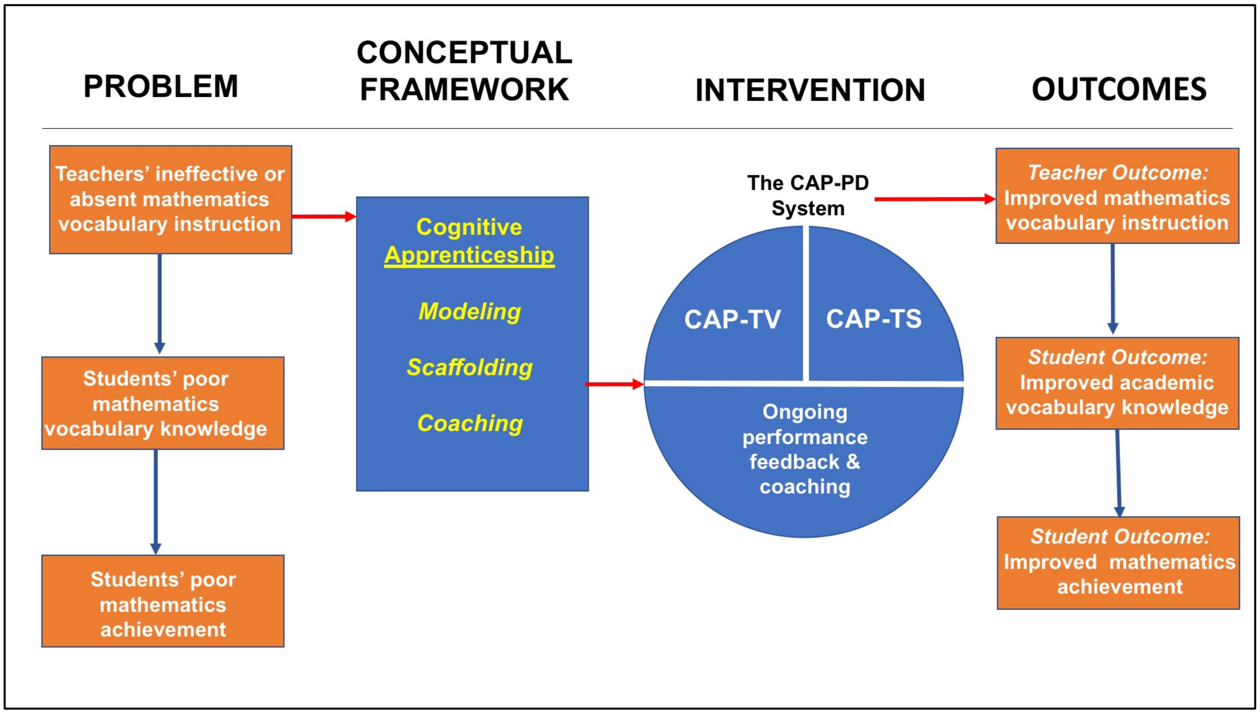


Figure 1.4. Overview of the current study.

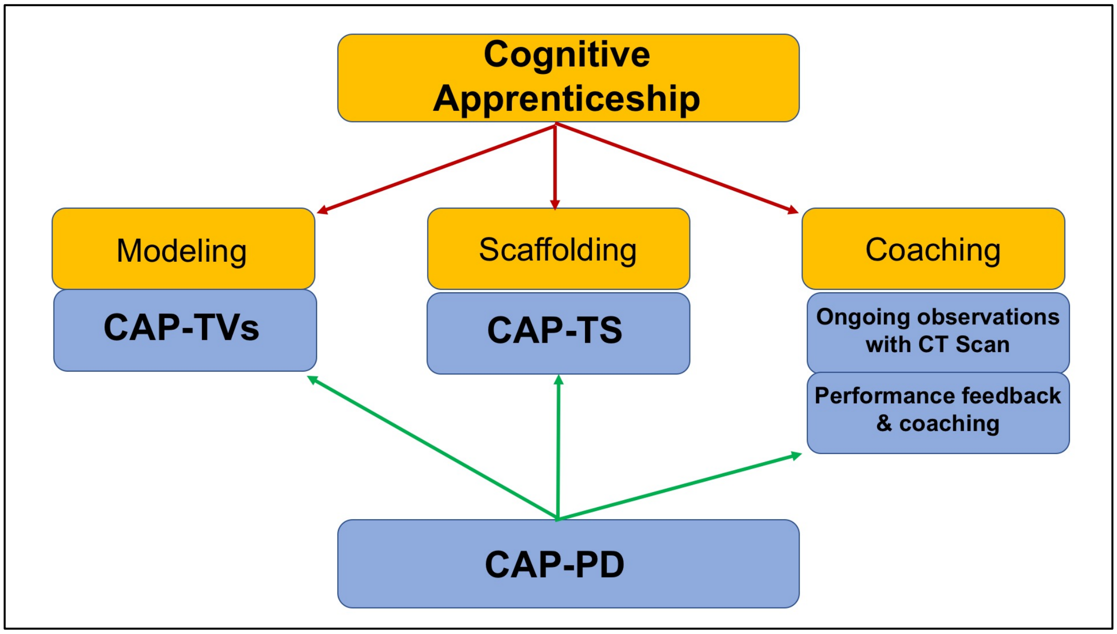


Figure 1.5. Mapping the CAP-PD system onto the cognitive apprenticeship framework.

The Current Study

The purpose of the current study was to investigate the specific role that the performance feedback and coaching component plays in the CAP-PD system. To do this, the current study used a single-case multiple baseline design with alternating treatments to examine the effects of CAP-TV and CAP-TS components with and without the performance feedback and coaching component on sixth grade mathematics teachers' explicit vocabulary instruction. This experimental design adheres to the guidelines stated in the What Works Clearinghouse single case design standards (Kratochwill et al., 2013). Four teachers were recruited from a local school system. Selection criteria for participants were that participants were currently (i.e., in the 2018-2019 academic year) teaching year-long grade 6 mathematics, and had at least one class period a day that included students with disabilities. Teachers were observed every other day (due to schedule constraints) throughout the study, excluding days when there was no planned vocabulary instruction (e.g., testing or special events).

In the baseline phase, teachers were observed and data was recorded, but none of the CAP-PD components were provided to them. Teachers moved from baseline to the intervention phase after a minimum of five documented observations and a stable, predictable baseline trend was established (Kratochwill et al., 2013). In cases where more than one teacher met these criteria on the same day, one was selected at random to start the intervention first, followed by the other following one more observation.

In the intervention phase, teachers were given access to the library of CAP-TVs for explicit mathematics vocabulary instruction and the CAP-TS template sent to them directly. They continued to have access to these resources throughout the intervention phase. For the first three observations in the intervention, teachers also received performance feedback and coaching on each of three consecutive lessons. The performance feedback consisted of outputs from CT Scan for that lesson, along with a written coaching email, consistent with previous CAP-PD studies. Following those three lessons were three more lessons when no performance feedback or coaching was given. This alternation between “treatment on” and “treatment off” continued throughout the intervention phase, for a minimum of three “cycles” per teacher (extenuating circumstances prevented some participants from meeting this minimum; more information is provided with the results in chapter 4). Finally, the experiment concluded with a maintenance phase in which none of the teachers received any performance feedback or coaching, and their access to the CAP-TVs was also removed.

The primary observational measure used in this study was the CT Scan. The variables of interest that the CT Scan will record are fidelity of implementation of vocabulary evidence based practices (as indicated by percentages of implementation markers recorded by the observer), the time spent teaching using each practice, and the Quality Vocabulary Index, or QVI, which expresses the teacher's implementation fidelity of vocabulary practices as a function of the time spent teaching vocabulary (the QVI and its derivation is explained in further detail later in this chapter). The experimental design of this study is fully explained in chapter 3.

This study provided a number of important insights about the role that performance feedback and coaching play in the CAP-PD approach to mathematics teachers' professional development in explicit vocabulary instruction. To date, CAP-PD has been used to improve a variety of teacher practices, especially explicit vocabulary instruction in middle school science, and the current study aimed to expand that body of research into middle school mathematics vocabulary instruction. Finally, the current study expands the work of previous studies in both the CAP-PD literature and the literature on the application of the cognitive apprenticeship framework to professional development settings.

Chapter 2: Literature Review

This chapter provides a review of the literature in each of the major constructs involved in this study: teaching academic language and mathematics vocabulary (especially to students with disabilities), the CAP-PD system, the cognitive apprenticeship approach to professional development, and the use of performance feedback in professional development. These reviews provide a rationale for the purpose of the current study, which is to investigate the specific role that the performance feedback and coaching component plays in the CAP-PD system by examining its effects on middle school mathematics teachers' vocabulary instruction. This chapter concludes with a summary of this rationale and the research questions that guide the current study.

Problem Statement

According to the National Center for Education Statistics' database of NAEP achievement, scores in eighth grade mathematics show a stagnant trend over the three most recent administrations of the assessment (NCES, 2018). This trend appears for students with disabilities as well as students without disabilities. On the 2017 eighth grade NAEP mathematics assessment, students with disabilities' average scale score remained in the "below basic" range; the average scale score of students without disabilities was higher, but was still in the "basic" range (NCES, 2018). Both of these scoring ranges indicate scores that are considered less than proficient. Combined, this means that 91% of students with disabilities who took the NAEP in 2017 and 63% of students without disabilities did not demonstrate proficiency on the eighth grade mathematics assessment.

There are a number of possible factors that could explain this phenomenon, but one area that merits examination is the quality of mathematics instruction in the middle grades, especially the quality of instruction for students with disabilities. The National Council of Teachers of Mathematics called for equity in instructional quality and expectations of all students regardless of status (i.e., English language learners, students with disabilities, or students from racial/ethnic minority groups; NCTM, 2012; NCTM, 2014). However, access to high-quality instruction is not adequate to improve outcomes for all students; as the NCTM points out, “educators must have the knowledge, skills, and disposition necessary to support effective, equitable mathematics teaching and learning” (NCTM, 2014, p. 1). Professional development that supports and expands teacher knowledge and ability to address the needs of all learners is perhaps one way to address the problem of poor achievement outcomes in mathematics.

Although the majority of students with disabilities receive most of their content instruction in general education settings (62.2%; NCES, 2016), teachers across content areas are often underprepared to implement the most effective research-based approaches to provide instruction for these students. Many content-area teachers simply lack training in teaching students with disabilities, as it is not typically an area of focus in content-area teacher preparation programs. As a result, they may not have a repertoire of effective strategies and approaches from which to draw (Holdheide & Reschly, 2008). Subsequent professional development opportunities to develop this repertoire may be limited, if available at all, as these opportunities tend to vary widely by state and district, not to mention quality (Barton, Whittaker, Kinzie, DeCoster, & Furnani, 2017; Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009; Desimone, 2009; Yoon et al., 2007).

A goal of the current study was to provide a high-quality professional development that will expand mathematics teachers' repertoire of effective, evidence-based practices for teaching academic vocabulary. The foundation for these practices is explicit instruction, which is one of the most effective approaches for teaching students with disabilities (Hughes, Morris, Therrien, & Benson, 2017). Research in mathematics teaching has explored and supported the use of explicit instruction for teaching a number of mathematics skills and concepts (e.g., Hudson, Miller, & Butler, 2006; Montague, Enders, & Dietz, 2011; Witzel, Mercer, & Miller, 2003; Xin, Jitendra, & Deatline-Buchman, 2005). Additionally, the report from the National Mathematics Advisory Panel (2008) indicated that systematic and explicit instruction is effective for students with difficulties in mathematics, and Gersten et al. (2009) recommended the use of explicit instruction as an intervention in mathematics, supported with strong evidence by What Works Clearinghouse standards.

Mathematics teachers as well as other content-area teachers, especially in the middle grades, may not have been adequately prepared to provide explicit vocabulary instruction and may find incorporating vocabulary instruction to be a challenge as far as content knowledge and time constraints (Fagella-Luby et al., 2009). These teachers often have a great deal of curricular ground to cover, and may feel that time devoted to vocabulary instruction may not be the best use of instructional time (Bay-Williams & Livers, 2009). In a study involving secondary social studies teachers, O'Connor and colleagues (2015) found that when the time spent on explicitly teaching academic vocabulary (i.e., student friendly definitions) was decreased to 5 minutes from 15 minutes, the researcher-developed vocabulary knowledge assessments showed that students both with and without disabilities learned a smaller percentage of the words they were taught in that reduced time frame. Other research has shown that secondary content area teachers are not the only ones who are not likely to spend a lot of time on vocabulary instruction. A study by Cunningham and colleagues (2009) surveyed first grade teachers about what components of language arts instruction they would spend time on if they were to design an "ideal" 90 minute block. The researchers identified 13 instructional categories across the teachers' responses and found that on average, the teachers would spend just 1.6% of that 90 minute block – approximately 2 minutes – on vocabulary instruction of any kind.

It is also challenging for teachers to incorporate explicit vocabulary instruction to established (often state- or district-adopted) instructional programs and curricula for middle school mathematics (Hudson et al., 2006). However, it is also likely that most secondary content area teachers lack training and preparation in effective vocabulary instructional strategies and how to implement them in ways that maximize the instructional time that they devote to explicitly teaching vocabulary (Bay-Williams & Livers, 2009). Thus, there is a pressing need for high-quality professional development that will support teachers' smooth integration of evidence-based practices for vocabulary instruction alongside their content and other pedagogical demands (Dingle, Brownell, Leko, Boardman, & Haager, 2011; McKenna et al., 2015; Siuty, Leko, & Knackstedt, 2018). Linking what is known about effective professional development with evidence-based instructional practices such as explicit instruction is crucial. The aim of the current study is to provide a high-quality professional development intervention designed to support and improve the explicit vocabulary instruction of middle school mathematics teachers who are serving students with disabilities in their classes.

The next sections of this chapter will describe the instructional areas of interest of the current study and the research base for each of those areas. As this study is intended to support and improve mathematics instruction for students with disabilities, the research evidence for this need is also discussed below. Following that, the conceptual framework (i.e., cognitive apprenticeship) for the professional development intervention used in the current study is described in detail, as well as the intervention itself. This chapter concludes with a summary of the study's rationale and research questions.

Teaching Academic Vocabulary in Mathematics

With an increasing emphasis on building students' conceptual understanding in mathematics, many programs and curricula are not focused on building students' academic language (Lesaux, Kieffer, Faller, & Kelley, 2010). Like conversational language, academic language includes its own syntax, grammar, and vocabulary, but these functions are limited to a specific discipline, such as mathematics (Cummins, 2000; Powell & Driver, 2015). Mathematical proficiency is not limited to procedural knowledge (i.e., computation and problem solving) or logical reasoning. Understanding and using academic language is also an important component of mathematical proficiency (Riccomini et al., 2015). However, without a strong foundation in the academic language of mathematics, the higher-order thinking and reasoning skills that many standards demand is not possible (Riccomini et al., Schleppegrell, 2007; van der Walt, Maree, & Ellis, 2008). The processes through which students solve mathematics problems, even "real-world" problems that do not necessarily directly involve language (i.e., word problems; Xin et al., 2005), require knowledge of language and vocabulary in order to apply the concepts and skills needed to solve the problems (Morin & Franks, 2009). Research has also suggested that middle school students with and without mathematics difficulties struggle to "translate" the literal meanings of terms into more conceptual communications, such as algebraic equations (Capraro, Capraro, & Rupley, 2010). There are a number of reasons why academic language and vocabulary are obstacles for students with disabilities in mathematics; understanding these obstacles is essential to addressing the issue of improving mathematics instruction for these students.

Students With Disabilities' Obstacles to Learning Academic Vocabulary

An area in which many students with disabilities often struggle with mathematics is language, yet it is not often taught directly in many mathematics classrooms (Fisher &

Frey, 2008; Hattie, Fisher, & Frey, 2017). Research in explicit instruction and in mathematics education suggests that taking an explicit instruction approach to teaching mathematical language and vocabulary could be effective for students with mathematics difficulty or those with disabilities (Pierce & Fontaine, 2009; Powell et al., 2017). Low performance on mathematics assessments that require understanding of academic language and other vocabulary may actually indicate the students' difficulty with language, rather than their actual mathematics competency (Pierce & Fontaine, 2009; Thompson & Rubenstein, 2000). Given that the majority of items on mathematics assessments such as the NAEP are text-based problems, it is reasonable to view students' difficulty with language and vocabulary as part of the reason for continued poor mathematics achievement outcomes (Powell et al., 2017; Powell & Nelson, 2017).

Furthermore, teaching standards such as the National Council of Teachers of Mathematics Principles and Standards (NCTM; 2000) and the Common Core State Standards (CCSS; 2010) expect that students will develop skills in argumentation, discourse, and communication about mathematics, beginning in the primary grades and continuing through high school (Powell et al., 2017). Providing explicit instruction in mathematics language and vocabulary is key to supporting all students' abilities to meet these standards and expectations (Doabler, Fien, Nelson-Walker, & Baker, 2012; Riccomini et al., 2015; Xin et al., 2005).

In general, students with disabilities do not employ word-learning strategies, such as using context clues, and do not often read a wide range of texts independently (Bryant, Goodwin, Bryant, & Higgins, 2003; Fagella-Luby et al., 2009; Jitendra, Edwards, Sacks, & Jacobson, 2004). Many students with disabilities also struggle with learning and applying cognitive and metacognitive strategies, such as problem-solving strategies in mathematics (Montague et al., 2011; van Garderen, 2008). Building students' ability to understand and use academic language is essential for developing understanding of mathematical concepts, ability to work with higher-order thinking skills, ability to communicate with others about their strategies and processes for problem-solving, and for increasing their overall achievement in mathematics (Hughes, Powell, & Stevens, 2016; Powell & Driver, 2015; Powell et al. 2017). Many students with disabilities, even those whose area disability is not formally identified as mathematics, struggle with language acquisition, fluency, and processing (Jitendra et al., 2004; Morin & Franks, 2010). This makes learning in a language or text-dependent environment, as almost all secondary classrooms are, very challenging for these students (Kennedy, Deshler, & Lloyd, 2015).

In secondary grades, the issue of ambiguous or confusing language in mathematics persists, but is compounded by the inclusion of domain-specific vocabulary (e.g., *divisor*) and more words with multiple meanings (e.g., *volume*). Domain-specific vocabulary consists of terms that are only used for a limited purpose, "decontextualized and cognitively demanding" (Powell & Driver, 2015, p. 222). These terms are also referred to as "technical" vocabulary (Baumann & Graves, 2010; Fisher & Frey, 2008) or "Tier 3" vocabulary (Beck, McKeown, & Kucan, 2013). Because these terms tend to be used more rarely (i.e., only in the specific domain), students have fewer opportunities to engage with the terms.

Mathematics Vocabulary Instruction in the Secondary Grades

A number of studies have examined interventions to support elementary students' acquisition and use of mathematical vocabulary (Kostos & Shin, 2010; McAdams, 2011; Powell & Driver, 2015). Interventions such as tutoring with explicit instruction (Powell & Driver, 2015), journaling (Kostos & Shin, 2010), and direct instruction in the general education classroom (McAdams, 2011). However, there are far fewer empirical, evidence-based studies in this area than there are articles that offer "suggestions" for teaching, most of which are adaptations of reading strategies (Hebert & Powell, 2016).

The literature base for empirically-tested interventions in middle grades mathematics classrooms is considerably more limited. Temple and Doerr (2012) examined the ways in which a 10th grade mathematics teacher encouraged her students to use precise, technical mathematics vocabulary. They found that there were two key "moves" that the teacher used that led to students' frequent and prolonged use of newly learned terms: activating students' prior knowledge (i.e., previously learned mathematical concepts and terms) and increased use of questions and feedback to "push" students to use technical vocabulary in conversations and discussions. During the lesson, the researchers counted the frequency of questions and feedback statements as well as the students' responses, and later coded these into more specific types of questions and feedback statements in order to look for patterns in the way the teacher engaged the students in discussions around a mathematical concept. They noted that the teacher used two different patterns of interaction depending on the type of knowledge she intended the students to acquire. When students were acquiring new knowledge, the teacher-led discussion had a "funneling" pattern; in other words, the teacher prodded and encouraged students to use new terms. When the students were building on previously learned material to construct new understanding, the teacher's discussion was intended to have students "explain their reasoning and build on each other's contributions," which the authors called a "focusing" pattern (p. 293). While the researchers found that students' accuracy and fluency in discussing the lesson's core mathematical concept did increase, they made this claim based off of a single observation of a single teacher.

Additionally, although their study was not specifically focused on mathematics vocabulary instruction, Capraro and Joffrion (2006) explored the ways in which middle school students struggle to build necessary conceptual understanding in algebra. This study included 668 seventh and eighth grade students and 25 teachers. The researchers focused on the ways in which students were able (or not) to recognize and use conceptual understandings of mathematical terms and procedures needed to solve a word problem, and turn them into mathematical symbols (i.e., linear equations). Students were assessed on a variety of problems that asked them to generate linear equations based on a sentence or two that described a scenario. For example, students were asked to generate a linear equation that would correctly solve a statement such as, "Sam is exactly 3 years older than Michael" (p. 157). After scoring the assessments and analyzing students' errors, they found that students whose procedural errors reflected their lack of understanding of concepts and technical vocabulary. Instead, they found through the error analysis that these students often relied on a key word strategy, shortcut, or “algorithm” they had been taught in order to “translate from English to math” (p. 152), which weakens the connections students make between mathematical concepts and procedures. They note that this finding points to the need for teachers to teach technical mathematical vocabulary directly and provide opportunities for students to apply it to mathematical procedures and use it in discussions, as students will likely not be exposed to these domain-specific, technical words otherwise.

Definition of Explicit Mathematics Vocabulary Instruction in the Current Study

For the purposes of this study, the operational definition of explicit mathematics vocabulary instruction and its component evidence-based practices are supported by research across reading, vocabulary, and mathematics instruction. The current study defines explicit mathematics vocabulary instruction as a set of evidence-based practices that includes: (a) directly teaching students a student-friendly definition of a word (Beck et al., 2013), (b) modeling the correct use of the word (Archer & Hughes, 2011; Beck et al., 2013), (c) providing a morphological analysis of the word, when appropriate (Baumann, Edwards, Boland, & Font, 2012; Bos & Anders, 1990; Graves, 2006; Harris, Schumaker, & Deshler, 2011), (d) providing examples, including representations and/or symbols of the word and (e) providing non-examples of the word when appropriate (Boardman et al., 2008; Frayer, Frederick, & Klausmeier, 1969; Greenwood, 2002; Stahl, 1999), (f) leading classroom discussions about mathematical concepts (Beck et al., 2013; Schleppegrell, 2007; Temple & Doerr, 2012), and (g) directly showing semantic word parts and semantic relationships between words (Baumann et al., 2002, Bos & Anders, 1990; Graves, 2006). In order to aid teachers' use of explicit mathematics vocabulary instruction, each practice has a number of implementation markers, or IMs, that come directly from the supporting research literature on each practice.

This definition brings together Archer & Hughes' (2011) explicit instruction framework (Hughes, Morris et al., 2017) with key elements of high-quality vocabulary instruction (Beck et al., 2013; Ebbers & Denton, 2008; Graves, 2006; Jitendra et al., 2004; Kame'enui & Baumann, 2012). These six evidence-based practices (see Table 2.1, below) will form the core content of the professional development delivered in the current study.

Table 2.1

Explicit Mathematics Vocabulary Instruction: Evidence-Based Practices and their Implementation Markers.

Explicit Mathematics Vocabulary Instructional Practice	Implementation Markers	Evidence Base
1. Student-Friendly Definitions	(a) Cue students' attention to instruction (b) Review background knowledge (b) Use clear language (c) Monitor students' understanding (d) Repeat essential information (e) Anticipate misconceptions or misinterpretations (f) Include visual cue or anchor	Beck, McKeown, & Kucan, 2013; Jitendra et al., 2004; Pierce & Fontaine, 2009
2. Use Examples (including mathematical representations)	(a) Cue students' attention to instruction (b) Use clear language (c) Monitor students' understanding (d) Give explicit explanation and connection to definition (e) Include visual cue or anchor	Boardman et al., 2008; Dunston & Tyminski, 2013; Frayer et al., 1969; Greenwood, 2002; Stahl, 1999; Thompson & Chappell, 2007
3. Use Non-Examples (when appropriate)	(a) Cue students' attention to instruction (b) Use clear language (c) Monitor students' understanding (d) Make clear connections to definition and correct examples (e) Anticipate misconceptions or misinterpretations (i.e., point out similarities to correct example) (f) Give explicit explanation of differences (i.e., why it is a non-example) (g) Include visual cue or anchor	Boardman et al., 2008; Frayer et al., 1969; Greenwood, 2002; Stahl, 1999
4. Use a Morphological Approach (when appropriate)	(a) Cue students' attention to instruction (b) Use clear language (c) Monitor students' understanding	Baumann et al., 2002; Bos & Anders, 1990; Ebbers & Denton,

	(d) Define all parts of term (e) Reunite term and connect to definition (f) Include visual cues or anchors for each part *and* whole term	2008; Graves, 2006; Harris, Schumaker, & Deshler, 2011; Thompson & Rubenstein, 2000
5. Indicate Semantic Relationships	(a) Cue students' attention to instruction (b) Use clear language (c) Monitor students' understanding (d) Explain similarities among terms (e) Explain differences among terms (f) Include visual cues or anchors for each semantic feature or related term	Baumann et al., 2002; Bos & Anders, 1990; Dunston & Tyminski, 2013; Ebbers & Denton, 2008; Graves, 2006
6. Facilitate Academic Discussion of Terms	(a) Establish relevance of discussion (b) Provide generative topics or questions (b) Use in appropriate context (b) Use clear language (c) Call on students all over the room (d) Give corrective feedback as needed (e) Make connections to students' prior knowledge and/or experiences (f) Encourage student use of relevant technical/precise terms	Beck et al., 2013; Capraro, Capraro, & Rupley, 2010; Schleppegrell, 2007; Temple & Doerr, 2012; Thompson & Rubenstein, 2000

The lack of current research on explicit vocabulary instruction in middle school mathematics illustrates the need for studies of high-quality professional development that can potentially help teachers incorporate explicit mathematics vocabulary instruction into instruction. The current study will examine the effect of a professional development system called Content Acquisition Podcasts for Professional Development (CAP-PD) that is rooted in cognitive apprenticeship (Collins et al., 1989) in order to provide ongoing instruction, observation, feedback, and coaching to middle school mathematics teachers.

In the sections that follow, I describe the CAP-PD process, and then explain its connections to the cognitive apprenticeship conceptual framework.

The CAP-PD System

Content Acquisition Podcasts for Professional Development, or CAP-PD, is a multimedia system of instructional modeling, including CAPs with Teacher Video (CAP-TVs), materials to support classroom instruction (CAPs with Teacher Slides, or CAP-TS), and an ongoing cycle of observation, performance feedback, and instructional coaching. The CAP-PD system is new, and so empirical evidence on its effectiveness is still somewhat limited, but it is very promising. CAP-PD was designed around the cognitive apprenticeship framework, and so many of the elements of the framework described in the previous sections are reflected here. In this section, each component and each relevant literature base of the CAP-PD system is described. In addition, this section reviews the existing studies to date of the CAP-PD system as a whole.

Content Acquisition Podcasts for Teachers with Embedded Video (CAP-TVs)

The instruction and modeling in the CAP-PD system is provided by Content Acquisition Podcasts – Teacher Videos, or CAP-TVs. Each of these brief instructional vignettes include direct, multimedia instruction about a specific evidence-based practice, followed by a number of short video clips, recorded in real K-12 classrooms, of teachers implementing that evidence-based practice. The multimedia instruction section is designed to build the learner's declarative, or factual knowledge of a specific topic or practice, similar to domain knowledge in the cognitive apprenticeship framework. This type of knowledge is key in any learning context, but many approaches to teacher professional development stop at the point of delivering factual information, without supporting additional types of knowledge that are just as necessary if the content is to be implemented correctly and effectively (Alexander, et al. 1991; Desimone, 2009). This section of a CAP-TV is essentially a narrated PowerPoint presentation, and engages the viewer through visual and auditory means while keeping his/her cognitive load processing at a productive level to enhance learning. This approach is rooted in cognitive load theory (DeLeeuw & Mayer, 2008) and in Mayer's cognitive theory of multimedia learning (CTML; 2009).

Cognitive load and Mayer's multimedia instructional design principles. CTML and Mayer's design principles for multimedia-based instruction (Mayer, 2008) drive the creation of each CAP-TV. These principles are applied in order to address the three aspects of cognitive load theory: extraneous, intrinsic, and germane (Sweller, Van Merriënboer, & Paas, 1998; Van Merriënboer & Sweller, 2005). Extraneous processing is detrimental to the learner's ability to take in the key content during instruction. Too much text on a screen, differences between text on screen and the narration, too much information in general, irrelevant images, or other distracting (i.e., extraneous) aspects of a multimedia presentation cause the learner to have to spend cognitive resources on sorting out the extraneous from the important, rather than on taking in the really important information (Mayer, 2008). The process of learning the content that is relevant to the actual learning objective is called intrinsic processing (DeLeeuw & Mayer, 2008).

The complexity of that content is generally what affects a learner's intrinsic processing load; the more complex the content, concept, or skill, the higher the learner's intrinsic cognitive load will be (Leppink, Paas, Van der Vleuten, Van Gog, & Van Merriënboer, 2013). If the demands on the learner's extraneous and intrinsic cognitive processing are lessened, then the learner is able to thoroughly process and organize the new information and relate it to previously learned material; this type of processing is associated with germane cognitive load.

In accordance with Mayer's cognitive theory of multimedia learning (2008), CAP-TVs mitigate extraneous and intrinsic cognitive processing and boost germane cognitive processing by presenting only essential content about each topic or teaching practice, and by limiting the text and images on screen as well as the narration to only those most relevant to the topic or teaching practice. In addition, text and images are carefully placed together on the screen in order to support the learner's association of the two. Narration is always synced to the text that appears on the screen, and is delivered in a conversational rather than formal tone. Finally, both visual and auditory cueing are used throughout each CAP-TV to signal to the learner that key content is going to be presented. The first half (approximately 8 minutes) of this example CAP-TV on using student-friendly definitions shows these principles in action (note: this CAP-TV is from a previous CAP-PD study): https://vimeo.com/143387419.

Use of authentic modeling in CAP-TVs. Once the CAP-TV has presented the essential content about an evidence-based practice and the steps necessary for its correct implementation, a number of short video clips are used to show the practice in action. Combined with the declarative knowledge provided in the first section of the CAP-TV, the modeling video in each CAP-TV builds both conditional and procedural knowledge, both of which are essential for supporting teachers' implementation of the evidence-based practice taught in the video (Alexander, et al. 1991; Kennedy, Rodgers et al., 2017). Conditional knowledge can be defined as the learner's knowledge about when and in what context the practice should be used, while procedural knowledge is concerned with the actual steps and sequence that the teacher needs to use when implementing the practice. In the same example video provided above, the latter half (approximately 8 minutes) provides an example of this kind of modeling.

Research base on CAP-TVs. There is already a strong body of empirical research on the effectiveness of CAP-TVs for both preservice teacher preparation and in-service professional development, both as a stand-alone intervention and as a part of the CAP-PD system. One of the first studies to examine stand-alone CAP-TVs was a randomized control trial that found that preservice teachers who learned about an instructional practice for vocabulary (Intensifying Vocabulary Instruction; IVI) by watching CAP-TVs implemented the practice with greater fidelity than those who learned about the practice by reading a text (Ely, Kennedy, Pullen, Williams, & Hirsch, 2014). Participants in the CAP-TV group also outscored the comparison group on a measure of their knowledge of IVI.

Additional follow-up studies of CAP-TVs and preservice teachers expanded this initial work, using other evidence based practices as the core instructional content and comparing CAP-TV to additional modes of learning (i.e., text-based and lecture formats) and found similar results. Alves et al. (2017) found that preservice teachers who received instruction on vocabulary evidence-based practices increased their knowledge of those practices and included more of the implementation steps of those practices in a sample lesson (d = .44), than did the participants that received instruction without CAP-TVs. Romig et al. (2018) completed a similar study with a writing strategy as the evidence-based practice of interest, and found large effect sizes for both a knowledge measure and an implementation steps measure (CAP-TV group vs lecture group d = 1.15; CAP-TV group vs assigned text group d = 1.92). By using a variety of instructional evidence-based practices as independent variables, this line of research on the use of CAP-TVs in preservice teacher education settings has demonstrated not only the effectiveness, but also the flexibility of the CAP and cognitive apprenticeship approach.

While the research to date on the use of CAP-TVs with preservice teachers has shown that CAP-TVs are an effective instructional tool, there is also evidence for its effectiveness with in-service teachers in professional development settings. Prior to the development of the CAP-PD system, Kennedy, Hirsch, Rodgers, Bruce, and Lloyd (2017) compared the use of CAP-TVs as professional development presentation to a traditional workshop-style PD. In both settings, the PD delivered instruction on the use of three evidence-based practices for classroom management: opportunities to respond

(OTRs), behavior-specific praise statements (BSPs), and pre-correction statements. Across all three of these practices, teachers in the CAP-TV group implemented them with greater frequency than the teachers who received traditional PD. Although the sample was relatively small (n = 6 in both groups), there were large effect sizes for the CAP-TV group on the frequency of each of the three practices when the teachers were observed (OTRs d = 2.03; BSPs d = 1.67; pre-correct statements d = 1.99).

In the current study, CAP-TVs were provided in a similar way to previous CAP-PD studies by Kennedy and colleagues. The practices that are the focus of the CAP-PD process in the current study are called, collectively, explicit mathematics vocabulary instruction. The practices are: (a) using a student friendly definition, (b) using examples (including mathematical representations), (c) using non-examples, (d) using a morphological approach, (e) identifying semantic relationships among words, and (f) facilitating academic discussions. Table 2.2 provides hyperlinks to each of the CAP-TVs that were developed for each of these practices for this study. The definitions of each practice are provided in Table 2.1. Based on the fact that CAP-TVs have already demonstrated effectiveness across a variety of settings, CAP-TVs will serve as a comparison treatment in the experimental design of the current study, while performance feedback and coaching will serve as the treatment of interest.

Table 2.2

List of Explicit Mathematics Vocabulary Instruction CAP-TVs and Current Web Links

Evidence-Based Vocabulary Practice	Current Web Link
Using a Student-Friendly Definition	https://vimeo.com/288619218/24d3fe4e82
Using Examples & Non-Examples	https://vimeo.com/290174251/0586221416
Using a Morphological Approach	https://vimeo.com/320610681/535b7da9fb
Highlighting Semantic Relationships	https://vimeo.com/320612903/8c23b4d16f
Facilitating Academic Discussions	https://vimeo.com/320614483/ca4963b434

Content Acquisition Podcasts Teacher Slides (CAP-TS)

The CAP-Teacher Slides (CAP-TS) are the scaffolding component of the CAP-PD system. Similarly to CAP-TVs, the CAP-TS are PowerPoint presentations and adhere to many of Mayer's (2008) design principles, but are not pre-recorded or narrated. Another key difference between these two components is that CAP-TS are designed for use in the teacher's actual classroom instruction, rather than as a PD content resource on teaching practice. CAP-TS are customizable to fit a teacher's instructional needs, such as pacing and content, and teachers can use them to provide explicit instruction on a new topic or to review background knowledge linked to a new activity or instructional unit (Kennedy, Rodgers, et al. 2017). CAP-TS grew from Kennedy's initial development of the Content Acquisition Podcasts (CAPs) for students (Kennedy, Thomas, Aronin, Newton, & Lloyd, 2014). In their current iteration in the CAP-PD system, CAP-TS are still built around Mayer's instructional design principles (2008), and incorporate elements of both cognitive apprenticeship (Collins, et al. 1989) and explicit instruction (Archer & Hughes, 2011). Each CAP-TS has a number of evidence-based practices already embedded, such as signaling or cueing new or important information, monitoring student understanding, student opportunities to respond, and using visual anchor images alongside text. When a teacher downloads a CAP-TS, they are able to input their own language, sequence, and content to fit their students' needs and learning objectives.

Research base on CAP-TS. Recent CAP-PD studies have used both fully completed CAP-TS lesson presentations on science vocabulary (e.g., Kennedy, Rodgers et al., 2017) as well as CAP-TS templates that allow teachers to input their own content (e.g., Romig, 2018). In both of those studies, the teachers responded positively to these materials, used them frequently, and were excited to use them in their classes even beyond the duration of the study.

In a single-case multiple baseline study of the full CAP-PD package on middle school science teachers' vocabulary instruction, Kennedy, Rodgers et al. (2017) provided full vocabulary PowerPoint presentations to teachers. They found that the teachers were able to easily incorporate these slides into their lessons, which in turn enhanced their explicit vocabulary instruction, largely because of the practices embedded in the PowerPoints. A CAP-PD study by Romig (2018) took a slightly different approach to providing this scaffolding, by giving teachers a template, rather than a fully completed lesson, to use in their planning and teaching of writing instruction. That study found that the teachers used the CAP-TS template to support their use of modeling in their writing lessons during the intervention and maintenance phases of the study, and continued to use it after the end of the study (Romig, 2018). It is important to note that neither of these studies examined the specific role of the CAP-TS as a component of the CAP-PD system, but employed the CAP-TS as an essential scaffold, according to the cognitive apprenticeship framework.

In the current study, the participants were given a CAP-TS template that they could complete with the relevant mathematics vocabulary for each lesson. They were also given a full mathematics vocabulary lesson as a CAP-TS presentation, to serve as an example. Like the CAP-TVs, CAP-TS were not an experimental variable of interest, but were provided as part of the CAP-PD system.

Observation, Performance Feedback, and Coaching.

The observation-feedback-coaching component in the CAP-PD system functions as its own cycle within the larger system. As teachers are receiving the instruction and materials from CAP-TVs and CAP-TS, they are being observed and given performance feedback, and in the coaching they receive they may be directed back to a specific section of a CAP-TV to review, or directed to a different CAP-TV to expand their practice. The observation-feedback-coaching cycle continues within the CAP-PD system throughout the intervention. In recent studies, teachers have received as many as six observations over the course of the CAP-PD intervention (Kennedy et al., in preparation).

Giving teachers ongoing observation, feedback, and coaching is not a feature of most traditional PD programs (Darling-Hammond, et al. 2017; Desimone, 2009), despite evidence that it has a positive impact on instructional quality (Fallon, Collier-Meek, Maggin, Sanetti, & Johnson, 2015; Scheeler et al., 2004; Solomon, Klein, & Politylo, 2012). For many teachers, the majority of the feedback they receive is for one-time evaluative purposes, and is not necessarily tied to ongoing improvement (Sanetti, Fallon, & Collier-Meek, 2012).

Cognitive Apprenticeship for Professional Development

While some research on teacher learning and professional development has found success in using some of the methods and components of cognitive apprenticeship, the body of that research is quite small. No studies in the documented literature have studied an application of the full cognitive apprenticeship framework, but the following sections review the research that has included components of the framework.

Collins et al. (1989) defined cognitive apprenticeship as a "focus of the learning-through-guided-experience on cognitive and metacognitive, rather than physical, skills and processes" (p. 457). They initially intended cognitive apprenticeship to be an approach to teaching K-12 students. Taking inspiration from traditions of apprenticeship as vocational training (such as an apprentice tailor, or printmaker, etc), they applied the idea to school subjects. Making problem-solving and other learning strategies transparent provides a model for the learner, and opportunities for the teacher to provide corrective feedback and scaffolding, as well as extensions of the skill when the student is ready.

Two major concepts lie at the heart of the cognitive apprenticeship approach to learning. First, cognitive apprenticeship strives to "make thinking visible" (Collins, Brown, & Holum, 1991). As the framework was initially intended as an approach to teaching K-12 students, cognitive apprenticeship as outlined by Collins et al. (1989) sought to externalize cognitive processes that make up complex tasks or knowledge, with the teacher making his/her thinking "visible" to students (i.e., apprentices). Second, the concept of *situated learning* is also a key component of the cognitive apprenticeship framework. This concept states that learning occurs within, and is entirely dependent on its physical and social contexts, and should ultimately be generalizable beyond declarative (or factual) knowledge (Collins et al., 1991; Dennen & Burner, 2008; Sawyer & Greeno, 2009).

Collins et al. (1989) laid out their framework as "characteristics of ideal learning environments," along four dimensions: content, methods, sequence, and sociology (p. 476). Within each of these domains they included a number of characteristics. The following sections briefly define these domains and characteristics and discuss how each might be applied to a professional development context. Then, the core components of cognitive apprenticeship that make up the conceptual framework for the current study – modeling, scaffolding, and coaching – are discussed in more detail.

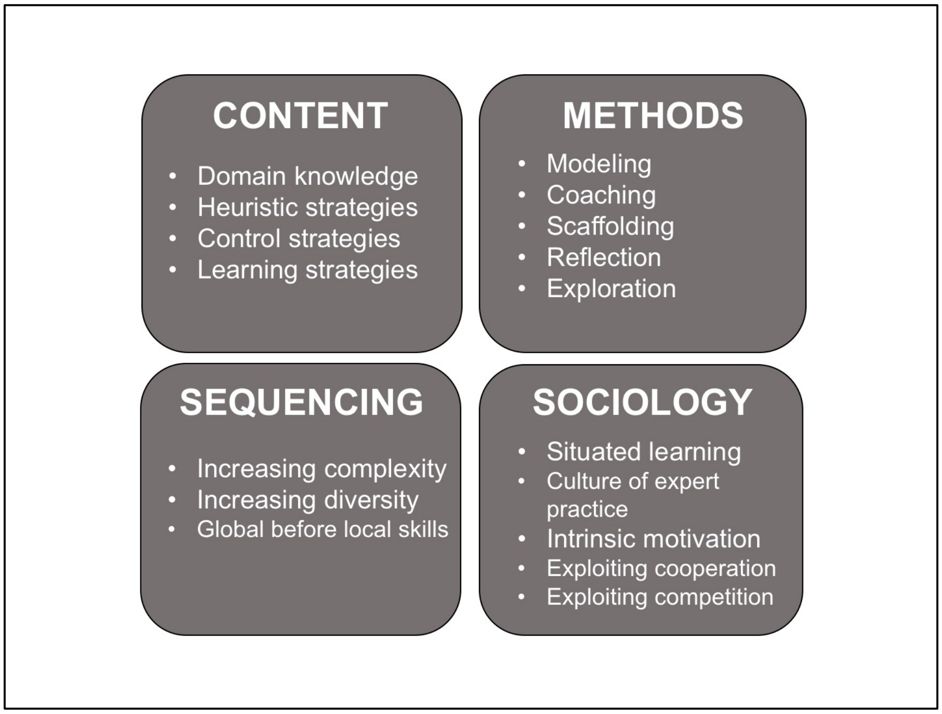


Figure 2.1. The cognitive apprenticeship framework domains and characteristics.

Content

This domain is of particular relevance to incorporating a cognitive apprenticeship approach into professional development. Collins et al. (1989) define content as the "knowledge required for expertise" (p. 477). In the context of designing and implementing professional development for teachers, this implies that decisions about content, or what teachers should be taught, should conform to the goal of developing their expertise. This calls for a narrow view of developing teachers' instructional practice and expertise; in other words, professional development for teachers should focus as specifically as possible on essential evidence-based practices and the specific steps or skills inherent in them. As the current literature base reveals, this is not a very commonly researched approach to teacher professional development. In the current study, the CAP-PD system is already deeply rooted in this part of the cognitive apprenticeship framework. The instruction and coaching provided in the CAP-PD system is intentionally narrowly focused on specific evidence-based practices and the essential implementation steps of each practice.

Domain knowledge. Domain knowledge refers to the factual knowledge and knowledge of procedures. Domain knowledge is taught and learned in such a way that it can be called upon when needed, but is generally isolated or "inert" from other skills and knowledge (p. 477). Other researchers have referred to this construct as *declarative* knowledge (Alexander, Schallert, & Hare, 1991). In a professional development context, simply defining or describing an evidence-based practice and then expecting teachers to know what to do with it and how and when to implement it in their classrooms would be insufficient – and yet, this is precisely the approach of many professional development programs that are widely used in schools and school districts (Ball & Cohen, 1999; Darling-Hammond, Hyler, & Gardner, 2017; Desimone & Garet, 2015; Hill, Beisiegel, & Jacob, 2013).

Heuristic strategies. Collins et al. (1989) refer to these strategies as "tricks of the trade," or techniques for completing tasks. These are almost always learned through experience and over time, and would not be known by outsiders to the context (i.e., profession, expertise, or trade). In a professional development context, heuristic strategies can be viewed as the ways in which teachers approach and make decisions, sometimes tacitly, about implementing the new evidence-based practices they are learning. For example, a teacher who has just learned about using student-friendly definitions (an evidence-based practice in vocabulary instruction) uses heuristic strategies to decide how and when to implement the practice and its essential steps with her students.

Control strategies. Control strategies are ways in which the learner manages the problem-solving process, i.e., selecting which procedures to use. This is similar to the construct of *conditional* knowledge as described by Alexander et al. (1991). Collins et al. (1989) further explain control strategies as incorporating the process, inherent to the learner, of monitoring, diagnosing, and remediating their approach to a problem. For teachers, this is a process of monitoring whether students are learning (or otherwise meeting a lesson objective) from the practice the teacher has chosen to implement, diagnosing or analyzing any obstacles that students are encountering, and remediating those obstacles or problems by amending, or expanding the selected instructional practice. In a professional development context, especially in this study, the researcher/coach employs these control strategies to make similar decisions about the effectiveness of the coaching.

Learning strategies. The final characteristic of the content domain is learning strategies. These are defined as “strategies for learning any of the other kinds of content” (p. 479). Teachers often talk about fostering these kinds of strategies with their K-12 students (e.g., having students select an independent reading book that is neither too easy nor too challenging). This approach assumes some autonomy and implicit understanding on the part of the learner, however, and students with disabilities may or may not have that autonomy or understanding, especially in mathematics. In Collins et al.'s (1989) original framework, students' own learning strategies are meant to make their content learning more efficient; as applied to professional development, these are strategies the teacher selects and applies in order to make that learning more efficient, such as a strategy for note-taking or planning a paragraph. With students with disabilities making up a significant portion of their rosters, general education teachers need a broader range of such strategies. Carefully planned professional development and coaching, such as the vocabulary instructional practices this study will use, can address this.

Methods

The teaching methods domain of the cognitive apprenticeship framework is also a cornerstone of the CAP-PD system. The modeling, coaching, and scaffolding characteristics in this domain are the heart of the CAP-PD approach. In the CAP-PD system, modeling is provided in the instructional modeling videos (CAP-TVs), scaffolding is provided in the materials to support classroom implementation (CAP-TS), and coaching is provided in the performance feedback from observation. Each of these components of the CAP-PD system is discussed in fuller detail later in this chapter.

Modeling. In the cognitive apprenticeship framework, as elsewhere, modeling refers to an expert carrying out an actual, meaningful task that students can observe and internalize as a concept. This is one way in which experts/models make their thinking visible to the learner, which according to Collins et al. (1989) is a necessary step for the learner to see. Collins et al. (1989) argue that many of the key components of a skill are hidden, and therefore are impossible to master without a model to make those hidden components visible to the learner. In a K-12 context, a teacher can model a skill using a think-aloud procedure. In a professional development context, teachers learning from other teachers is a powerful learning tool, as several recent empirical studies have found (e.g., Borko et al., 2008; Darling-Hammond et al., 2017; Seidel, Stürmer, Blomberg, Kobarg, & Schwindt, 2011). Seeing a model of an evidence-based practice being used is key to teachers' moving beyond knowing what the practice "is," (i.e., declarative or domain knowledge) to knowing how and when it is most effective. In the CAP-PD system (and the current study) specifically, modeling is a component of every instructional video that teachers watch, and most of these videos include a variety of different models and classroom contexts (i.e., tiers of instructional intensity, grade levels, etc) in order to develop teachers' understanding of the practice even further.

Coaching. In the coaching step, the learners carry out the task that was modeled to them, and receive hints, reminders, and other feedback from the expert. This is a very common and widely supported component of preservice teacher preparation and student teaching experiences (Darling-Hammond, 2014; Maheady, Smith, & Jabot, 2014), though it is far less common as a component of professional development programs, despite empirical evidence that it is effective for both improving teacher quality and student achievement outcomes (e.g., Kraft & Blazar, 2017). In many ways, coaching is a hallmark of the cognitive apprenticeship framework, and by extension it is the driving force of the CAP-PD system explored in the current study.

By conducting classroom observations and examining the observational data, the researcher/coach makes diagnostic decisions about what information or support the teacher needs in order to implement the targeted practice effectively. For example, if a math teacher is implementing student-friendly definitions in a vocabulary lesson on measures of central tendency, an observer/coach (in the case of the current study, the researcher/coach) can observe the lesson and collect data on whether she implemented the steps (i.e., implementation markers) of the practice accurately and effectively; for this practice, the steps or markers would include: (a) cueing the definition for students, (b) using clear, simple language, (c) reviewing key background knowledge, (d) clarifying potential misconceptions, (e) repeating essential information, (f) using anchor images, and (g) monitoring student understanding of the definition. In a typical coaching scenario, if the teacher included all of those markers except for monitoring understanding, the observer/coach would likely note its absence and would suggest ways to include it in future lessons.

In the above scenario, the coach's advice could be enhanced by providing the teacher with performance feedback on his lesson. Performance feedback is defined as data and other information collected via observation and provided to the teacher as a means of strengthening his/her practice (Duchaine, Jolivette, & Fredrick, 2011; Sweigart et al., 2015). In this case, performance feedback could show the teacher and coach that he did not ask any questions to monitor students' understanding when he gave a student-friendly definition for the term *median*, based on the number of opportunities to respond (OTRs; in this example, zero OTRs) the teacher used.

In the CAP-PD system and in the current study, these observations and data are collected using the Classroom Teaching (CT) Scan instrument (Kennedy, Rodgers et al., 2017). Figure 2.2 provides an illustration of the example scenario. The research base of the CT Scan and the performance feedback it provides are discussed in more detail in a later section of this chapter.

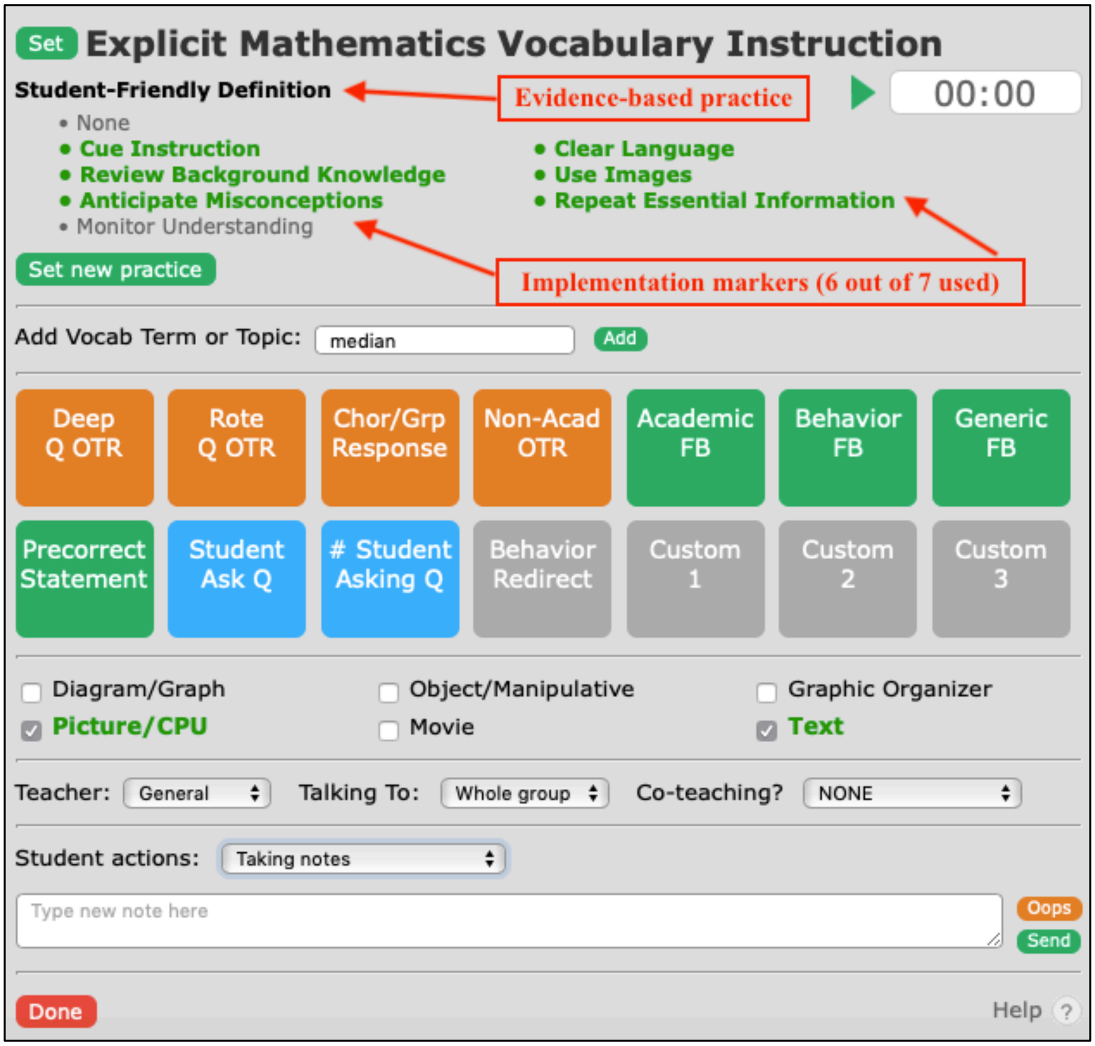


Figure 2.2. Example of CT Scan observer interface. In this example, the user is observing the teacher's implementation of using a student-friendly definition.

Scaffolding. In a K-12 context, scaffolding refers to the ways in which the teacher determines what kind of supports are needed by individual students, such as graphic organizers or anchor charts (Collins et al., 1989). These scaffolds are eventually faded until students can carry out the task or demonstrate mastery independently. In a PD context, tools or sample materials that guide the teacher's implementation of evidence-based practices are considered scaffolds. Providing teachers with a variety of models and materials to support implementation is also a kind of scaffolding in this framework. In the CAP-PD system, teachers receive PowerPoint slide sets to accompany their lessons. These pre-made slides include many of the implementation steps (i.e., slides that contain visual instructional cues, or that contain comprehension monitoring questions) that make up evidence-based practices such as explicit vocabulary instruction.

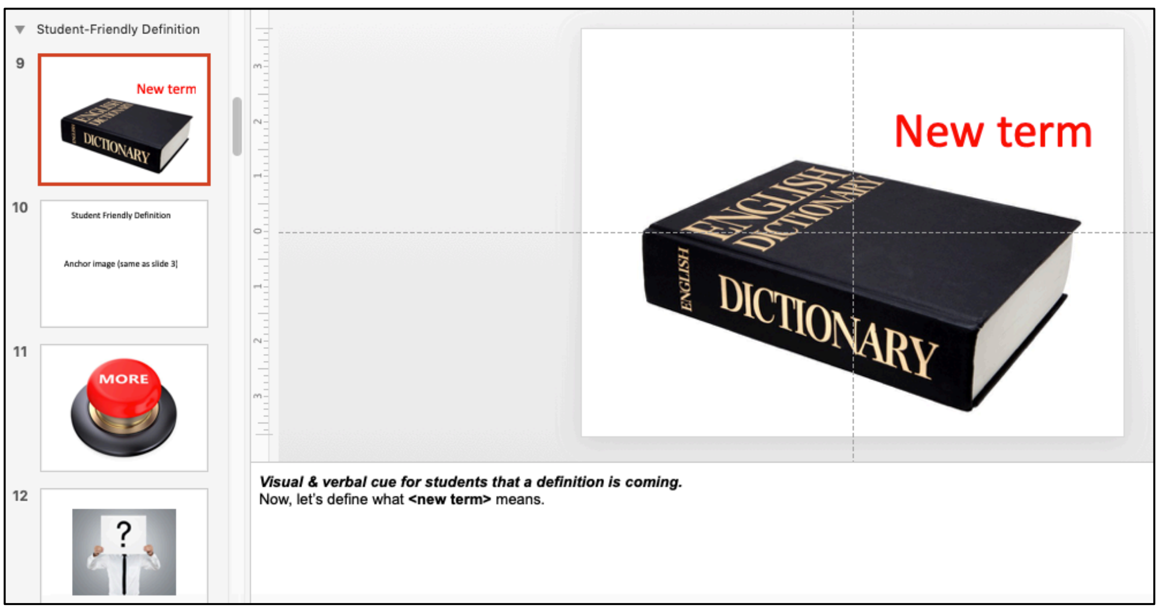


Figure 2.3. Sample CAP-TS slide for cueing instruction using a student-friendly definition. This figure shows editable slides with anchor images already embedded, with notes to the teacher about what to insert/edit and why.

Reflection. This step requires the learner to observe and compare his/her process with that of other learners as well as with that of the expert. In both K-12 and professional development contexts, there is a wide variety of ways in which this reflection step can take place, from video self-reflection to written reflection or journaling or group discussion. In the CAP-PD system, the interactions between teacher and his/her feedback and coaching emails, as well as between teacher and researcher/coach, provide opportunities for reflection (for an example of this from the current study, see Appendix B). This is an under-researched area in both the CAP-PD and broader teacher professional development literature bases.

Exploration. Collins et al. (1989) describe this crucial step as "the natural culmination of the fading of supports" (p. 483). In a K-12 setting, the exploration stage allows the teacher to push learners to apply the process they have mastered to new contexts or problems, gradually releasing the responsibility of performing the task to from the teacher to the learner. In a professional development context, as the scaffolds from models, materials, and coaching on evidence-based practices are faded, teachers should be able to use the practices in a wider variety of instructional situations.

Sequencing

In K-12 classrooms with or without the cognitive apprenticeship approach, the order in which students are taught specific skills is important. For example, students learn computational skills in mathematics (i.e., addition, subtraction, multiplication, and division) before learning more complex skills such as pre-algebra, in which they would apply the computational skills and use the correct order of operations to solve an equation. However, the importance, if any, of sequencing is not well documented in the teacher professional development research base.

Increasing complexity. This characteristic requires some organization of task complexity on the teacher's (or, in a professional development context, the observer/coach's) part, to ensure that scaffolds are applied effectively. The CAP-PD system does not address this explicitly, but in the current study of sixth grade mathematics instruction, the teachers themselves will, presumably, be teaching math skills and concepts at increasing levels of complexity. The models, scaffolds, and coaching from CAP-PD will support teachers' integration of explicit vocabulary instruction alongside those increasingly complex skills and concepts.

Increasing diversity. Collins et al. (1989) note that this characteristic is important because as learners approach an expert level of a certain skill, it is important that the teacher introduce tasks that require additional skills and strategies, "so the student learns to distinguish the conditions under which they do (and do not) apply" (p. 485). In mathematics instruction in particular, this is relevant to explicit vocabulary instruction. Many terms used in mathematics are also used in other contexts and with other meanings (e.g., *volume*), and teaching students to understand these different usages is key. In a broader professional development context, increasing diversity can refer to providing a greater variety of models of evidence-based practices, such as across varying grade levels or instructional levels of intensity. The current study will address these issues by providing that variety in the CAP-TVs.

Global before local skills. The purpose of this component is to assign broader tasks that, when completed, give students a "conceptual map, before attending to the details of the terrain" (p. 485). A K-12 teacher might demonstrate an entire procedure, for example, before breaking that procedure down into smaller steps, thus giving students a target or purpose for learning those steps. In the CAP-PD system, this decomposition of steps is the approach of most of the CAP-TV videos the teachers will watch.

Sociology

A core tenet of the cognitive apprenticeship approach is that learning is context-dependent (Brown, Collins, & Duguid, 1989). The situation in which learning takes place is as responsible for the knowledge that is gained as any amount of instruction or strategy. Situated learning employs "real-world" problems, rather than abstract or isolated ones. With regard to professional development, Putnam & Borko (2000) argue that, from a situative perspective within a cognitive apprenticeship approach to professional development, "the learning of teachers is intertwined with their ongoing practice, making it likely that what they learn will indeed influence and support their teaching practice in meaningful ways" (p. 6).

Situated learning. Extending Collins et al. (1989)'s description to a practice-based PD context, teachers' learning experiences (i.e., professional development) should take place at schools, and to the extent possible, in actual classrooms. This situated learning approach draws on earlier, qualitative work by Borko and colleagues (1997). It includes having teachers observe their colleagues and participate in discussions around insights from the observed teacher about his/her planning and practice, and what the colleagues observed. Some studies have examined teacher professional development models that use a collective, collaborative, situated learning approach, and they generally have mixed results (e.g., Grossman, Wineburg, & Woolworth, 2001; Sherin & Han, 2002). However, using a situated learning approach and providing approximate real-world examples of a teacher implementing each evidence-based practice are key features of the CAP-TV videos that were used in the current study.

Culture of expert practice. Collins et al. (1989) describe this characteristic as the way in which learners work and communicate in ways that approximate the ways that experts work and communicate. In a K-12 context, it requires the teacher not only to provide the expert modeling of a process or task, but also to model how they think about, interpret, and use their expertise, and encourage learners to do the same. As noted above, a group or collaborative learning approach to professional development is not always effective, and the current study does not explicitly employ a group dynamic. In fact, one of the unique features of the CAP-PD system is that it is individualized to meet a teacher's specific instructional and professional needs. However, as Romig (2018) found, teachers were working in the same school and taught the same subject; as a result, they planned lessons together, and the CAP-PD materials (especially CAP-TS) became an important part of their collaborative planning.

Intrinsic motivation. This sociological characteristic allow learners to "attempt to carry out realistic tasks in the spirit and for the purposes that characterize adult expert practice" (p. 489). Motivating learners, whether they are students in a K-12 classroom or teachers in a PD program, is key to developing expertise. The CAP-PD system (especially in each of the CAP-TV videos) is clear and deliberate in stating what the goals of its instruction are and why teachers should use the practices and materials being taught to them in order to motivate them to use the system to its full advantage. Receiving personalized performance feedback and individual coaching on could also be a source of motivation for teachers. The CAP-PD system is iterative and supportive, rather than evaluative; in other words, it focuses on implementation of observable, actionable steps and practices for improving instruction, rather than focusing on a rating or scoring system for the teacher's performance that could diminish a teacher's intrinsic motivation.

Exploiting cooperation. According to Collins et al. (1989), "learning through cooperative problem solving is both a powerful motivator and a powerful mechanism for extending learning resources" (p. 489). In classroom cooperative groups, students can take on different roles and carry out different tasks that provide additional scaffolding to their learning. In typical professional development programs, teachers often work in groups arranged by subject area or grade level. As CAP-PD is an personalized approach to teacher PD, there is not a cooperative component in its framework.

Exploiting competition. This characteristic actually refers to comparing processes of carrying out a given task, rather than competition or comparisons between individuals. Collins et al. (1989) note specifically that the comparisons should only be between learners' processes, and not between the product of those processes. One approach to coaching that is available to the CAP-PD observer/coach is to compare the teacher's current performance to his/her past performance, as the CT Scan tool is capable of storing and organizing a large database of observations.

Current research base for cognitive apprenticeship. The cognitive apprenticeship framework is complex, and to implement it fully would require a fundamental shift in how both researchers and practitioners think about the processes of teaching and learning. This has made the framework difficult to implement in the K-12 settings for which it was intended; of course, the same difficulties will apply to its potential use in teacher learning and professional development settings. To date, no studies in the documented literature have studied an application of the full cognitive apprenticeship framework in teacher professional development.

While some research on teacher learning (including preservice teacher preparation as well as in-service professional development) has found success in using some of the framework's methods and components, the body of that research is quite small. Many of these few studies lack either the methodological rigor or a careful focus on evaluating teacher quality to allow causal claims about the impact of a cognitive apprenticeship approach to PD on improving teacher practice. For example, many studies in this body of work rely on teacher self-reporting or qualitative coding of teachers' discussions or interviews (as opposed to lessons) to provide information about the effectiveness of the professional development, rather than collecting direct observational data (e.g., Borko et al., 2008; Peters-Burton, Merz, Ramirez, & Saroughi, 2015). Borko et al. (2008) tested an extended (multiple sessions), iterative professional development program called the

Problem Solving Cycle (see also Borko et al., 2011; and Koellner et al., 2007) that included opportunities for math teachers to view video of themselves teaching as well as others' teaching. The teachers' responses were positive, as they felt that they were able to get insights and feedback on their teaching, and to see other practices and techniques modeled by their colleagues (Borko et al., 2008). The researchers used video recordings of teachers' lessons to create "artifacts of practice" (Borko et al., 2008, p. 418) that provide the models of teacher learning, situated in teachers' own physical and social contexts, as called for by the cognitive apprenticeship framework and according to the researchers' situative perspective approach.

Peters-Burton et al. (2015) applied cognitive apprenticeship as their professional development model. The science teachers in their study participated in an extended (year-long) professional development to support their use of inquiry-based teaching methods. The researchers collected both quantitative and qualitative data, but did not record or collect data through direct observation of teachers' implementation; rather, the research questions were aimed at whether the professional development improved teachers' self-efficacy with using inquiry-based teaching methods in their science classes. While both this study and Borko et al. (2008) directly applied elements of the cognitive apprenticeship framework to their research in professional development, both have a number of limitations. First, neither study used direct observation or fidelity of implementation as the primary measures of the impact of the professional development on actual teacher practice. Second, neither study employed methods or design that would allow causal claims about the effectiveness of their respective professional development programs. In both studies, all teachers received the professional development and as a result there was no control group or staggered implementation by which the researchers could compare their results. In order to address these issues of methodological quality in studies of the cognitive apprenticeship approach to teacher professional development, the current study employed a single-case research design, and the constructs of interest – impact of performance feedback in PD for improving explicit vocabulary instruction in mathematics – were measured through direct observation.

Some of the most commonly studied components of the cognitive apprenticeship framework in the teacher professional development research base are its teaching methods, especially modeling, coaching, and scaffolding. While it is rare that these methods are applied in empirical studies explicitly aligned with the cognitive framework, each has a strong empirical base as an approach to teacher learning. The current literature points to each of these methods as effective approaches for designing and implementing PD (e.g., Darling-Hammond et al., 2017; Desimone & Garet, 2015; Kraft & Blazar, 2017; Yoon et al., 2007). The CAP-PD system was designed to employ these methods, as well as several other characteristics from the cognitive apprenticeship framework to improve teacher professional development as well as instructional practice. The following section describes the importance of studying the performance feedback and coaching component of CAP-PD in greater detail.

Performance Feedback in Professional Development Contexts.

Receiving feedback on one's performance on a task is one of the key components of effective learning (Collins et al., 1989; Brown et al., 1991). Receiving feedback is also a crucial component of adult learning and is common in most professional fields (Cannon & Witherspoon, 2005). Performance feedback refers specifically to data and other information collected via observation and provided to the teacher as a means of strengthening his/her practice (Duchaine et al., 2011; Sweigart et al., 2015). Performance feedback provides information about behaviors or instructional components that the teacher is executing well, and information about behaviors or instructional components that they are not executing as desired or planned. Finally, performance feedback supports professional learning; as such, it should be a crucial part of professional development, and has thus become an area of interest in recent research on teacher professional development.

A meta-analysis by Solomon and colleagues (2012) evaluated a body of single-case intervention studies on performance feedback delivered to in-service teachers, ultimately analyzing 36 published studies. Their analysis determined that performance feedback is effective for teachers at all grade levels and for both general and special education teachers, and that feedback provided within a day at most is more effective than delayed feedback. In another meta-analysis, Fallon and colleagues (2015) applied the What Works Clearinghouse (WWC; Kratochwill et al., 2013) standards for single-case study design and evidence to a body of 47 studies of performance feedback delivered to in-service and preservice teachers. While their analysis found somewhat mixed results for studies that met WWC design standards, 29 of the 47 studies demonstrated at least moderate evidence that performance feedback was effective, according to WWC evidence standards.

In terms of effective dosage, research on performance feedback with in-service teachers has shown that receiving the feedback as immediately as possible (within 24 hours is ideal in most cases) has the most effect on changing teacher practice and behavior (Scheeler, Congdon, & Stansbery, 2010; Solomon et al., 2012; Sutherland, Wehby, & Copeland, 2000; Sweigart et al., 2015). Research has also shown that, in general, providing follow-up (i.e., coaching) to a PD session is more effective than not providing follow-up (e.g., Darling-Hammond et al., 2017; Guskey & Yoon, 2009), but this is still an area where additional research is needed.

Approaches to Providing Performance Feedback to In-service Teachers

A number of approaches to how performance feedback is provided to teachers have been explored in the research literature. Two of the most common and most effective forms of performance feedback are immediate or live feedback provided by "bug-in-ear" (BIE) technology, and visual performance feedback. Studies of performance feedback delivered via BIE have found positive and encouraging results for coaching both preservice and in-service teachers, including para-educators. Many of these studies of the BIE technology have been conducted by Scheeler and her colleagues (Scheeler et al., 2010; Scheeler, McAfee, Ruhl, & Lee, 2006; Scheeler, McKinnon, &

Stout, 2012; Scheeler, Morano, & Lee, 2018; Scheeler & Lee, 2002). These studies demonstrated improvement in a variety of teaching behaviors and practices as a result of receiving performance feedback via BIE. An additional study of a BIE intervention with preservice special education teachers was conducted by Rock and her colleagues (2009). That study also found that receiving immediate performance feedback via BIE improved teachers' practice (in that study, rates of praise statements was the practice of interest) and consequently student engagement and on-task behavior.

Another approach to providing performance feedback is the use of visuals and graphics to present data collected during a lesson observation. This data can consist of rates and types of OTRs or praise statements, frequency and duration of the use of specific practices, fidelity of implementation of specific practices, and overall time use in the lesson, and can be presented as pie graphs, line graphs or timelines, percentages, rates, and frequency counts. Reinke, Lewis-Palmer, and Martin (2007) provided teachers with graphs of their use of behavior-specific praise statements daily, and found that teachers increased the number of these statements after receiving the visual performance feedback. Myers, Simonsen, & Sugai (2011) provided teachers in a professional development intervention with data collected from observation of their lessons, including rates of praise statements, ratio of positive to negative interactions, and intervals of students' on- and off-task behaviors. They found that receiving this data along with one-on-one coaching improved teachers' practice as demonstrated by increased number of praise statements, improved ratio of positive to negative interactions, and more intervals of on-task behaviors (Myers et al., 2011). However, their professional development intervention used a response to intervention (RtI) approach, and so was differentiated among tiers according to teachers' response to the feedback and coaching. As a result, the timing and dosage of the performance feedback varied at different tiers of intensity (i.e., weekly at Tier 2, daily at Tier 3).

Myers et al. (2011) also note that simply "exposing" the teachers to information about the practices they were targeting for improvement was not sufficient to improve their practice (p. 51). The CAP-PD system takes a similar approach, tying the practices on which instruction is provided to specific menu options in the CT Scan observation instrument, such that the performance feedback that teachers receive is directly connected to what they learn in the CAP-TVs and what they are using in their instruction with the CAP-TS.

Using CT Scan to Provide Performance Feedback in the CAP-PD System

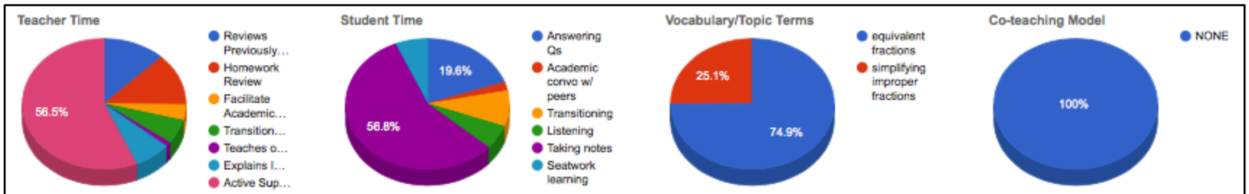
The existing literature on the use of the CT Scan is predominantly set in professional development settings (e.g., Kennedy, Rodgers et al., 2017; Kennedy, Rodgers et al., 2018). The CT Scan and the performance feedback it produces are unique features of the CAP-PD system, as most teacher professional development does not provide teachers with specific, individualized performance feedback on their implementation of the content of the professional development (Kennedy, Rodgers et al., 2017). The CT Scan, developed by Kennedy and colleagues at the University of Virginia, is a web-based, low-inference observation tool that records teachers' moves in real-time, and allows the observer to record descriptive data about various aspects of the lesson, such as duration, frequency, and implementation fidelity (Kennedy, Rodgers et al., 2018; see also www.thectscan.com).

The CT Scan uses an explicit instruction framework, as its original purpose was to observe and record data about general education teachers' instruction of students with disabilities. This tool was conceived and developed out of a need for greater specificity in teacher observation. As a result, the CT Scan is primarily a descriptive tool, and does not produce evaluative information such as a rating or a score. The implementation markers (described below) provide a measure of instructional quality, but the goal of those markers is to provide a basis for instructional coaching, rather than to evaluate overall performance or quality. Additionally, the CT Scan computes a measure called the Quality Vocabulary Index, or QVI, which expresses the teacher's implementation fidelity of vocabulary practices as a function of the time spent teaching vocabulary. Like the implementation markers, which are expressed as a percentage, the QVI is not used to provide a "score" on the teacher's performance, but is primarily used for research purposes. Another unique feature of the CT Scan is that it stores observation data to allow coaches and teachers to view how a teacher's performance changes over time. Finally, the CT Scan can be customized and used in any content area or instructional setting, such as writing instruction, mathematics instruction, or vocabulary instruction.

Using the CT Scan. The observer first selects a category of instruction, such as vocabulary instruction. This refers to the instructional topic the teacher is using at the time; the CT Scan also includes menu options for transitional time and non-specific general content instruction (such as reviewing homework, etc). Once an instructional category is selected, as the lesson moves forward, the observer next selects from a list which evidence-based practice the teacher is using. For example, in vocabulary instruction, this would include practices such as using a student-friendly definition (see Figure 2.2). Meanwhile, the CT Scan records the duration of the teachers' use of each selected practice, which later appears as a section of one of the graphs in the performance feedback output, and provides a start and stop time for each practice on the lesson timeline piece of the performance feedback. Each time the teacher implements a new practice, the observer selects it from the list on CT Scan. For each practice the observer selects, there are a number of implementation markers (IMs) listed on the user screen (see Figure 2.2). These IMs serve as a fidelity checklist for the practice; in the performance feedback and coaching pieces, they inform the teacher and coach about which steps the teacher used and which he/she did not use. The list of IMs for each evidence-based practice corresponds to the steps that are taught in the CAP-TV for that practice; as a result, if a teacher misses one or more IMs, the coach can refer the teacher back to a specific part of the CAP-TV where the missed steps are discussed.

Regardless of which category or practice is selected, the observer can also record frequency counts of various types of opportunities to respond (OTRs; deep, rote, choral, behavioral), feedback statements (generic, academic, behavioral), student questions, and behavior redirect statements. These counts also appear in the performance feedback, both as raw counts and as rates (e.g., OTRs per minute), and also appear in the lesson timeline time-stamped to when they occurred in the lesson. This provides even deeper detail than the rates or frequency counts, as it allows the teacher and coach to see trends in how and when the teacher is providing these statements or questions (i.e., if they are all at one point of the lesson and hardly anywhere else). Finally, the observer can also record what the students are doing and their level of engagement during the lesson. Each of these elements is reported in the visual performance feedback that teachers receive (see Figures 2.4 and 2.5), and form the basis of the coaching email (Figure 2.6) that the observer sends to the teacher along with the visuals.

CT Scan and performance feedback in the current study. While the CT Scan and the performance feedback it provides have been used in every CAP-PD study to date, its specific impact on teachers' improvement has not yet been examined. The current literature on performance feedback shows that it can improve various aspects of teachers' practice and implementation fidelity. The current study seeks to add to that literature base by demonstrating the impact of the performance feedback and coaching as part of the CAP-PD system.



Setting	Value
Session Id	649
Observer	Kate
Observer email	kpeeples@virginia.edu
Teacher Id	B2
Grade	6
Subject	Math
Date	2018-09-11
Block	3
Setting	Inclusive
Video	Live
Reminder	2
Research	Baseline
Template	menus.txt

Event	(#)	#/M
Deep Q OTR	(0)	0.0
Rote Q OTR	(21)	0.3
Choral Response	(13)	0.2
Social Behavioral OTR	(0)	0.0
Academic Specific FB	(3)	0.0
Behavior Specific FB	(2)	0.0
Generic FB Statement	(22)	0.3
Precorrect Statement	(0)	0.0
Student	(3)	0.0
Ask Q	(2)	0.0
Num Student Ask Q	(12)	0.2
Behavior Redirect	(0)	0.0
Custom 1	(0)	0.0

Vocabulary Practices	Time	Markers used	Total
Facilitate Academic Discussion	03:00	75%	4%
Total vocabulary practices used	1		
Total time when all markers were used	00:00		
Quality Vocabulary Index	0.07		

Figure 2.4. Example of data-based performance feedback output from CT Scan. In this display, the teacher can see what proportions of the observed lesson were devoted to various practices and activities (i.e., pie graphs) as well as specific counts and time use.

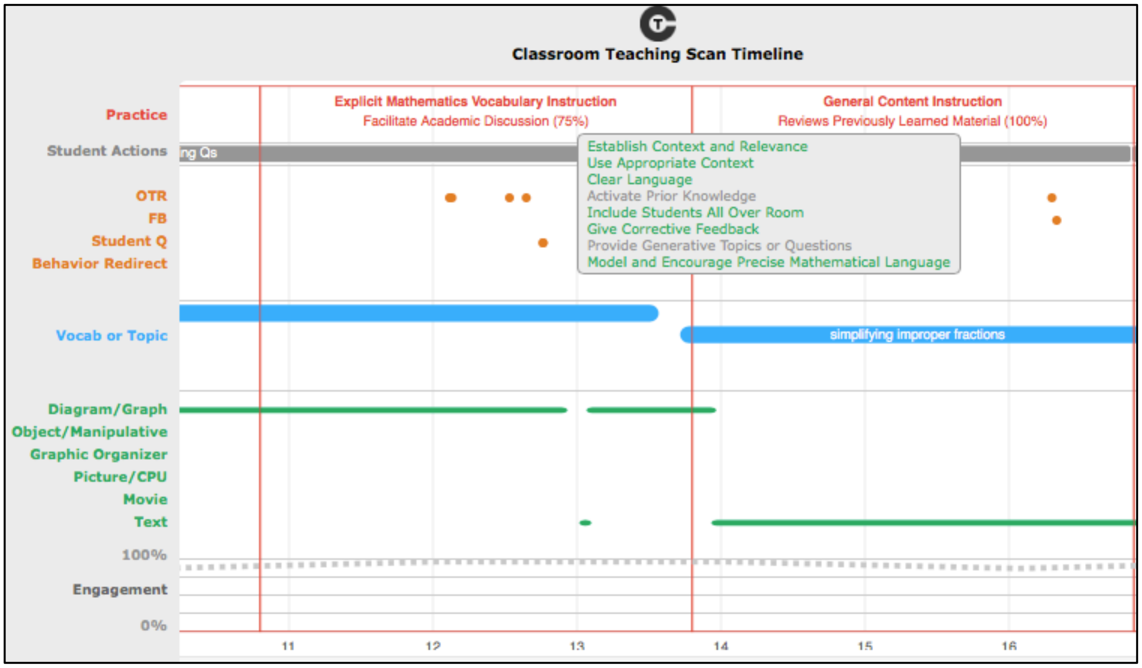


Figure 2.5. Example of timeline performance feedback output from CT Scan. This display presents the same data in a different format. Here, the teacher can see what practices were used in the observed lesson, what IMs were used with each practice, when and for how long each term was taught (e.g., improper fractions), and when other behaviors occurred, such as opportunities to respond (OTRs, represented by the orange dots).

Practices	What You Did	Comments
Vocabulary Instruction	**Student Friendly Definition** *Provided for* mean median mode	Providing students with student-friendly definitions is probably the most important and effective practice at your disposal for supporting student learning. Today you were reviewing measures of central tendency. You provided student friendly definitions for 3 terms and about 7 minutes in this lesson. I like how you were asking lots of questions to confirm student learning, and using the slides with images to help students develop cognitive anchors for each term.
	Example of Term Meaning *Provided for* mean median mode	Providing students with rich examples of what terms mean is also a great choice and powerful practice. In this lesson you included visual representations with your student-friendly definitions of mean, median, and mode. The representations were clear and mathematically accurate.
Opportunities to Respond (OTRs)	18 questions 14 rote 4 choral 0 deep	In this lesson, you asked a total of 15 questions. 14 of the questions were rote, for a rate of .3 questions per minute. You also provided 4 choral response questions.
Feedback Statements (FB)	15 feedback statements 15 generic 0 academic 0 behavioral	During the observation, you gave 15 total feedback statements to students, which were all generic feedback statements such as "Good job," or repeating back a student's response.

Coach's Comments	
The Big Picture	Great job using slides with images and text to review and explicitly teach the meaning of these terms! I liked seeing the combination of student-friendly definitions and then using examples to teach these terms. Using combinations of evidence-based practices is so important to helping students understand complex terms. Using the visuals on the slides provided further support for your students. You spent about 10 minutes (about 18% of your total class time) on explicit vocabulary instruction in class today. The rest of class was spent on students completing the test review study guide.
One Big Thing	You continue to do a good job of merging the vocabulary teaching slides template with your approach and style. To take it a step further, check out Video 6 on having a discussion, and see if you can find ways to incorporate discussions (including deep questions and feedback, which are featured at around the 3-6 minute marks) into future review lessons as well as when you teach new terms.

Figure 2.6. Example of coaching email based on CT Scan performance feedback, from a previous CAP-PD study. Using the data recorded and presented in the performance feedback (Figures 2.4 and 2.5), the coach writes up a summary of the lesson in terms of targeted practices such as explicit vocabulary instruction, and emails it to the teacher along with a link to view the graphs and timeline. An updated version is included in Appendix B.

Research Questions

The current study built upon each of the research bases discussed in this chapter. First and foremost, this study expands on the research literature on the use of performance feedback and instructional coaching in teacher professional development. More specifically, this study contributes to the existing CAP-PD research base by examining the specific role that the performance feedback and coaching component of the system has on teachers' improved use of evidence-based practices. There is already a solid research base supporting the use of performance feedback for teachers, and other components of the CAP-PD system, namely CAP-TVs and CAP-TS, have already demonstrated their effectiveness with regard to improving teachers' use of evidence-based practices. On the other hand, the CT Scan tool provides a unique kind of performance feedback; as such, its role in the CAP-PD system should be investigated.

By investigating the effects of CAP-PD on instructional quality, this study expands the research on the use of a cognitive apprenticeship approach to teacher professional development. Additionally, this study contributes to the very limited research literature on evidence-based practices for explicit vocabulary instruction in middle school mathematics. The teachers participating in the study were current middle school mathematics teachers whose classes include students with disabilities. The instruction that these teachers will receive is rooted both in the literature on explicit vocabulary instruction and in current mathematics teaching practices and standards (i.e., NCTM, 2000).

The current study focused on the performance feedback and coaching component of the CAP-PD system, and its effect on the quality of middle school mathematics teachers' use of evidence-based practices for explicit vocabulary instruction. The primary research question that will drive this study is:

(1) What is the effect of CAP-TVs and CAP-TS with and without performance feedback and coaching on mathematics teachers' use of evidence-based practices for effective vocabulary instruction?

Chapter 3: Methodology

In order to investigate the effects of performance feedback and coaching as a PD intervention to improve the vocabulary instruction of middle school mathematics teachers, this study used a single-case multiple baseline across participants design, and implemented the performance feedback and coaching intervention as alternating treatments. Interventions in single case research designs investigate measure the effects of making deliberate changes to the subject's environment (i.e., manipulations of an independent variable). These measures represent the dependent variables of the investigation, and data are collected repeatedly and systematically from each subject throughout the study. The goal of single-case research designs is to show a functional relationship between the independent and dependent variables (Horner et al., 2005; Kazdin, 2011).

Broadly, multiple baseline designs can answer causal research questions by evaluating the effects of treatment (interventions) on a dependent measure of performance replicated across individuals. More specifically, in multiple baseline designs across individuals, the effect of the intervention is determined by the demonstration of the desired change in performance when the intervention is introduced to each individual and only at that point (Kazdin, 2011, p. 150). In the current study, the manipulated (independent) variable was the performance feedback and coaching "treatment." The primary dependent variable is the quality of the teachers' explicit mathematics vocabulary instruction as measured by the QVI statistic.

Participants

Prior to any recruitment activities, I gained approval for this study from the university's Institutional Review Board (IRB). As part of that process I also gained permission from the school division's administrative leadership as well as the principal at the participating school. In this study, individual teachers were the units of analysis; therefore, I only collected demographic information about the teachers and not about the students. Each teacher was asked to provide the number of students in the observed class who have current Individualized Education Plans (IEPs) or Section 504 Plans. No individually identifiable student information or data was collected. Each teacher was provided with letters to send to parents of students in the observed classes, informing them of the purpose and timeline of the study, as well as stating that no identifying student information will be collected.

Teachers

Participants in this study were sixth grade mathematics teachers at one middle school in a local school division. With the support and assistance of the school's instructional leaders, four teachers (the entire sixth grade mathematics team) was successfully recruited. Selection criteria were as follows: (a) fully certified to teach the mathematics course to which they are assigned; (b) currently teaching general education sixth grade mathematics courses; (c) have a number of students with disabilities (i.e., current Individualized Education Plans or Section 504 Plans) in at least one of these courses; and (d) can readily agree and consent to participate for the duration of the study and complete all related measures and activities (e.g., completing participant demographic survey and social validity survey). Each teacher was informed of these measures and activities before they were asked to consent to participate, and each teacher signed a participant consent form approved by the university IRB. Each teacher was assigned a numeric research code, and all information provided and data collected from each teacher throughout the study was associated with that code. Additionally, pseudonyms have been assigned to each teacher in this dissertation which will also be used in any subsequent manuscripts for publication. As compensation for their participation, each teacher received $250 at the conclusion of the study. This funding was provided from two generous awards given by the Curry School of Education and Human Development: The John and Florence May Fellowship, and an Innovative, Developmental, Exploratory Award (IDEA) grant.

Demographic Information by Teacher

Each of the teachers responded to a brief demographic survey prior to the start of observations. Below I summarize the results of that survey. The full demographic survey is included in Appendix A.

Ada. Ada is 28 years old and identifies as a white female. During the study, she was in her third year of teaching overall, and her first at Mae Jemison Middle. Her licensure area is Elementary (PK-6), which includes the Mathematics 6 course she was teaching at Jemison. In addition to teaching sixth grade mathematics, she has also taught first, third, and fifth grade mathematics. Her undergraduate degree is in elementary education, followed by a master's degree in curriculum and instruction with a mathematics focus.

Florence. Florence is 38 years old and identifies as a female of Native American and European descent. During the study, she was in her 16th year of teaching overall, and her 13th at Mae Jemison Middle. Her licensure area is PK-8 Mathematics, which includes the Mathematics 6 course she was teaching at Jemison. In addition to teaching sixth grade mathematics, she has also taught seventh and eighth grade mathematics. Her undergraduate degree is a bachelor of science in liberal studies, followed by a master of education degree. Florence also holds a mathematics specialist licensure endorsement.

Emilie. Emilie is 32 years old and identifies as a white and Latina female. During the study, she was in her fourth year of teaching overall, all of which have been at Mae Jemison Middle. Her licensure area is K-8 Mathematics, which includes the Mathematics 6 course she was teaching at Jemison. She holds an undergraduate degree as well as a master of arts degree in teaching.

Sophie. Sophie is 28 years old and identifies as a white female. During the study, she was in her first year of teaching. As a "career switcher," her middle grade (6-8) license was provisional at the time of the study. In addition to teaching sixth grade mathematics, she was also teaching seventh and eighth grade mathematics remediation during the study. Sophie's undergraduate degree is in criminal justice.

Dependent Variables

This study relies on precise observational data in order to determine the effect of the performance feedback and coaching component of the CAP-PD system on mathematics teachers' explicit vocabulary instruction. To accomplish this, I used the CT Scan, developed by Kennedy and colleagues (Kennedy, Rodgers et al., 2017). In order to provide a richer context for the participants, the experiment itself, and its findings, I also collected demographic information from the participating teachers at the beginning of the study and a social validity survey at the end of the study.

Observation Instrument

The CT Scan was used to investigate teacher practice and the quality of specific aspects of vocabulary instruction. The CT scan is a descriptive, low-inference observational instrument that uses an explicit instruction framework to record data about instructional practice. It records data in real time, as the observer is able to record practices and implementation steps of practices as the teacher is using them. For each evidence-based vocabulary practice in the CT Scan, there is a list of steps that need to be included in order to implement the practice with fidelity. These are called implementation markers, or IMs. All vocabulary practices and IMs in the CT Scan are based on relevant research and expert review. At the end of the observation, the CT Scan can produce detailed performance feedback that shows the practices and implementation steps the teacher used and did not use, for how long, and for what proportion of the whole lesson. This feedback includes frequency counts, graphs, and a timeline view of the entire lesson. A full background and overview of how to use the CT Scan is provided in chapter 2. The instrument can be viewed online at www.classroomteachingscan.com/ctscan/, and screenshots of the user interface as well as the performance feedback and coaching can be seen in Figures 2.3 - 2.6).

Quality of vocabulary instruction. The CT Scan does not produce a score in the evaluative sense, but it does produce a statistic called the Quality Vocabulary Index, or QVI (Kennedy et al., 2018). The QVI is primarily intended as a research statistic, rather than as a measure used for feedback or coaching. The QVI is a weighted measure of implementation fidelity. Rather than simply using a percentage of implementation markers, the QVI takes duration of each evidence-based vocabulary practice into account. In its mathematical expression (below), x is the percentage of implementation markers for one instance of an evidence-based practice, y is the duration for which the teacher used that practice (measured in seconds), and z is the duration of the whole observation or lesson (Kennedy et al., 2018). In a vocabulary lesson, a teacher might use a variety of evidence-based practices or may use relatively few practices, but the observer can record each of these and the implementation fidelity markers for each instance using CT Scan. To create the total QVI score, the CT Scan sums all of these “weighted” fidelity measures.

$$QVI = \Sigma\left\{(x+1)\frac{y}{z}\right\}$$

In the current study, the QVI was the basis for making phase change decisions. I hypothesized that the teachers would likely not use many of the explicit mathematics vocabulary instructional practices or any of the implementation markers, so a stable, predictable trend in their QVI scores would be easily observable by visual analysis. I further hypothesized that QVI scores would increase across both the "treatment on" and "treatment off" periods within the intervention phase, but that QVI scores during "treatment on" period would exceed those in the "treatment off" periods. I expected that as teachers became more knowledgeable about the explicit mathematics vocabulary instructional practices and the implementation markers (IMs) for each practice, the quality of their vocabulary instruction (i.e., implementation fidelity) and the amount of instructional time they spent on it would increase, thus increasing the QVI scores.

Fidelity of implementation. While the QVI will be the variable used for making phase change decisions, the CT Scan can provide additional, detailed information on teachers' fidelity of implementation of the explicit mathematics vocabulary instruction. As noted above, this data is also included in the calculation of the QVI statistic. With the CT Scan, the researcher can examine which implementation markers are used, if some are used more frequently than others; additionally, this information can be used in the coaching component. For example, one of the evidence-based practices for explicit mathematics vocabulary instruction is using a student-friendly definition. There are six implementation markers in the CT Scan menu for this practice (see Table 2.1). If a teacher cued her instruction and used clear language according to the observer, but did not use any of the other markers, then she reached 33% on that instance of using a student-friendly definition. When expressed as a percentage, I calculated implementation fidelity simply as the number of IMs used divided by the number of IMs available, multiplied by 100. Like the QVI, this data was recorded and graphed throughout the study (see Figures 4.1 and 4.2), as a means of illustrating and examining changes in the teachers' instruction. The number of IMs used for each vocabulary practice is available on the CT Scan outputs that the teachers received, but was not directly reported to or discussed with the teachers in their performance feedback or coaching emails.

Time spent on explicit mathematics vocabulary instruction. In addition to implementation fidelity, the QVI statistic is also a function of the time the teacher spends using explicit vocabulary instructional practices. This data is also collected and recorded in the CT Scan and was graphed throughout the study as another means of examining changes in teachers' instruction (see Figure 4.3). In another expansion on previous CAP-PD studies, I included a simple line graph in each of the coaching emails that the teachers received with their CT Scan outputs, showing the total minutes that the teacher had spent on vocabulary instruction on each day of the study. This was then tied in to the recommendation provided to the teacher elsewhere in that document, especially with regard to how many minutes the teacher should try to spend on vocabulary in future lessons (e.g., "You spent about 3 minutes on vocabulary today, which was about 2% of the class time. Next time, try to go for 5-7 minutes."). Research on the vocabulary learning of students with disabilities has shown that devoting more instructional time for direct, explicit vocabulary instruction is essential in improving academic outcomes for these students (Bryant et al., 2003; Jitendra et al., 2004; Kennedy et al., 2015).

Additional data provided by CT Scan. The CT Scan also collects information based on frequency counts, including opportunities to respond (OTRs), feedback statements to students, and behavioral redirect statements. The instrument also provides the observer the opportunity to differentiate among different types of OTRs and feedback statements. These counts appear in the performance feedback as raw counts and as rates (calculated per minute of the lesson), and the distribution of these counts also appear on the lesson timeline. Finally, the observer using CT Scan can also record information about students' actions, time use, and engagement during the lesson. All of this information appears in the teacher's performance feedback.

Table 3.1

Data Provided by CT Scan in the Current Study

Data Type	Measure	Role(s) in Current Study
Quality of explicit mathematics vocabulary instruction	Quality Vocabulary Index QVI; calculated by CT Scan)	Primary decision variable in experiment
Duration of vocabulary instruction	Recorded by observer, aggregated by CT Scan	• Highlighted in intervention (performance feedback & coaching) • Presented in results to illustrate changes in teacher practice
Proportion of vocabulary instructional time to whole lesson	Calculated by CT Scan	• Highlighted in intervention (performance feedback & coaching) • Presented in results to illustrate changes in teacher practice
Fidelity of implementation	• Number and percentage of implementation markers used per practice (calculated by CT Scan) • Number and percentage of implementation markers used across all vocabulary practices (calculated by CT Scan)	• Shown in performance feedback & coaching, but not highlighted • Presented in results to illustrate changes in teacher practice

Technical specifications of CT Scan. The CT Scan has not yet been subjected to technical review as it is still a relatively new tool. It has been used recently as the primary observation measure in CAP-PD studies, and has received additional input from experts on its design and utility; as such, it is still evolving. In the studies that have used it as the primary measure, it has performed well. For example, Kennedy and colleagues (2017) used the CT Scan as the primary dependent measure in an initial study of the CAP-PD system, and reported 87% inter-observer reliability on class time spent on vocabulary instruction (p. 221). Romig (2018) also used the CT Scan as the primary dependent measure in another CAP-PD study, and reported 92% inter-observer reliability on class time spent on modeling in writing instruction.

Inter-Rater Reliability with the CT Scan

As the principal investigator, I have completed hundreds of hours of observation in the field using CT Scan, and have also done multiple observations and trainings for inter-observer reliability with other researchers using the CT Scan. However, in order to ensure the reliability of observations and data related to the primary dependent measure in this study, I included two additional observers to ensure inter-rater reliability (IRR). One observer was already trained and had also completed hundreds of hours of observation, and was also the lead developer of the instrument. I had established a high rate of inter-rater reliability on the CT Scan in previous studies with this observer (although IRR was calculated separately with this observer for the current study). The other observer was a doctoral student to whom I provided training and practice sessions using the CT Scan prior to and during study observations. That training consisted of a detailed review of the evidence-based practices and implementation markers for explicit mathematics vocabulary instruction. Then, this observer watched and coded previous videos (i.e., not from the current study) using the CT Scan until 80% inter-observer reliability was reached. Finally, after conducting two live observations for practice, the doctoral student's observations were included in IRR evaluations.

These two trained observers participated separately in live observations, also using the CT Scan. 15.8% (16 observations) of the total observations for each teacher were double coded by one of these two additional observers. Each of these two observers completed 8 observations in person with me. At the individual lesson level, inter-rater reliability was calculated as overall percent agreement on when vocabulary instruction occurred and did not occur. On this measure, the reliability for observations with Observer 1 was 94.8%, and 95.6% with Observer 2.

Inter-rater reliability for implementation fidelity of vocabulary instruction was calculated on all double coded observations that included vocabulary instruction (a total of 6 observations). I calculated percent agreement on the teachers' use of the implementation markers for all vocabulary practices in CT Scan (not just explicit mathematics vocabulary instructional practices). Cohen's kappa was also calculated for the reliability on each practice, by combining the numbers of agreements and disagreements for all the implementation markers (see Table 3.2).

Table 3.2

Inter-Rater Reliability for Implementation Fidelity of Vocabulary Instructional Practices.

Vocabulary Practice (overall agreement)	Cohen's kappa (SE)	Implementation Marker	Percent agreement
Asks student(s) to state definition (83.3%)	0.5 (0.23)	Cue instruction	83.3%
		Clear language	100%
		Require interpretation	66.7%
		Provide feedback	83.3%
Prompts student(s) to apply (94.4%)	N/A	Cue instruction	83.3%
		Clear language	100%
		Provide feedback	100%
Uses a student-friendly definition* (82.1%)	0.58 (0.17)	Cue instruction	75%
		Clear language	75%
		Review background knowledge	75%
		Use images	100%
		Anticipate misconceptions	100%
		Repeat essential information	75%
		Monitor understanding	100%
Uses examples*† (83.3)	N/A	Cue instruction	0%
		Clear language	100%
		Use images or visual representations	100%
		Make connections explicit	100%
		Anticipate misconceptions	100%
		Monitor understanding	100%
Facilitates academic discussion* (88.9%)	0.46 (0.36)	Establish relevance	100%
		Use appropriate context	100%
		Clear language	100%
		Activate prior knowledge	100%
		Include students all over classroom	100%
		Use generative topics and questions	100%
		Give corrective feedback	66.7%
		Model academic language	66.7%

* Denotes an explicit mathematics vocabulary instructional practice.
† Denotes a practice that was only observed once during the double coded observations.

Social Validity

At the conclusion of the study, teachers were asked to complete a social validity survey. This survey was modeled after Hirsch (2016) and Romig (2018). The items on the survey asked teachers about the extent to which the CAP-TVs improved their knowledge and application of explicit mathematics vocabulary instruction, as well as the extent to which the performance feedback and coaching led to similar changes in their knowledge and practice. Additionally, teachers rated their overall time commitment to reading and understanding their performance feedback and coaching, watching the CAP-TVs, and their satisfaction with the CAP-PD approach as a new model of delivering PD. The full social validity survey is provided in the Appendix (Table A2) and the results are reported in chapter 4.

Experimental Design

This study used a single-case multiple baseline design across participants, with alternating treatments. The strength of this hybrid design lies in demonstrating replication of treatment effects in each individual participant and, in the case of this study, effects of a single treatment compared to brief periods of no treatment. This study will essentially allow for comparative analysis of two CAP-PD components, the CAP-TVs on their own and the CAP-TVs combined with the performance feedback and coaching component. The study was designed in order to better understand the effects of each and their role in the CAP-PD system as a whole. Each phase of the experiment is described in detail later in this section, and a sample visual analysis is provided in Figure 4.2.

Some additional provisions and protocols were set in place prior to the start of the experiment in order to address some potential threats to validity and experimental control. First, the teachers were randomly assigned to enter the intervention phase once sufficient data points are collected and those points present a stable, predictable trend. Second, the use of email to deliver performance feedback and coaching to individual teachers provided some control over diffusion of treatment, as teachers will be asked not to share or discuss their feedback with colleagues. Additionally, prior to sending, the coaching email will be checked against a fidelity checklist protocol (Figure 3.1) established by Kennedy, Rodgers et al. (2017) to ensure that each email contains the same content.

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

Figure 3.1. Coaching email fidelity checklist (Romig, 2018).

In the following sections, each phase of the experiment is explained.

Baseline (A)

During this phase, all teachers were observed a minimum of five times, in accordance with WWC standards for single-case research (Kratochwill et al., 2013); additional observations were conducted as needed in order to demonstrate a stable, predictable trend. The QVI was used as the decision variable in moving teachers to the first intervention phase. In order to move to the first intervention phase, teachers must have a minimum of five observations, and a reliable, stable trend is established (based on visual analysis of the graphed totals of implementation markers). Once these conditions are met, teachers were randomly selected to enter the first intervention phase.

Intervention Phase (Alternating Treatments)

The baseline phases concluded with each teacher gaining access to the CAP-TVs and my sending them the CAP-TS template. Teachers were given access to the library of CAP-TVs for vocabulary instruction (www.mathvocabsupport.com; also see Table 2.2 for direct links to each CAP-TV). Each CAP-TV pertains to a specific evidence-based practice for explicit mathematics vocabulary instruction. The CAP-TVs contain information that directly mirrors the practices and implementation markers in the CT Scan tool (i.e., what teachers received performance feedback and coaching on). Based on information provided by the teachers about their instructional sequence and pacing, I created a CAP-TS template that teachers could complete with terms they plan to teach, and that they could also customize to fit their specific needs. These materials were be available to teachers throughout the intervention phase as well as during the maintenance phase (i.e., as digital files, they could not be "removed").

Alternating treatments. Within the intervention phase were two alternating treatments, referred to as "treatment on" and "treatment off." During "treatment on" periods (i.e., the first three observations after ending baseline), the teachers received the performance feedback and coaching intervention. During "treatment off" periods (i.e., the next three observations), teachers received no feedback or coaching. In other words, a "treatment off" period is equivalent to a "CAP-TV/CAP-TS only" condition. These alternations of three observations each continued at least three times, yielding nine data points in each condition for each teacher. In many alternating treatments designs, balance between intervention phases is needed in order to determine effectiveness. In this study, this balance was achieved through the use of this "three on, three off" design.

During "treatment on" periods, teachers received the visual performance feedback (i.e., outputs) from CT Scan and the coaching email, sent to them individually after each observation (see Figures 2.3 – 2.5). In accordance with the literature on effective performance feedback, these emails were sent within 24 hours of the class observation (e.g., Scheeler & Lee, 2002; Scheeler, McAfee, Ruhl, & Lee, 2006). This is possible in part because the CT Scan produces the visual performance feedback and other data automatically as soon as the observation is concluded. The coaching narrative is then written based off of that output, in essence explaining the key "findings" from the lesson to the teacher.

Maintenance (D)

Finally, teachers entered the maintenance phase. Here, it was not possible to remove the effects of the professional development, but teachers no longer had access to the CAP-TVs, and did not receive any more performance feedback or coaching. One teacher, Florence, opted to leave the study before maintenance phase data was collected for her. For each of the remaining three teachers, three observations were conducted in this phase. Time constraints and further possible attrition prevented a more extensive maintenance phase.

Data Analysis

The observational data from this study was analyzed through visual analysis of graphed individual teacher data. The primary dependent measure is the Quality Vocabulary Index (QVI), but additional data from CT Scan, including implementation fidelity (i.e., percentage of IMs used) and time (i.e., duration of explicit mathematics vocabulary instruction are also presented as graphs and were similarly analyzed to determine results (see Table 3.1 for a list of data sourced from CT Scan, and Figures 4.1, 4.2, and 4.3). The analysis applied in the current study adhered to the What Works Clearinghouse Single-Case Design Guidelines (Kratochwill et al., 2013). This document is a widely used and highly regarded guide for evaluation of single-case designs.

In single-case designs, there are three key features of visual analysis that determine whether a functional relationship is present: level, trend, and variability (Holcombe, Wolery, & Gast, 1994; Kratochwill et al., 2010). These components of visual analysis are not interdependent; an effect of either one of the interventions could be indicated by one component, but not the others.

Level

Level generally refers to the mean performance of an individual participant, usually within a specific phase of the experiment. This can be estimated by the researcher interpreting the graph, or calculated. The level in one phase is compared to the level in the adjacent phase in order to determine whether the level has increased or decreased. In the current study, a change in level indicated a change (i.e., increased or decreased) time spent on explicit mathematics vocabulary instruction, or a change in the number of IMs used, or a change in the time spent on explicit mathematics vocabulary instruction.

Trend

As level provides a one-dimensional measure of change, trend allows the researcher to see the rate of change, rather than just a single point. A trend drawn on a graph represents the line of best fit through the data points. Trend allows the researcher to see change (or lack of change) in the participants' performance or behavior over time (Franklin, Gorman, Beasley, & Allison, 1996). A trend line that slopes gradually would indicate slower progress (or decline), while a trend line that slopes more sharply would indicate rapid progress. In the current study, trend demonstrates the magnitude and relative speed of changes in each teacher's use of explicit mathematics vocabulary instruction.

Variability

Finally, variability in visual analysis refers to the consistency of the observed performance or behavior. This is another way in which the researcher can observe the effect, or lack thereof, of the intervention. Highly variable data, usually displaying a “zig-zag” pattern, indicates that a teacher did not use explicit mathematics vocabulary instruction consistently in terms of time or quality. if a teacher used an evidence-based practice, but used it inconsistently.

Chapter 4: Results

In the current study, I examined the effects of a performance feedback and coaching intervention on changes in three aspects of sixth grade mathematics teachers' use of explicit vocabulary instruction: (a) the Quality Vocabulary Index (QVI); (b) implementation fidelity as measured by percentage of "implementation markers" (IMs) used; and (c) time spent on explicit vocabulary instruction as measured by the percentage of the total lesson time. I introduced the performance feedback and coaching according to a multiple baseline design (Kazdin, 2011) to assess whether the independent variable induced changes in the instructional behaviors listed above. In this chapter, I present the results of the study.

Using the Classroom Teaching Scan (CT Scan), I collected data for those three dependent variables of interest. During the experiment, the QVI served as the variable on which I primarily based decisions about experimental variations, but in terms of results, it is important to look also at the components of the Quality Vocabulary Index: implementation fidelity and duration of vocabulary instruction. Examining changes in implementation fidelity allows for patterns in teachers' learning to emerge; that is, whether they have learned the steps necessary to implement a new vocabulary practice. Examining duration of vocabulary instruction, as well as proportions of lessons spent on vocabulary instruction, reveals to what extent teachers were willing and able to provide explicit, direct instruction on vocabulary. As reviewed in Chapter 2, this kind of instruction is essential to supporting students with disabilities and others who struggle with the academic language demands of mathematics.

Throughout the experiment, I conducted live observations of the participating teachers' regular mathematics instruction using the CT Scan. During these observations I recorded all of the instructional moves of the teacher, but paid particular attention to recording the use of vocabulary instruction. This focus on vocabulary instruction during observation was mirrored in the performance feedback that each teacher received during the intervention phase. An example of this performance feedback is provided in Appendix B.

For all teachers, the first three observations of the intervention phase were Treatment 1 (i.e., they received the performance feedback & coaching intervention, also noted as "treatment on"). This was followed immediately by Treatment 2 for the next three observations (i.e., no performance feedback or coaching, also noted as "treatment off"). This cycle was then repeated, a total of three or four times for each teacher, depending on her overall response to the treatment. These rapid "on-off" alternations were essential to the initial alternating treatments component of this study's hybrid experimental design. However, as I visually inspected the data over the course of the experiment, it became clear that the measures (i.e., CT Scan menu of vocabulary practices) were not sensitive enough to capture any changes that occurred due to the rapid alternating treatments (i.e., replicated effects or maintained effects). As a result, the graphs presented in this chapter have been collapsed into a simpler, multiple baseline presentation.

In this chapter, I present the results for each teacher in succession (i.e., the order in which they entered the intervention phase), in terms of within-subject findings. For each teacher, I present descriptive data for the Quality Vocabulary Index first, followed by descriptive data for the percent of implementation markers achieved by each teacher in each lesson (i.e., fidelity), and finally followed by descriptive data regarding the total amount of time each teacher spent on any vocabulary instruction, expressed as a percentage of the total lesson time. The tables that accompany the graphs later in this chapter display the descriptive statistics of each of these dependent variables. I then report results in terms of between-subject findings, including one of the two effect sizes used to evaluate the results (percent of non-overlapping data, or PND; both PND and Tau-U are presented in depth later in this chapter). In the final section of this chapter, I present results of the social validity survey, including teachers' expressions of their experiences in the study with regard to their reception of the performance feedback and coaching as well as the professional development on explicit mathematics vocabulary instruction.

Changes in Teachers' Explicit Mathematics Vocabulary Instruction

In the following sections, each teacher's performance is presented in the order in which they entered the intervention phase. Each teacher has been assigned a pseudonym, based on noteworthy women in the fields of mathematics and science. I present each of three dependent variables across the four teachers: Quality Vocabulary Index (QVI), percent of implementation markers for explicit mathematics vocabulary instruction, and overall time spent on vocabulary instruction (in terms of percent of each lesson). In the sections that follow, I present each teacher's performance in each of the three phases of the experiment. For each of these descriptions of teacher performance, I address the first three of Kratochwill et al.'s (2013) steps of visual analysis for single-case research designs, and address the fourth (i.e., demonstration and replication of effect) as a summary to this section:

- Establish a stable, predictable data pattern in the baseline phase.
- Examine the data in each experimental phase for within-phase and between-phase patterns (i.e., level, or mean; trend, or slope; variability, and consistency).
- Examine the data points immediately adjacent to phase changes in the experiment (i.e., immediacy of effect).
- Determine whether there is a demonstration of effect, and if the effect is replicated three times at three different points in time (Kratochwill et al., 2013).

The graphs for visual analysis, as well as tables of descriptive results, are presented on the pages that follow. The x-axis of each graph represents the observation day. It is important to point out that each teacher's observations were conducted on alternate days in order to accommodate the school's block schedule. In addition, observation days may not align to calendar days due to inclement weather (i.e., school closing or delay), full-period testing, or teacher absence. The solid vertical line on each graph represents the change from baseline to the intervention phase. The dashed vertical line on each graph represents the change from intervention to the maintenance phase (n.b.: Florence chose to leave the study early and therefore did not have a maintenance phase; this is discussed later in this chapter).

The y-axis of each graph represents the variable of interest. I present three sets of graphs, each featuring a different dependent variable: The Quality Vocabulary Index (QVI; Figure 4.1), percent of implementation markers for explicit mathematics vocabulary instruction achieved in each lesson (Figure 4.2), and percent of lesson time spent on vocabulary instruction (Figure 4.3).

Table 4.1

Descriptive Statistics for QVI by Participant Across Intervention Conditions

	Baseline				Intervention 1					Intervention 2				
Teacher	M (SD)	Median (Range)	Mode	Slope	M (SD)	Median (Range)	Mode	Slope	PND	M (SD)	Median (Range)	Mode	Slope	PND
Ada	0 (0)	0 (0)	0	0	0.07 (0.14)	0 (0-0.35)	0	0.04	33.3%	0.01 (0.02)	0 (0-0.05)	0	0.01	16.7%
Florence	0.04 (0.06)	0 (0-0.14)	0	-0.02	0.17 (0.21)	0.04 (0-0.57)	0	-0.05	44.4%	0.09 (0.10 4)	0.07 (0-0.24)	0	0.06	50%
Emilie	0.01 (0.03)	0 (0-0.09)	0	0	0.15 (0.14)	0.1 (0-0.43)	N/A	0.03	66.7%*	0.1 (0.16)	0 (0-0.26)	0	0.04	33.3%
Sophie	0.01 (0.02)	0 (0-0.05)	0	0	0.13 (0.1)	0.07 (0.36)	0.07	0.07	77.8%*	0.09 (0.1)	0.06 (0.026)	0	0.05	55.6%*

Note. During “Intervention 1,” teachers were receiving performance feedback and coaching. During “Intervention 2,” teachers were not receiving performance feedback or coaching.
* $p < 0.05$

Table 4.2

Descriptive Statistics for Implementation Fidelity by Participant Across Intervention Conditions

	Baseline				Intervention 1					Intervention 2				
Teacher	M (SD)	Median (Range)	Mode	Slope	M (SD)	Median (Range)	Mode	Slope	PND	M (SD)	Median (Range)	Mode	Slope	PND
Ada	0 (0)	0 (0)	0	0	22% (0.44)	0% (1-100)	0	0.17	22.2%	8% (0.2)	0% (0-50)	0	0.13	16.7%
Florence	29% (0.46)	0% (0-100)	0	-0.19	43% (0.42)	57% (0-100)	0	0.28	0%	34% (0.39)	29% (0-89)	0	0.22	0%
Emilie	7.1% (0.19)	0% (0-50)	0	0	37% (0.44)	0 (0-100)	0	0.1	44.4%	17% (0.35)	0 (0-100)	0	0	11.1%
Sophie	0 (0)	0 (0)	0	0	29% (0.36)	0% (0-88)	0	0.24	44.4%*	36% (0.37)	43% (0-92)	0	0.23	66.7%*

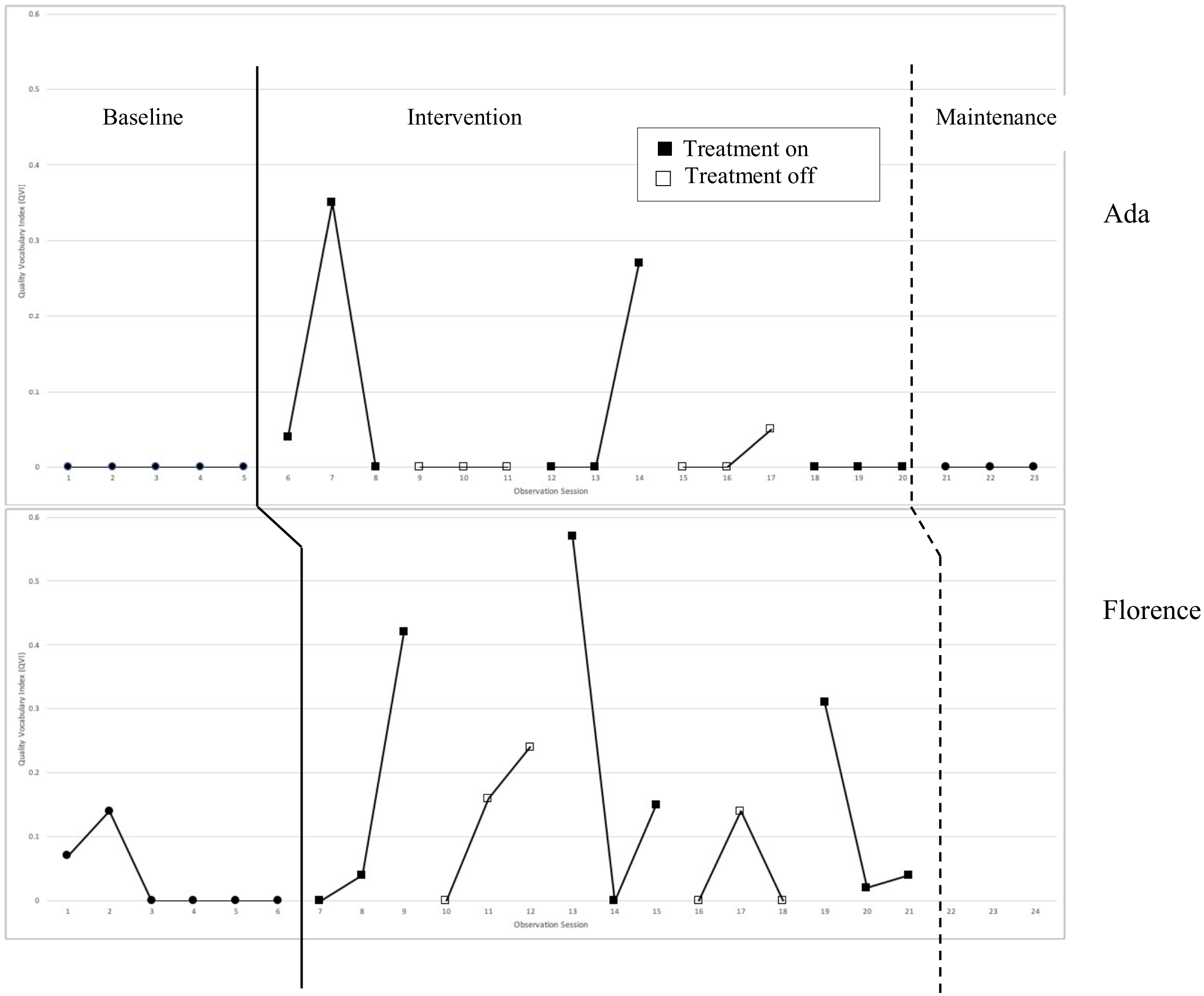
Note. During "Intervention 1," teachers were receiving performance feedback and coaching. During "Intervention 2," teachers were not receiving performance feedback or coaching.
* $p < 0.05$

Table 4.3

Descriptive Statistics for Percent of Lesson Spent on Vocabulary Instruction by Participant Across Intervention Conditions

	Baseline				Intervention 1					Intervention 2				
Teacher	M (SD)	Median (Range)	Mode	Slope	M (SD)	Median (Range)	Mode	Slope	PND	M (SD)	Median (Range)	Mode	Slope	PND
Ada	3% (0.06)	0 (0-18)	0	0	4% (0.07)	0 (0-18)	0	0.17	33.3%	1% (0.01)	0 (0-3)	0	0.13	16.7%
Florence	1.8% (0.03)	0% (0-7)	0	-0.01	10% (0.12)	2% (0-30)	0	-0.03	44.4%	5% (0.06)	4% (0-13)	0	0.03	50%
Emilie	0.9% (0.02)	0% (0-6)	0	0	7% (0.07)	5% (0-24)	N/A	-0.01	44.4%	5% (0.07)	0% (0-17)	0	0.02	33.3%
Sophie	1% (0.02)	0% (0-4)	0	0	7% (0.05)	5% (0-16)	13%	0.04	55.6%*	5% (0.05)	4% (0-14)	0	0.03	44.4%*

Note. During "Intervention 1," teachers were receiving performance feedback and coaching. During "Intervention 2," teachers were not receiving performance feedback or coaching.
* $p < 0.05$


Baseline
Intervention
Maintenance
Treatment on
Treatment off
Ada
Florence
Quality Vocabulary Index (QVI)
Observation Session

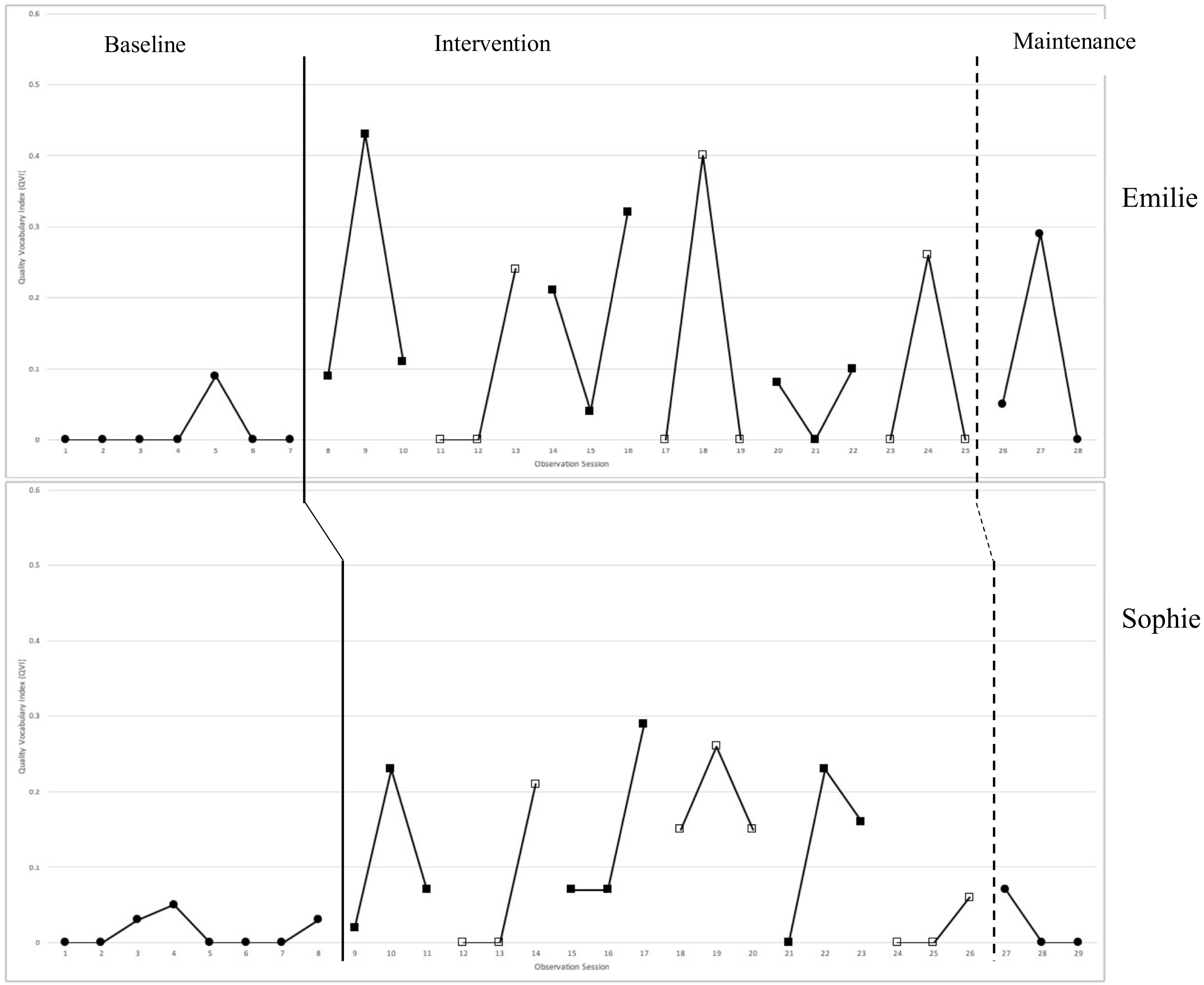
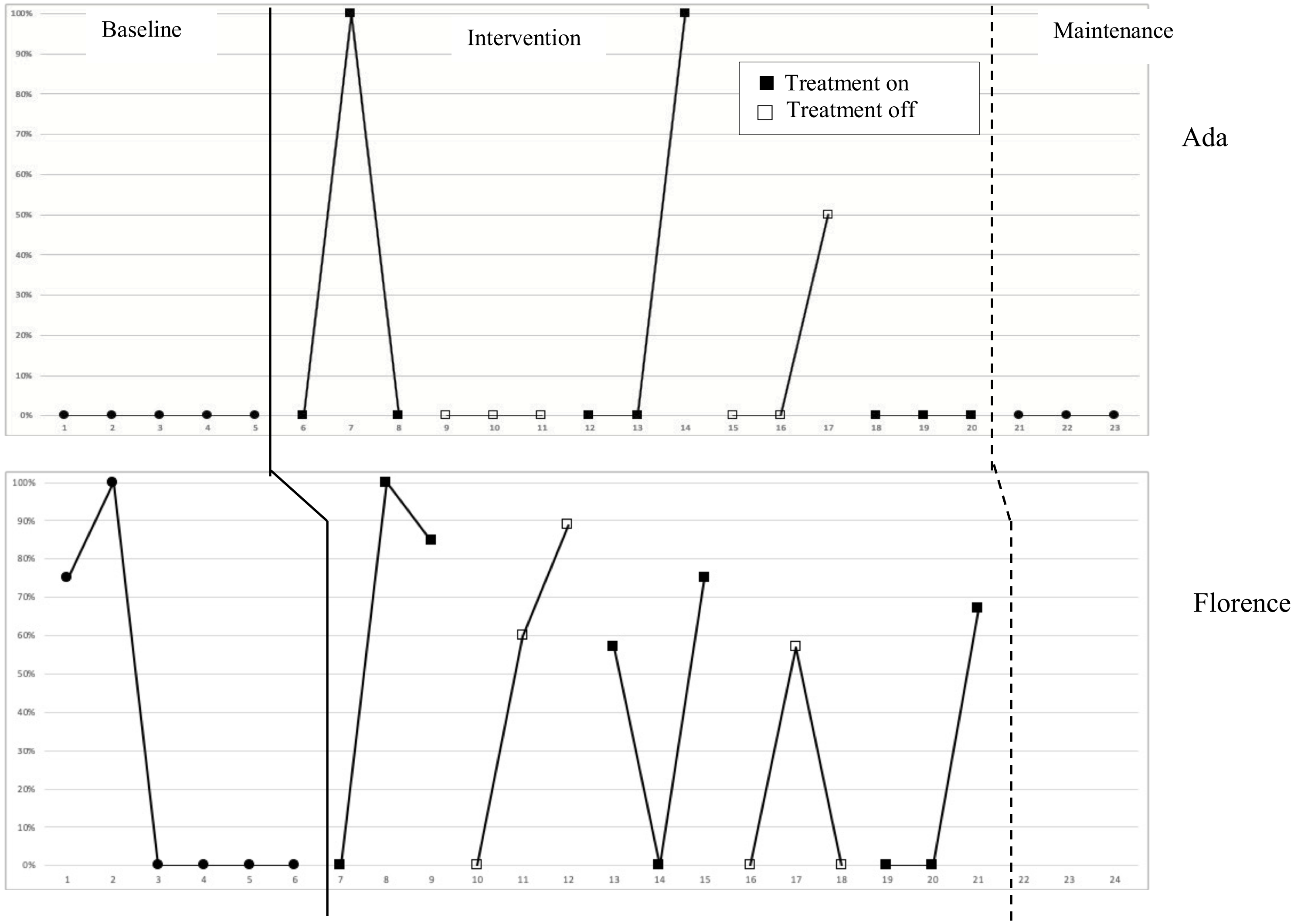


Figure 4.1. Teachers' Quality Vocabulary Index (QVI).


Baseline
Intervention
Maintenance
Treatment on
Treatment off
Ada
Florence

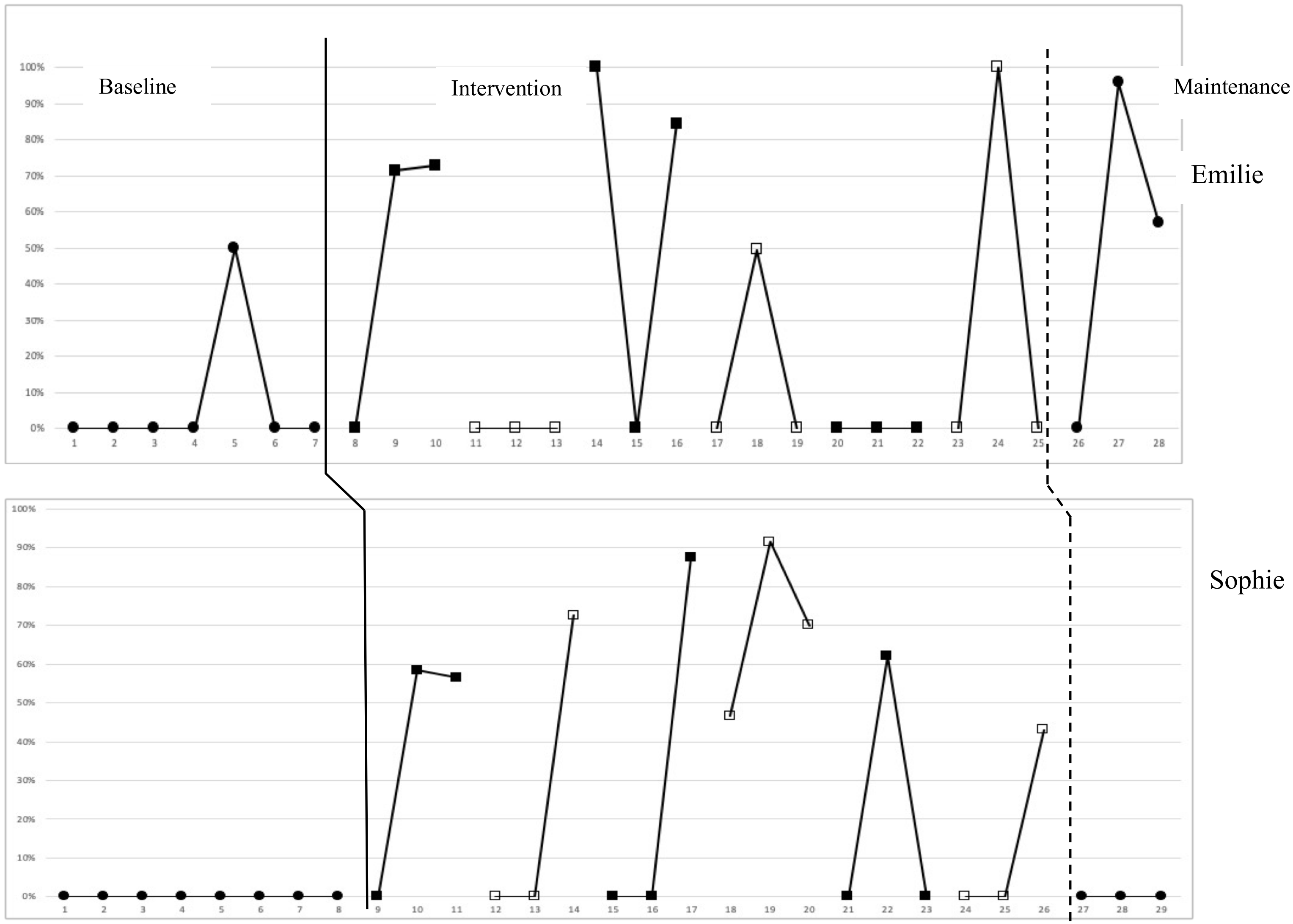
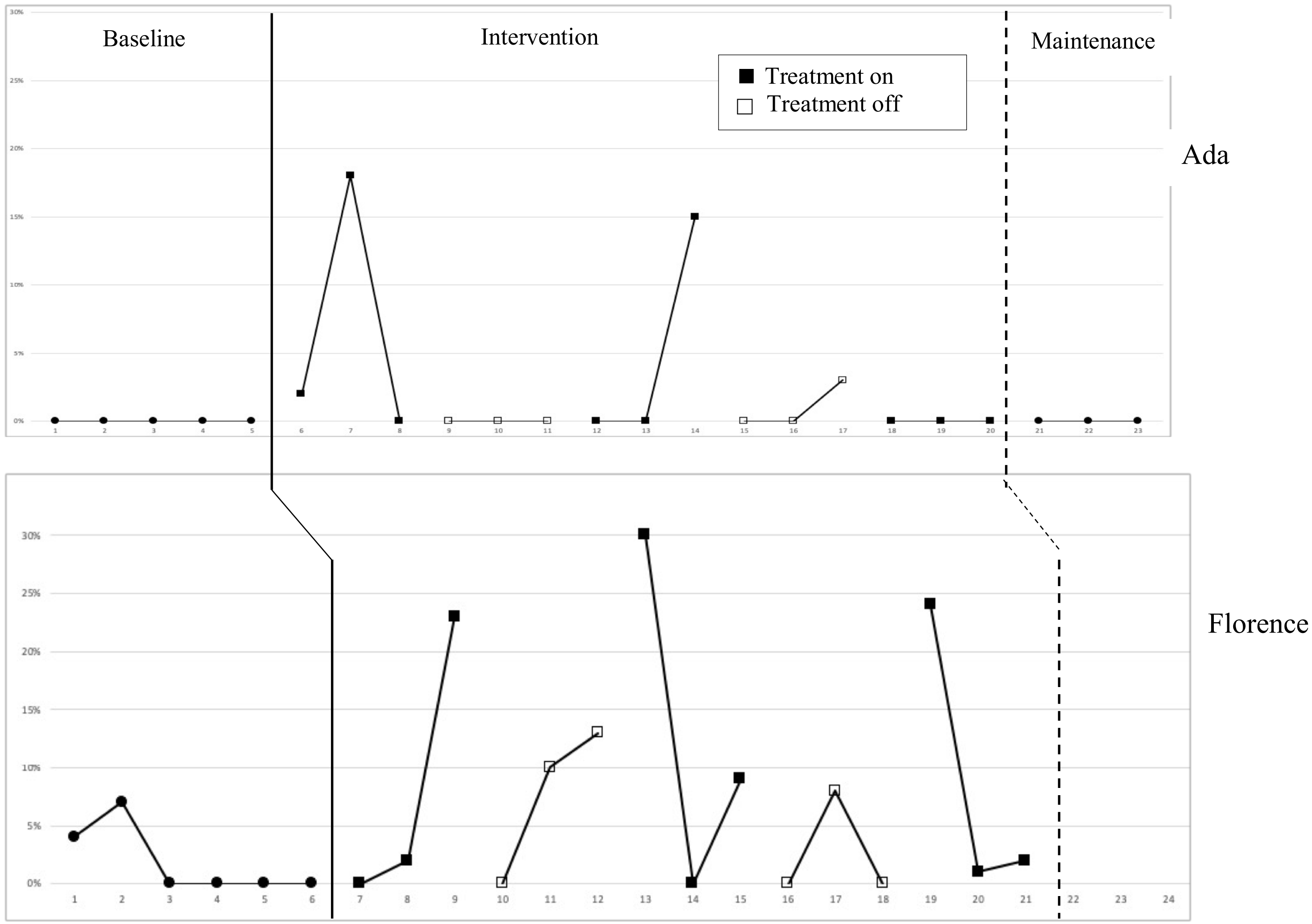


Figure 4.2. Teachers' Percent of Explicit Mathematics Vocabulary Instruction Implementation Markers.


Baseline
Intervention
Maintenance
Treatment on
Treatment off
Ada
Florence
30%
25%
20%
15%
10%
5%
0%

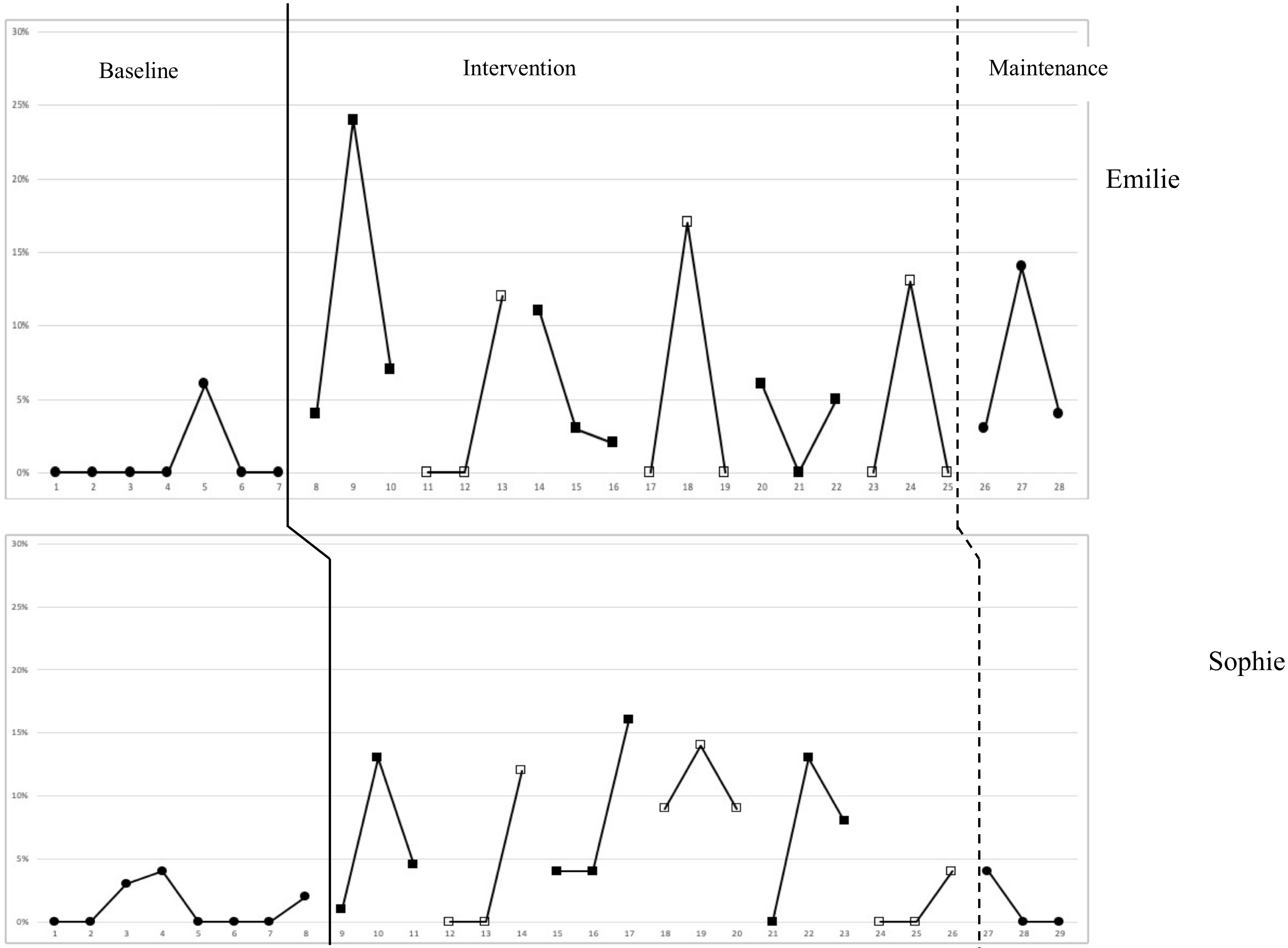


Figure 4.3. Teachers' Percent of Lesson Spent on Vocabulary Instruction (Any Practices).

Ada

Baseline. Ada was the first of the four teachers to enter the intervention phase. She was observed five times during the baseline phase. None of those lessons included vocabulary instruction of any type. Ada started receiving the intervention after Observation #5; thus, the data from Observation #6 forward reflect her performance during intervention.

Intervention. Ada's explicit mathematics vocabulary instruction remained minimal throughout the study. Ada's mean QVI during "treatment on" alternations was .05, and .03 during "treatment off" alternations. With little change in level, the overall trend (slope) in QVI data was also minimal. The net slope (i.e., combined average slopes of "treatment on" and "treatment off" mini-phases) in QVI across the entire intervention phase was only .05 for Ada. With vocabulary instruction happening only sporadically, Ada's QVI data appear to be less variable than her colleagues', but this is mainly because in most of her observed lessons there was no vocabulary instruction.

When she first received the intervention, Ada began to implement explicit mathematics vocabulary instruction in the first two observations during "treatment on" (QVI of .04 and .35, respectively). However, she did not replicate this effect; in fact, only two observed lessons throughout the rest of the study contained any vocabulary instruction. With regard to implementation fidelity, those same first two observations also show a somewhat immediate effect of the intervention. Ada did not use an explicit mathematics vocabulary practice in the first "treatment on" observation (Observation #6);

instead, she used another vocabulary practice (Ask student(s) to state a definition) with 100% fidelity.

In the next observation, however, she did use one of the explicit mathematics vocabulary practices (Facilitate an academic discussion), again with 100% fidelity. With regard to percent of the lesson spent on vocabulary, Ada spent only 2% (almost 1.5 min.) of the lesson on vocabulary instruction in Observation #6, but 18% (nearly 12 min.) of the lesson in Observation #7. In summary, there was an immediate effect for the intervention following baseline for Ada, but that effect was not replicated.

Maintenance. Ada was observed three times during the maintenance phase. No vocabulary instruction was observed in those three lessons.

Florence

Baseline. Florence was the second teacher to enter the intervention phase. In her first two observations, she used brief academic class discussions as part of her mathematics vocabulary instruction, prior to receiving any professional development. Following those two observations, her baseline trend remained stable at zero on all variables, for four consecutive observations. Following the multiple baseline design, Florence received the performance feedback and coaching intervention after Observation #6, so the data from Observation #7 on shows the influence of the intervention.

Because of those first two lessons that included discussions, her mean QVI for baseline was 0.04; based on the research to date that has used the QVI as a dependent measure, this is close enough to zero as to be non-existent (Kennedy, Rodgers et al., 2018). In those instances, Observation #1 included approximately 3 min. of explicit vocabulary instruction (total observation time was 76 minutes) and Observation #2 included 5 minutes and 29 s. of explicit vocabulary instruction (total observation time was 80 min). The median and mode QVI in the entire baseline phase was 0. With regard to implementation, Florence's baseline mean was 29% but this is again skewed by her first two observations (i.e., 75% and 100% respectively, followed by four consecutive points of 0%). Both the median and mode for this variable in baseline were zero. With regard to time spent on explicit mathematics vocabulary instruction during baseline, Florence's mean percent of lesson spent on this was 1.8%.

Intervention. Florence's mean QVI score during the "treatment on" alternations was .17, and .09 when she was not receiving the intervention. The trend in QVI scores over this entire phase was positive but minimal, as measured by a net slope of only .02. With regard to the percent of implementation markers achieved per lesson, Florence's mean percentage was 43% while receiving the performance feedback and coaching intervention, and 34% when not receiving it. The overall trend in each of those treatment conditions was positive, at .28 during "treatment on" and .22 during "treatment off," and the net slope of .5 for the entire intervention phase showed that Florence gradually, and somewhat consistently, improved her implementation fidelity of the explicit mathematics vocabulary practices over the course of the intervention. This pattern of increasing implementation fidelity does not appear in the time spent teaching with explicit mathematics vocabulary instruction, however, as Florence's mean percent of lessons during "treatment on" was 10%, and 5% during "treatment off." The trend in this variable has a net slope of 0 for the entire phase, due to a negative slope coefficient of -.03 in "treatment on" offset by a positive slope coefficient of .03 during "treatment off."

Florence's performance in terms of both QVI score and percent of lesson devoted to explicit mathematics vocabulary instruction was highly variable (see Figures 4.1 and 4.3). The lessons in which she provided vocabulary instruction were somewhat sporadic and inconsistent. As the QVI statistic takes duration of vocabulary instruction into account as a proportion of the entire observation, the presence of this variability across phases is not surprising. However, as Figure 4.2 indicates, there is somewhat less variability in Florence's implementation fidelity during each instance of explicit mathematics vocabulary instruction.

As shown in Figure 4.1, there is a very small increase in Florence's QVI score from the last data point in baseline (QVI = 0) to its adjacent point in the intervention (QVI = .04), but as noted elsewhere, a QVI of .04 is, for measurement purposes, virtually zero. In that instance, Observation #8 included 1 min 48 sec of explicit vocabulary instruction (total observation time was 81 min). There is not a notable "jump" in QVI score from baseline to intervention until the third intervention observation (QVI = .42). That observation corresponds to the third consecutive lesson on which Florence was receiving performance feedback and coaching. During the adjacent "treatment off" mini-phase, her QVI score reverted immediately back to zero, and increased only minimally. When she received the treatment again, the QVI score jumped up again (QVI = .57).

However, in all but one of the treatment alternations, Florence had at least one observation with a QVI score of 0; as a result, there is not a replicated effect as measured by QVI.

With regard to implementation, there is not an immediate effect shown from the last baseline data point to the first intervention data point (Figure 4.2), but similarly to the pattern in QVI previously described, there is a “jump” in the second and third intervention data points (100% and 84.8%, respectively). Throughout the rest of the intervention phase, Florence’s implementation percentage is consistently medium to high, when present, but the presence of vocabulary instruction itself is not consistent, limiting the extent to which the effect could be replicated.

The apparent relationship between the QVI for these data points and percent of implementation markers used is that in her second intervention observation, Florence used one of the explicit mathematics vocabulary practices with 100% of the practice’s implementation markers used, but only for about two minutes. As a result, that implementation data point is high and it occurs quickly relative to baseline, but the corresponding QVI for that point is very low (.04), because the vocabulary instruction lasted only two minutes out of an 80-minute observation. This example is typical of Florence’s overall performance with regard to explicit mathematics vocabulary instruction: When instruction was present, she generally implemented it with a medium to high level of fidelity, but did not implement it for very long in most instances.

With regard to time spent on explicit mathematics vocabulary instruction per lesson (Figure 4.3), the cross-phase pattern (i.e., from baseline to intervention) for Florence is nearly identical to that of her QVI data. This is largely because both variables are functions of duration; that is, the less time that is spent on explicit mathematics vocabulary instruction, the lower the results from these two variables will be, and the more similar the effect, if any, would be. In summary, there was a somewhat immediate effect (i.e., within the first three observations) of the intervention following baseline for Florence. This effect was replicated twice, although with high variability, on subsequent alternations of receiving the performance feedback and coaching intervention.

Maintenance. Florence requested to leave the study after Observation #21 (this was at the start of the school's winter holiday break). As a result, I collected no maintenance data for her. She did agree to complete the social validity survey, and her responses from that measure are discussed later in this chapter.

Emilie

Baseline. Emilie was the third teacher to enter the intervention phase. Following the multiple baseline design, Emilie received the performance feedback and coaching intervention after Observation #7, so the data from Observation #8 onward shows the influence of the intervention. Emilie had one observation during baseline when she used one of the explicit mathematics vocabulary practices. All other baseline observations had a QVI of 0. For the one baseline observation in which Emilie used one of the explicit mathematics vocabulary practices (Use a morphological approach), she did so for just over 4 minutes, with only 50% of its implementation features, bringing her mean for implementation during baseline to 7.14%, but the median and mode remained at zero. The slope coefficients for the baseline phase remained very near zero on all three variables.

Intervention. During the intervention phase, Emilie's mean QVI score was .13 during the "treatment on" alternations, and .15 during "treatment off." This difference is too small to consider it an effect. The trend in QVI scores over this entire phase was mostly neutral, as measured by a net slope of .07. With regard to the percent of implementation markers achieved per lesson, Sophie's mean percentage was 26.6% while receiving the performance feedback and coaching intervention, and 36.5% when not receiving it, which shows some potential carryover effect of the intervention. The overall trend for this variable in each of the treatment conditions was positive but small, at .10 during "treatment on" and 0 during "treatment off," and a net slope of .10 for the entire intervention phase. Finally, the mean percentage of her lessons that Sophie spent on vocabulary instruction during "treatment on" was 7%, and 5% during "treatment off." The trend in this variable has a net slope of just 0.01 for the entire phase, indicating that her time spent on vocabulary instruction most likely did not increase as an effect of the intervention.

Visual analysis of Emilie's data show that there was a small immediate effect from the last baseline observation to the first intervention observation, where her time spent on vocabulary instruction went from none to 4% of her lesson (approximately 3 minutes). She taught vocabulary on each of the three days of the first “treatment on” mini-phase, once spending as much as 21% (about 15 minutes) of a lesson (Observation #9) on explicit mathematics vocabulary instruction. This effect was replicated in the next “treatment on” phase (i.e., when Emilie began receiving performance feedback and coaching again), albeit to a lesser degree than the initial effect. In summary, there was an immediate effect for the intervention following baseline for Emilie. That effect was replicated twice during the intervention phase, although to a lesser extent each time.

Maintenance. Emilie continued to use explicit mathematics vocabulary instruction during each of her three maintenance observations. Her mean QVI for that phase was .13 (median = .06). Emilie also attained her highest mean implementation fidelity during maintenance, at 51%. Also during maintenance the mean percent of lesson time spent on vocabulary instruction was 7%, meeting her highest corresponding mean from the intervention phase.

Sophie

Baseline. Sophie was the fourth and final teacher to enter the intervention phase. She entered the intervention phase on Observation #9, but received her first performance feedback and coaching intervention after Observation #8. This was done so that her first observation in the “treatment on” mini-phase would have the influence of the intervention. Sophie had two baseline observations in which she taught vocabulary, which was recorded with the CT Scan and thus influenced her QVI scores, but the instruction she provided (e.g., Asks student to state definition) did not correspond to any of the practices or implementation markers for explicit mathematics vocabulary instruction.

Intervention. During the intervention phase, Sophie taught vocabulary fairly regularly, with only three lessons where no vocabulary instruction was observed at all. Sophie's mean QVI score was higher during the "treatment on" alternations (.13) than her mean QVI score during "treatment off" (.09). The trend in QVI scores over this entire phase was positive, as measured by a net slope of .12. With regard to the percent of implementation markers achieved per lesson, Sophie's mean percentage was 33% while receiving the performance feedback and coaching intervention, and 29% when not receiving it, which shows some carryover effect of the intervention. The overall trend for this variable in each of the treatment conditions was positive, at .24 during "treatment on" and .23 during "treatment off," and the net slope of .47 for the entire intervention phase showed that Sophie improved her implementation fidelity of the explicit mathematics vocabulary practices over the course of the intervention, but still only had a medium level of fidelity overall. The mean percentage of lessons that Sophie spent on vocabulary instruction during "treatment on" was 7%, and 5% during "treatment off." The trend in this variable has a net slope of 0.07 for the entire phase, indicating that her time spent on vocabulary instruction most likely did not increase as an effect of the intervention.

Like most of the other teachers, Sophie's performance in terms of both QVI score and percent of lesson devoted to explicit mathematics vocabulary instruction is highly variable (see Figures 4.1 and 4.3). One explanation for the high variability on QVI and percent of lesson spent on vocabulary is that there were several lessons in which Sophie did some vocabulary instruction, but still did not use the explicit mathematics vocabulary instructional practices (e.g. Observations #15, #16, and #23). For example, she often asked students to define a vocabulary term, which is recorded in the CT Scan as vocabulary instruction (and therefore contributes to QVI score and percent of lesson), but her QVI score would otherwise be a "0" if the QVI only included the explicit mathematics vocabulary instructional practices. On the other hand, Sophie's implementation remained low throughout the intervention phase, but appears to show more consistency than the other two dependent variables (see Figure 4.2).

Visual analysis of Sophie's data show that there was a somewhat immediate effect (i.e., within the first three observations) from baseline to intervention. Sophie's QVI of zero at the end of baseline (Observation #8) increased to .23 by Observation #10 and .07 in Observation #11. The percent of lesson time that Sophie spent on vocabulary instruction went from none in baseline to 13% of her lesson (just over 8 minutes), with 58% implementation fidelity, by the second intervention observation. This effect was replicated in the next "treatment on" mini-phase, but to a lesser degree than the initial effect. During her last "treatment on" mini-phase (Observations #21-23), Sophie actually spent less time and implementation fidelity on vocabulary instruction than the "treatment off" phase immediately prior. In summary, there was an immediate effect for the intervention following baseline for Sophie. That effect was replicated twice during the intervention phase.

Maintenance. On the first maintenance observation, Sophie did not use any explicit mathematics vocabulary instruction, but did spend 4% of the lesson (about three minutes) on other vocabulary instruction (e.g., Asks student(s) to state a definition). Sophie did not use any vocabulary instruction during the last two of her three maintenance observations. As a result, Sophie's mean QVI for the maintenance phase was just 02.

Determining the Presence of a Functional Relationship

According to Kratochwill et al. (2013), the fourth and final step in visual analysis of results in single-case research designs is to "integrate the information from all phases of the study" (p. 31) to assess whether a functional relation exists between the dependent variable(s) and independent variable. For a functional relation to exist, an effect must be demonstrated at three different time points. In this section, I will assess each teacher's results, as well as across all of their results, for the presence of a functional relationship.

In Ada's case, the slope of her performance in terms of QVI showed a minimally positive trend during the intervention phase as a whole (from .04 during "treatment on" to 0.01 during "treatment off," net slope of .05), and visual analysis does suggest the presence of a minimal functional relationship. The level of her QVI performance across the intervention phase was 0.05, which she met or exceeded at three separate time points. This pattern and replication of effect is also present in Ada's results on the other two dependent variables. However, at the end of the intervention and into the maintenance phase, Ada's mean QVI returned to 0 as she did not implement any more vocabulary instruction. Overall, these data do not indicate that providing performance feedback and coaching influenced Ada's instructional practice with regard to explicit mathematics vocabulary instruction.

With regard to Florence's results, the slope of her performance in terms of QVI remained fairly neutral during the intervention phase as a whole, with a combined (net) slope of 0.02 for both treatment on and treatment off. Visual analysis suggests the presence of a moderate functional relationship. The level of Florence's QVI performance across the intervention phase was 0.11, which she exceeded at seven separate time points during that phase, exceeding the minimum of three such points needed for a functional relationship to exist. This pattern and replication of effect is also present in Florence's results on the other two dependent variables. Because she opted to leave the study after completing the intervention, no maintenance data were collected. Overall, these data indicate that providing performance feedback and coaching influenced Florence's instructional practice with regard to explicit mathematics vocabulary instruction.

With regard to Emilie's results, the slope of her performance in terms of QVI trended upward over the intervention phase, with a combined (net) slope of 0.12 for both treatment on and treatment off. Visual analysis suggests the presence of a moderate functional relationship. The level of Emilie's QVI performance across the intervention phase was 0.15, which she exceeded at five separate time points during that phase, exceeding the minimum of three points needed for a functional relationship to exist. This pattern and replication of effect is also present in Emilie's results on the other two dependent variables. Additionally, Emilie continued to implement explicit mathematics vocabulary instructional practices at the end of the intervention phase and into the maintenance phase. Overall, these data indicate that providing performance feedback and coaching influenced Emilie's instructional practice with regard to explicit mathematics vocabulary instruction.

With regard to Sophie's results, the slope of her performance in terms of QVI trended upward over the intervention phase, with a combined (net) slope of 0.12 for both treatment on and treatment off. Visual analysis suggests the presence of a moderate functional relationship. The level of Emilie's QVI performance across the intervention phase was 0.13, which she exceeded at eight separate time points during that phase, exceeding the minimum of three points needed for a functional relationship to exist. This pattern and replication of effect is also present in Emilie's results on the other two dependent variables. However, at the end of the intervention and into the maintenance phase, Sophie's QVI returned to near zero, as she implemented any more vocabulary instruction only twice during the last six observations. Overall, these data indicate that providing performance feedback and coaching influenced Sophie's instructional practice with regard to explicit mathematics vocabulary instruction.

The multiple baseline element of the study's design allows for demonstration of effects across participants at separate (i.e., independent) time points. Looking across participants, each of the four teachers demonstrated an immediate or nearly immediate effect (i.e, within the first three intervention observations) when I introduced the performance feedback and coaching intervention. Visual analysis shows that this effect is replicated the first time that the intervention is re-introduced (i.e., when teachers being receiving performance feedback and coaching again, after the first "treatment off" alternation).

In summary for this section, the results across all four teachers does suggest a functional relationship, albeit small, between providing performance feedback and coaching alongside CAP-TVs and the teachers' implementation of explicit vocabulary instruction. In the next section, I present evaluations of effect size for these results in order to uncover more specific information about the timing and magnitude of the intervention effects.

Measures of Effect Size

Percent of non-overlapping data. Effect size is essentially a measure of the magnitude of a relationship between dependent and independent variable (Allison & Gorman, 1994; Tarlow & Penland, 2016). In single-case research designs, a common way to assess effect size is by calculating the percent of non-overlapping data (PND). Researchers can also calculate statistical significance (*p* value) for PND. A statistically significant PND result indicates that there is a strong probability that the apparent effect is as it appears in the calculation.

While PND is not without its flaws and limitations, it is presented here in order to illustrate the differences between the teachers' performance in terms of their vocabulary instruction. PND also helps to show the extent to which each teacher implemented vocabulary instruction as a result of the performance feedback and coaching intervention as well as the training they received on explicit mathematics vocabulary instruction (i.e., CAP-PD), especially because each teacher started the study with virtually no explicit vocabulary instruction at all. PND and statistical significance are reported for each teacher in Tables 4.1, 4.2, and 4.3.

On the QVI, both Emilie and Sophie had moderate PND (66.7% and 77.8%, respectively) that were statistically significant when they were receiving the performance feedback and coaching intervention (i.e., Intervention 1 on Table 4.1). With regard to implementation fidelity, only Sophie showed a statistically significant PND, but it was present in both intervention phases. However, at only 44.4% and 66.7%, these effects are debatable. Finally, with regard to percent of lesson time spent on vocabulary instruction, Sophie was again the only teacher with statistically significant PND in both intervention phases, and again with effects that are minimal at best (55.6% and 44.4%). It is important to note that PND calculation, as well as its *p* value, is sensitive to the number of data points in the baseline phase (Tarlow & Penland, 2016), and as Sophie had the longest baseline phase, this could be affecting the results of this effect size calculation.

Tau-U. In addition to PND, I also calculated Tau-U statistics for within- and between-phase trends for each participant as well as for the study as a whole. Tau-U is essentially a matrix of several different levels, or partitions, of contrasts of nonoverlapping data (Parker, Vannest, Davis, & Sauber, 2010). The usefulness of PND is limited in results in which trends are minimal, as is the case in the current study

(Vannest & Ninci, 2015). As a result, a report of the effects of this study's intervention in terms of Tau-U is presented here as a complement to PND. Tau-U results were calculated using the QVI statistic. Using the Tau-U online calculator (available for free at http://www.singlecaseresearch.org/calculators/tau-u), I first calculated "single contrasts," in which I did not differentiate between "treatment on" and "treatment off" conditions during the intervention phase (i.e., all data from the intervention phase was combined as if it were a single treatment). I did this in order to get a picture of the overall effects of the intervention throughout the course of the study (i.e., changes between phases). I then calculated "multiple contrasts" by comparing phase trends within each participant in order to get a closer, sharper look at the intervention effects (i.e., changes between treatments). Results for Tau-U effect sizes are also reported in Table 4.4.

Changes between phases. The calculation of Tau-U at this level of contrast shows changes in each participant between each phase and over the course of the entire study. None of the teachers had statistically significant Tau-U results across the entire study (i.e., the contrast of baseline to maintenance). Similarly to the PND results, only Emilie and Sophie had statistically significant effects in this calculation (using QVI data). Emilie showed a Tau-U trend of .52 (p = .05) from baseline to intervention, and .81 (p = .05) from intervention to maintenance. Sophie had a statistically significant result from baseline to intervention (Tau-U = .58, p = .02) but not from intervention to maintenance. Tau-U results can be interpreted as "improvement trends," (Parker et al., 2010); thus, Emilie's QVI improved approximately 52% from baseline to intervention and 81% from intervention to maintenance. Sophie's QVI improved approximately 58% from baseline to intervention, but the improvement from intervention to maintenance was actually negative and was also not statistically significant (Tau-U = -.5, p = .17).

Effects of the intervention for Ada and Florence's as measured by Tau-U are less promising. Looking at the trend from baseline to intervention, Ada showed a Tau-U of .27 (p = .48). Florence showed a Tau-U of .49 from baseline to intervention (i.e., nearly 50% improvement in QVI score), but this result was non-significant at p = .09. Looking at the trend from intervention to maintenance, Ada's Tau-U was actually negative, at -.27 (p = .48). This is not unexpected given the observed data (i.e., she stopped attempting to use the explicit mathematics vocabulary practices before starting the maintenance phase). Finally, because Florence did not complete a maintenance phase, this level of Tau-U could not be calculated for her.

Changes between treatments. The calculation of Tau-U at this level of contrast shows changes in each participant between "treatment on" and "treatment off" periods. Again, neither Ada nor Florence showed much of an effect of treatment when these periods were contrasted; in fact, both teachers' Tau-U was negative (-.04) and not statistically significant (p = .86). Emilie also showed a negative Tau-U (-.13) between treatments that was also not statistically significant (p = .55). However, when Emilie's baseline QVI data was contrasted with just the QVI from days when she was receiving the intervention, there was a statistically significant Tau-U effect (Tau-U = .57, p = .03).

This can be interpreted as there having been a moderate effect on her explicit mathematics vocabulary instruction (57% improvement) due to receiving performance feedback and coaching along with the CAP-TVs, compared to her vocabulary instruction when not receiving any of those interventions. Finally, Sophie also showed positive, moderate effects when her baseline QVI was contrasted with each of the treatment periods. Contrasting baseline and the days when she was receiving the intervention ("treatment on"), Sophie showed an improvement trend of .68, or 68% (p = .01). Contrasting baseline and the days when she was not receiving the intervention (but had access to CAP-TVs), she showed an improvement trend of .54, or 54% (p = .04).

Table 4.4

Tau-U Effect Size Results Across and Between Phases

	Baseline to Intervention	Int. 1 to Int. 2	Intervention to Maintenance	Baseline to Maintenance	Weighted Average
Ada	.27 (p = .38)	-.04 (p = .86)	-.27 (p = .48)	0 (p = 1)	.11 (p = .52)
Florence †	.49 (p = .09)	-.04 (p = .86)	N/A	N/A	.30 (p = .06)
Emilie	.52* (p = .05)	-.13 (p = .55)	.15 (p = .69)	.81* (p = .05)	.26 (p = .09)
Sophie	.58* (p = .02)	-.12 (p = .56)	-.05 (p = .17)	.08 (p = .84)	.33* (p = .02)_

* denotes statistically significant result at p < .05

† Florence's baseline data was adjusted when used in this calculation, in order to correct for a positive baseline trend.

As Table 4.4 shows, all four of the teachers showed some amount of improvement as far as QVI score over the course of the study, but the highest improvement as well as the only one with statistical significance was Sophie's. Over the course of the study, Sophie's QVI improved approximately 33% ($p = .02$). Florence and Emilie's improvements in QVI were approaching statistical significance, with 30% and 26% improvement trends, respectively. In Chapter 5, I discuss some of the contextual factors that may have contributed to these effects.

Social Validity Results

At the conclusion of their respective maintenance phase, each teacher also completed a social validity survey, administered and collected via Qualtrics. The social validity measure was a survey made up of a total of 27-29 items, (depending on responses to some Likert-type items, teachers were then presented with an optional item to explain their response). This survey consisted primarily of Likert-type items, with opportunities for elaboration on some of those items in open-ended responses. All Likert-type items were required, but open-ended elaboration responses were optional.

In general, the teachers were satisfied with the professional development, teaching materials, performance feedback, and coaching that they received in the study. All four of them reported that they increased their time on vocabulary instruction to some degree (i.e., from "none at all" to something more, or from "minimal; only when questions came up" to something more). All four of the teachers indicated that they would include explicit mathematics vocabulary instruction in future lessons.

Table 4.5 presents the responses for survey items related to the performance feedback and coaching intervention. Table 4.6 presents the responses for survey items related to the CAP-PD (i.e., the CAP-TV videos and CAP-TS slide template, and their relationship to the performance feedback). Following these tables is a brief summary and explanation of the results.

Table 4.5

Social Validity Survey Results by Teacher, Related to Performance Feedback & Coaching Intervention

Item	Teacher			
	Ada	Florence	Emilie	Sophie
How often did you look at your CT Scan data output (the link included in your feedback document)? 1 = I didn't look at it very often. 2 = Most of the time 3 = Every time I received it (and sometimes more than that)	1	2	1	3
Overall, to what extent did receiving the feedback influence your planning and/or instructional decisions with regard to vocabulary instruction? 1 = It wasn't really relevant or helpful. 2 = It was sometimes relevant or helpful. 3 = It was very relevant and/or helpful.	2	2	2	2
Which components of the feedback did you find ***most*** relevant or helpful for your instructional planning (especially for vocabulary instruction)? Rank each component, with 1 being the ***most*** relevant.	Frequency counts (OTRs, etc)	"One Big Thing"	N/A*	Minutes spent on vocabulary (line graph)
In general, which feedback component did you think was ***easiest*** to understand quickly?	Frequency counts (OTRs, etc)	Graphics (pie graphs, timeline)	Written	Frequency counts (OTRs, etc)
In general, which feedback component did you think was ***most*** informative?	Frequency counts (OTRs, etc)	Written	Graphics (pie graphs, timeline)	Frequency counts (OTRs, etc)

* This item was not displayed to the teacher because she answered "not very helpful" to a previous item.

Table 4.6

Social Validity Survey Results by Teacher, Related to CAP-PD

Item	Teacher			
	Ada	Florence	Emilie	Sophie
What is your impression of the overall quality of the CAP-TV videos? 1 = Not easy to follow or informative 2 = Most are not that informative. 3 = They're OK. 4 = Most are easy to follow and informative. 5 = They are generally easy to follow and informative.	5	5	4	4
What is your impression of the overall relevance of the instructional practices you were taught about in the CAP-TV videos, ***in terms of mathematics instruction***? 1 = The vocabulary practices weren't relevant to my content. 2 = They were mostly relevant, but I felt I had to make a lot of adjustments to make them work for my content. 3 = They were mostly relevant, and I felt I could make them work for my content somewhat easily. 4 = They were relevant and I felt I could easily incorporate them into my content.	3	2	3	3
What is your impression of the overall relevance of the instructional practices you were taught about in the CAP-TV videos, ***in terms of your students' academic needs***? 1 = I do not think these vocabulary practices would support my students' math learning and growth. 2 = I think these vocabulary practices might support my students' math learning and growth. 3 = I think these vocabulary practices did (or could) support my students' math learning and growth.	3	3	2	2
To what extent were the vocabulary practices and implementation features from the CAP-TVs *easy to implement* in your teaching? 1 = Most were not easy to implement.	2	4	3	2

2 = Some were easier than others. 3 = Most were easy to implement, with some adjustments. 4 = Very easy to implement.				
To what extent were the vocabulary practices and implementation features from the CAP-TVs *easy to learn*? 1 = I didn’t understand any of them. 2 = I didn’t really understand 1-2 of them. 3 = Most were easy to learn. 4 = Very easy to learn.	3	4	4	4
How likely are you to use the materials from the CAP-PD again? 1 = Extremely unlikely (I definitely won’t) 2 = Somewhat unlikely (I probably won’t) 3 = Not sure 4 = Somewhat likely (I probably will) 5 = Very likely (I definitely will)	4	5	4	5

Impressions of the performance feedback and coaching intervention. All four of the teachers rated the performance feedback and coaching as "sometimes helpful," but there was some variation in how often and for how long each teacher looked over her feedback. Ada said that she did not look at her feedback very often and added that she "thought it was a cool way to keep track of what goes on in the class, I just didn't take the time to look at it very often because there's always so much to do as a teacher." Emilie, who co-taught her class with an English as a Second Language (ESOL) teacher, felt that the feedback, which only tracked Emilie's instructional moves, was not very helpful because "the ESOL collaborator often presented & practiced vocabulary with students." Interestingly, Emilie had significant effects for her use of explicit mathematics vocabulary instruction when she used it, but despite receiving the performance feedback and coaching, the ESOL co-teacher continued to do the vast majority of the vocabulary instruction in their class.

Another concern that the teachers had with regard to the intervention was that they did not get opportunities to discuss their performance feedback. The teachers were asked at the beginning of the study not to share their feedback with the others, but because they frequently plan together (typically on at least a twice-weekly basis), some of them felt that they might have planned to teach more vocabulary if they could have discussed the feedback from their previous vocabulary instruction. Ada described feeling "limited in how I could use [the feedback]. I just found it hard to motivate myself to work to incorporate some of those practices because I had to do it myself and couldn't plan with my team."

Perceptions of overall effectiveness of CAP-PD. The teachers responded positively to the overall experience of the professional development and training on explicit mathematics vocabulary instruction. On Likert-type items (see Table 4.6), all four teachers rated the CAP-TVs as easy to follow, easy to learn the practices from, and informative. Florence added, "this study helped me recognize a need for explicit vocabulary instruction and to include words that I didn't necessarily see as math words but would assist students in their development of concepts." Both she and Emilie noted, however, that they sometimes struggled to budget instructional time for vocabulary instruction. Emilie explained that "sometimes I made a decision not to teach as much vocabulary in favor of other priorities." Florence cited student attention and "time on task" as barriers to being able to spend a great deal of time on vocabulary. Despite the challenges, all four teachers responded that they would probably use the explicit mathematics vocabulary instructional practices again in future lessons.

Chapter 5: Discussion

Deep understanding of mathematical concepts and procedures is hindered by a lack of mathematics vocabulary knowledge (Capraro & Joffrion, 2006; Hughes, Powell, & Lee, 2018; Schleppegrell, 2007). By the time students reach the middle grades, they are expected to know a huge amount of mathematics vocabulary, yet middle school teachers – including the ones in the present study – often find themselves having to re-teach not only forgotten concepts and procedures, but also vocabulary (Powell et al., 2017). Adding to this challenge is the fact that secondary mathematics teachers are often not prepared to teach literacy skills, including vocabulary (Fagella-Luby et al., 2009; Powell et al., 2017; Siuty et al., 2018). Thus, addressing the need to support middle school mathematics teachers' vocabulary instruction for students with disabilities or who are otherwise struggling in mathematics classes was the focus of the current study.

This study represents a step in a new direction for the Content Acquisition Podcasts for Professional Development process and the cognitive apprenticeship framework in which it is rooted. Previously, empirical CAP-PD studies have focused on middle school science teachers focused on teaching content-area vocabulary (Kennedy, Rodgers et al., 2017; Kennedy et al., 2018) and English teachers focused on writing instruction (Romig, 2018). This study builds on its predecessors by bringing the CAP-PD approach to middle school mathematics teachers. This study also builds on the growing CAP-PD research base by specifically examining the role of the performance feedback and coaching component of the CAP-PD process.

In this study, four sixth-grade mathematics teachers who taught students with IEPs in their classes were provided with access to Content Acquisition Podcasts with embedded Teacher Video (CAP-TVs; i.e., multimedia instruction) and Content Acquisition Podcasts Teacher Slides (CAP-TS; i.e., a PowerPoint template to facilitate the implementation of explicit vocabulary instruction). Each of the CAP-TVs focused on a distinct teaching practice in a group of practices called explicit mathematics vocabulary instruction. Following a randomized, staggered entry from a baseline observation phase to intervention (i.e., multiple baseline), the teachers were given these materials and asked to watch the videos. They began receiving the performance feedback and coaching, specifically directed at their implementation of the explicit mathematics vocabulary instructional practices. Each teacher received this intervention on three consecutive observed lessons, followed by three consecutive observations where no feedback or coaching was provided. This alternating treatment design element continued throughout an extended intervention phase.

Teacher observations were conducted with the Classroom Teaching Scan (CT Scan) instrument. With regard to the teachers' use of explicit mathematics vocabulary instruction, three types of data were collected and/or calculated in each observation: the Quality Vocabulary Index (QVI) statistic, the level of implementation fidelity for the instructional practices (expressed as a percentage), and the proportion of class time that teachers spent using the instructional practices (also expressed as a percentage). In this chapter, I discuss the results from this study as well as several implications for future research and for teacher professional development and practice.

Reflection on Results

This study found that there was a functional relationship between the performance feedback and coaching intervention and the teachers' use of explicit mathematics vocabulary instructional practices. Visual analysis of each of the four teachers' performance reveals that at least a weak relationship exists at the individual (within-subject) level, and that across subjects, there is also evidence of a weak to moderate functional relationship. These results indicate that it is possible that the performance feedback and coaching component of the CAP-PD system makes a significant contribution to changes in teachers' explicit vocabulary instruction. However, the trend in those changes is minimal, mitigating the effectiveness of the intervention.

The primary experimental variable in the study was the Quality Vocabulary Index (QVI). The calculation of a QVI score for a given lesson takes into account the level of implementation fidelity (i.e., how many implementation markers the teacher used per iteration of a vocabulary practice) and the duration of the vocabulary instruction. In the previous chapter, I presented results for this variable as well as data regarding these two components of the QVI. A closer look at these components here provides some additional context for the study's findings.

Implementation Markers Used in Explicit Vocabulary Instruction

In the CT Scan observation tool, there are two categories of vocabulary practices that an observer can use and record. First is a broad explicit vocabulary instruction menu of practices such as "Asks student(s) to state definition," "Connection made to other content," and "Demonstration." An additional vocabulary menu was added specifically for this study so that the explicit mathematics vocabulary instructional practices could be recorded. Those six practices and their implementation markers are presented in Table 2.1.

Results at both the within-subjects level and between subjects showed that when explicit mathematics vocabulary instruction was used, teachers' implementation varied considerably. For example, Sophie often spent at least a few minutes on vocabulary instruction of some kind (not always explicit mathematics vocabulary instruction) during the beginning of the class period, but the number of implementation markers (IMs) that she used each time ranged from 43% to 92%. She did not implement any vocabulary instruction with 100% fidelity. Across all four teachers, there is a similar pattern of (a) infrequent vocabulary instruction and (b) highly variable implementation when vocabulary instruction was used.

It is possible that certain vocabulary practices were easier to implement with greater fidelity than others. Returning to the example of Sophie's vocabulary instruction, one of the practices she used most often was "Asks student(s) to state definition," which is not an explicit mathematics vocabulary instructional practice and has only 4 implementation markers. In contrast, the most frequently used practice across all teachers from the explicit mathematics vocabulary instruction menu was "Use a student-friendly definition," which has 7 implementation markers. The performance feedback and coaching that teachers received included a broad look at the lesson (e.g., frequency counts of opportunities to respond, feedback statements, and behavior redirect statements) but focused more on the presence or lack, as well as quality, of their vocabulary instruction. When teachers implemented fewer than 100% of the IMs for a vocabulary practice, the written coaching ("One Big Thing") encouraged them to continue using the practice and to focus on including all of the markers, usually naming the specific markers that were missed.

There is little indication in the data on implementation fidelity that this specific coaching spurred the teachers to (a) implement the practice more frequently or (b) increase the number of IMs used. The social validity survey provides some more information about this result. All four teachers stated that they thought the performance feedback was clear, of high quality, and potentially helpful, but some also expressed some reservations about it. For Ada, time constraints and competing demands impeded her from paying close attention to her feedback or using it to implement more vocabulary instruction. She also indicated on the survey that she preferred the broader data in the feedback (i.e., frequency counts) than the written sections (i.e., "The Big Picture" or

"One Big Thing"). Emilie felt that the feedback she received wasn't always necessary or relevant, as she and her ESOL co-teacher often decided together to prioritize other types of instruction.

Time Spent on Explicit Vocabulary Instruction

One of the most important outcomes of the present study is that although it varied in quality, all four teachers went from spending no time at all on explicit vocabulary instruction to spending at least a few minutes on it, at least 1-2 times a week. The only exception to this was Ada, whose vocabulary instruction dropped off about halfway through the study; her reasons for this are discussed later in this chapter. Collecting and examining this type of data was important to this study because research on the vocabulary learning of students with disabilities has shown that devoting more instructional time for direct, explicit vocabulary instruction is essential in improving academic outcomes for these students (Bryant et al., 2003; Jitendra et al., 2004; Kennedy et al., 2015).

In secondary mathematics classes it is especially rare to see instructional time devoted to direct or explicit vocabulary instruction (Morin & Franks, 2009; Pierce & Fontaine, 2009; Riccomini et al., 2015). The current study was no different. The teachers in the current study responded positively to the ways in which explicit mathematics vocabulary instruction might support their students who were struggling with mathematical concepts and procedures, but this support was not indicated by their actual teaching behaviors. In short, teachers saw the value of the explicit instruction approach to teaching vocabulary, but found it difficult to prioritize it over what they felt were more high-leverage instructional choices.

There are a number of possible reasons for this occurrence, many of which are beyond the scope of the current analysis, but the social validity survey responses reveal some insights for why this might be the case. The main reasons that teachers gave for not using the explicit mathematics vocabulary instructional practices mostly had to do with a perceived need to adhere to curriculum and pacing (and, consequently, a need to prepare students for end-of-year achievement testing) and current expectations of the school- and district-level instructional leadership (i.e., a focus on "inquiry-based" instruction). For example, Florence explained, "Our primary emphasis will continue to be centered on developing math concepts through discovery but this study helped me recognize a need for explicit vocabulary instruction." Interestingly, in Emilie's case, her co-teacher did most of the vocabulary instruction, and Emilie noted that as a result of this, "there was more vocabulary being taught than the study reflected." In contrast to the results of the study, however, all four teachers also indicated that they felt that explicit mathematics vocabulary instruction would be either moderately or very helpful for their students.

Limitations

The results of the present study must be viewed in light of some important limitations. In terms of methodology, this study lacks both a reliable secondary dependent measure and a measure of student outcomes. In the absence of a secondary dependent measure to provide additional context for the quality of the mathematics instruction. I originally planned to include the Mathematics Scan observation instrument (M-Scan; Berry, Rimm-Kaufman, Ottmar, Walkowiak, Merritt, & Pinter, 2013) alongside the CT Scan. However, both time and logistical constraints prevented me from completing the required rater training. The rater training for the M-Scan is intensive and thorough, appropriate for a measure so robust. In addition, the M-Scan is designed for use with video recorded lessons, so that raters can score 30-minute segments of instruction ("soft codes") before assigning ratings for an entire lesson. The present study prioritized live observations in the teachers' classrooms, which would have been a considerable challenge for a second M-Scan observer unaccustomed to using the instrument in live teaching settings. Using the M-Scan during live observations also would have proved less reliable, compared to scoring recorded lessons, as it is intended to be used.

A second methodological limitation of the present study is the absence of a measure of student outcomes. In the literature on teacher professional development, student outcomes are essential to determining its effectiveness (Borko, 2004; Penuel et al., 2007; Yoon et al., 2007). A recent study by Hughes, Powell, and Lee (2019) developed and evaluated a vocabulary assessment specifically for middle grades students. Their aim in developing the measure was to provide a measure of vocabulary knowledge in mathematics that could help teachers identify areas of need and then target those needs with focused instruction. Future studies that seek to expand on the present study should make measuring student learning outcomes a priority. In order to meaningfully contribute to the limited research on the effects explicit vocabulary instruction in mathematics for students with disabilities, future studies must measure those effects in terms of changes in students' mathematics vocabulary knowledge as a result of that instruction.

Threats to Internal Validity

There are two sources of threats to the internal validity of the present study that should be addressed. First, two threats originate from the observer: Observer drift, and observer bias. More than 100 live observations were conducted over the course of this study, and I was the primary observer in each of them. In addition, I was the only researcher involved in data analysis. Although inter-rater reliability was high, only 15% of the observations were double coded by a second observer (also conducted live). Future replications or other future studies should address this threat by having at least one additional primary observer who is also involved in compiling and analyzing the data collected, and by double coding at least 20% of the total observations.

Attrition was another threat to the internal validity of the present study's results. As noted previously, one of the teachers, Florence, opted out of the study before the maintenance phase could be initiated. There were enough data points to justify including her in the final results of the study, but the absence of a maintenance phase further limits the internal validity of her results specifically. Although she did complete all phases of the study, there was a small attrition effect in Ada's results as well. From Observation #14 onward, she did not teach vocabulary at all, with just one exception of about a minute of vocabulary instruction on Observation #17. She later expressed that she did not feel that most of the explicit mathematics vocabulary instructional practices were a good fit for her preferred teaching style, and so she stopped trying to include them.

Implications for Future Research

The results from the present study point to a number of future research areas. In the sections that follow, I discuss some of these areas for future research in terms of (a) the study's dependent measure, the CT Scan instrument; (b) the study's primary dependent variable, the Quality Vocabulary Index (QVI); (c) the social validity of the performance feedback and coaching intervention; and (d) expanding upon what we know about how students might learn vocabulary through explicit instruction.

CT Scan

The CT Scan has been used in only one other single-case experimental study to date (i.e., Kennedy, Rodgers, Romig, Lloyd, & Brownell, 2017). With the exception of Kennedy et al. (2017) and the current study, the empirical studies that used the CT Scan as a dependent measure have focused on group designs and/or randomized control trials, and have focused on secondary teachers (particularly science). While the instrument is quite robust and has some evidence of reliability and validity, it is still somewhat untested in many settings and with a variety of observers. Additionally, the practices in the explicit mathematics vocabulary instruction "menu" and their implementation markers were developed for this study. While they were carefully researched and then briefly reviewed by one expert, they need further refinement with input from additional experts, including classroom teachers, and further evidence of their validity needs to be collected.

Quality Vocabulary Index

The QVI statistic is essentially a function of implementation fidelity and duration of vocabulary instruction; however, a notable limitation of its use in this study is its lack of sensitivity to very short durations of vocabulary instruction. In their introduction of the QVI statistic, Kennedy et al. (2018) acknowledge this, but it is particularly evident in the present study. The longest iteration of an explicit mathematics vocabulary practice in this study was 16 minutes 53 seconds (Florence, Observation #9). In addition, Florence implemented the instructional practices in that iteration with 85% fidelity, but because the lesson itself was over 80 minutes long, the resulting QVI was only .42. On the other hand, if the class periods in the current study were not on a block schedule and were therefore considerably shorter (e.g., 50 minutes rather than 80), QVI results would look markedly different and could lead to more encouraging interpretations. Future studies that use the QVI statistic as a dependent variable could attempt to control for the large amount of class time with no vocabulary instruction, as it appears to be a major influence on how the quality of teachers' vocabulary instruction is interpreted. As an example, if vocabulary instruction tends to take place earlier in longer, block scheduled class periods, the duration variable in the QVI calculation could be standardized to 20 or 30 minutes for all teachers in the study, shifting the statistical emphasis from duration to implementation and frequency.

However, despite its lack of measurement sensitivity in this study, this does reveal some implications for future research using the CT Scan instrument and especially the QVI statistic. First, at least as Kennedy et al. (2018) and many others define it, more vocabulary instruction is better than minimal or no vocabulary instruction. That was an underlying assumption of the present study as well, and this study was able to achieve that. Future studies would need to operate under a similar assumption. Second, because many secondary teachers, including the ones in the present study, are likely to favor other types of content-based instruction over vocabulary instruction (i.e., mathematical procedures), the calculation and/or interpretation of a teacher's QVI score may need to be adjusted in those cases where vocabulary instruction is present and of high-quality, but is limited to a few minutes of class time. This is especially important if the QVI is to be used in coaching scenarios or in studies where the QVI is reported and tracked as performance over time.

Third, the QVI is a very new statistic, with just a handful of empirical studies that have used it as a dependent variable, and so performance standards (i.e., what defines a high or low, "good" or "bad" QVI score) have yet to be established. In the present study, the QVI is listed on the CT Scan output (i.e., the performance feedback that teachers received), but was not mentioned or discussed in the written coaching. Recent updates to the CT Scan are aimed at refining another performance feedback output that allows researchers or coaches to make comparisons to a single teacher's past performance or between a teacher's performance and that of other teachers in similar settings. This is potentially a powerful addition to the coaching component of the CAP-PD system, but a good deal more research is needed in order to explore that potential.

Social Validity

Results from the social validity survey also point to some areas where future research is needed. First, some of the teachers expressed some concern about the amount and timing of the feedback, as well as the time it took to thoroughly take in the data and the coaching. On the other hand, Sophie indicated on the survey that she spent a lot of time reviewing her performance feedback and coaching materials (i.e., at least once every time she received it), and had arguably the most encouraging changes to her explicit vocabulary instruction according to measures of effect size. Future CAP-PD studies should investigate what dosage and/or frequency of the performance feedback and coaching is most impactful on changing teacher practice without overwhelming the teachers to the point of saturation. Understanding this is key to refining the CAP-PD process into an optimal professional development package for secondary teachers.

Second, three of the four teachers expressed a desire for more support in understanding and acting upon the performance feedback. A video guide to "reading" the CT Scan outputs was provided for them, but it was clear in the social validity results that a more individualized experience was desirable. Teachers expressed this concern as both a coaching issue (i.e., wanting in-person coaching) and a logistical issue, as they were not allowed to discuss their feedback with each other and this restriction is not reflective of actual teacher practice. Ada especially noted some frustration with not being able to discuss her feedback or options for vocabulary instruction when planning with the other sixth grade mathematics teachers. She added that she "found it hard to motivate [herself] to incorporate some of those practices because I had to do it myself."

Explicit Vocabulary Instruction in Mathematics

With a handful of exceptions, most of the existing research that supports the use of an explicit instruction approach to teaching vocabulary has come from studies of students in primary grades or even pre-kindergarten (e.g., Biemiller & Boote, 2006; Coyne, Simmons, Kame'enui, & Stoolmiller, 2004; Nash & Snowling, 2006). The gap in this research literature is considerable when it comes to students in secondary grades, with or without disabilities. The same is true for secondary content areas such as mathematics. While a strictly explicit instruction approach to disciplinary literacy may not always be appropriate in the middle and secondary grades (e.g., Faggella-Luby, Graner, Deshler, & Drew, 2012), more research is needed in order to pinpoint what effective vocabulary instruction looks like in those secondary content area settings, as well as what content area teachers need from professional development to support that instruction.

Another question about the role of explicit vocabulary instruction in content areas such as mathematics is how much time teachers should spend on these practices in order to make an impact on students' learning. The research literature on explicit vocabulary instruction has not pinpointed or endorsed a specific, "ideal" amount of time that teachers should spend on teaching vocabulary. While it may be impossible to determine that ideal number, more research is needed that explores the impact of time spent explicitly teaching vocabulary on a variety of student learning outcomes. This research is especially needed at the secondary level, where mathematics teachers such as the ones in this study often encounter students with limited background knowledge and other obstacles (such as disabilities in reading or limited English proficiency) that can hinder mathematics learning.

A final suggestion for future research in the area of content-area vocabulary instruction (especially mathematics) is to examine the reasons behind why teachers may or may not incorporate vocabulary or other literacy-oriented instruction into their existing content. In the current study, changes in teachers' use of explicit mathematics vocabulary instruction were moderate, but no data was collected that could explain why their implementation remained at a low to medium level as far as fidelity and duration. Exploring the instructional decision-making processes of secondary mathematics teachers, especially those who teach students with disabilities, could potentially provide valuable information about (a) what is missing from mathematics instruction that could support the academic growth of students with disabilities as well as others who struggle with language and literacy, and (b) what might be the most effective ways to provide training and professional development for those teachers to include literacy-oriented instruction alongside and integrated with mathematics content.

Implications for Professional Development and Practice

This study also has important implications for teacher professional development and practice. In this section I discuss some considerations for applying these results to professional development and practice, and some avenues for future research in those areas.

Performance Feedback and Coaching

The present study demonstrated that the sixth grade mathematics teachers did respond to receiving performance feedback and coaching on their vocabulary instruction, following some an online professional development (CAP-TVs and CAP-TS). By their own description in the demographic survey at the start of the study, none of the teachers were previously teaching vocabulary explicitly in their classes. All of them incorporated some explicit mathematics vocabulary instruction during the study following the intervention, and all but one had at least one replication of this effect.

Research in teacher professional development has already shown that this kind of feedback can be highly effective (e.g., Solomon et al., 2012; Sweigart et al., 2015). The present study contributes to that body of research literature; however, there are a number of avenues for future research. First, it is unclear what amount or frequency (i.e., dosage) of each component is optimal. Second, with regard specifically to the CT Scan and CAP-PD, more research is needed to determine what types of support teachers need in order to maximize the desired outcomes (i.e., increased use of evidence-based practice, increased fidelity of implementation, improved student outcomes, etc). The level of support teachers received in the present study was, according to the teachers themselves, not ideal.

The results of this study also point to a need for more intensive training for secondary mathematics teachers in the area of vocabulary instruction. For the teachers in the present study, this was the first time they had been asked to teach vocabulary explicitly. Although they did receive some training and professional development via CAP-TVs, their response to the performance feedback and coaching (which was geared towards their vocabulary instruction specifically) was minimal. It is likely that there was a considerable gap between their knowledge of the practices, gained from the CAP-TVs, and their confidence or ability to apply that knowledge. Future research should explore this gap and ways to remedy it, given that vocabulary knowledge is so essential to students with disabilities' acquisition of academic language in mathematics, and subsequently, their deep understanding of mathematical concepts.

CAP-PD and the Cognitive Apprenticeship Framework

The current study extends a line of research that utilizes the core concepts of the cognitive apprenticeship framework to address teacher professional development. The results of the current study point to the effectiveness of personalized performance feedback and coaching as a means of changing teacher practice with regard to explicit vocabulary instruction. The inclusion of the performance feedback and coaching component in the CAP-PD process is one of its most important connections to the cognitive apprenticeship framework. Generally, teachers in the current study responded positively about their experience with the CAP-PD approach training them in the use of explicit mathematics vocabulary instruction. The empirical results of the current study as well as the social validity results regarding the performance feedback and coaching component of the process indicate some areas for future CAP-PD development and research.

In addition, the current study is the first to investigate the effect of the CAP-PD approach in mathematics classrooms. As a result, the materials designed for the study were essentially prototypes, and need further review and refinement. For example, Emilie noted that she "would have liked to see real examples from real teachers" in the CAP-TVs, adding that "the examples in the videos did not fit the reality of the classroom environment." The most complete version of the CAP-PD system to date is for middle school science teachers, and the CAP-TVs for that population of teachers underwent a number of prototypical versions before including "real teachers" in its current iteration.

Emilie's concern is valid, and future work on the CAP-PD approach should consider developing the system for use with middle school mathematics teachers as well as science teachers. In addition to refining the CAP-TVs, future CAP-PD research with middle school mathematics teachers should include some key improvements to experimental design in order to address limitations in the present study. In future studies, more teachers should be included in the study population, for two main reasons. First, a group design would allow the option of having observers who are blind to the condition of each teacher. This would strengthen the generalizability of results. Second, a group design could address the current study's misalignment of treatment conditions with the actual, sociological aspects of teacher practice and instructional planning. A group design would allow all teachers at the same school to receive the intervention at the same time, and allow them to discuss and plan vocabulary instruction together, while a control group would be located at another school entirely. This would address the concern that Ada expressed in not being able to discuss her feedback with colleagues when planning her instruction. It is possible that she, as well as her colleagues, would have included more explicit mathematics vocabulary instruction had they been allowed to work and plan together using their performance feedback as a guide.

Conclusion

Explicit vocabulary instruction in mathematics has potential to support and improve the mathematical conceptual and procedural understanding of students with disabilities. Students with disabilities often struggle with reading and literacy skills including vocabulary, and require support in content-area classes where vocabulary can be challenging. In general, middle school mathematics teachers are not provided with training in literacy instruction. This study aimed to address this need by examining the effect of the performance feedback and coaching component of the Content Acquisition Podcasts for Professional Development (CAP-PD) system on middle school mathematics teachers' use of explicit mathematics vocabulary instruction. Results of the study indicate that performance feedback and coaching had moderately positive effects on teachers' use and quality of explicit mathematics vocabulary instruction. This study contributes to a growing body of research that utilizes a cognitive apprenticeship framework to address teacher professional development, especially teachers of students with disabilities. Additional single-case replications of this study are needed in order to further refine and generalize these results. The results of this study also show a need for future research to examine the additional supports that mathematics teachers may need in order to use explicit vocabulary instruction more thoroughly and confidently.

References

Alexander, P. A., Schallert, D. L., & Hare, V. C. (1991). Coming to terms: How researchers in learning and literacy talk about knowledge. *Review of Educational Research*, *61*, 315–343.

Alves, K. D., Kennedy, M. J., Kellems, R. O., Wexler, J., Rodgers, W. J., Romig, J. E., & Peeples, K. N. (2017). Improving preservice teacher vocabulary instruction: A randomized control trial. *Teacher Education and Special Education*. https://doi.org/10.1177/08884064177270

Archer, A. L., & Hughes, C. A. (2011). *Explicit instruction.* New York, NY: Guilford Press.

Ball, D. L., & Cohen, D. K. (1999). Developing practices, developing practitioners: Toward a practice-based theory of professional development. In G. Sykes & L. Darling-Hammond (Eds.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3–32). San Francisco, CA: Jossey-Bass.

Barton, E. A., Whittaker, J. V., Kinzie, M. B., DeCoster, J., & Furnani, E. (2017). Understanding the relationship between teachers' use of online demonstration videos and fidelity of implementation in MyTeachingPartner-Math/Science. *Teaching and Teacher Education*, *67*, 189–201. https://doi.org/10.1016/j.tate.2017.06.011

Baumann, J.F., Edwards, E.C., Font, G., Tereshinski, C.A., Kame'enui, E.J., & Olejnik, S. (2002). Teaching morphemic and contextual analysis to fifth-grade students. *Reading Research Quarterly, 37*(2), 150–176.

Baumann, J. F., Edwards, E. C., Boland, E., & Font, G. (2012). Teaching word-learning strategies. In E. J. Kame'enui & J. F. Baumann (Eds.), *Vocabulary instruction: Research to practice* (2nd ed., pp. 139–166). New York, NY: Guilford Press.

Baumann, J. F., & Graves, M. F. (2010). What is academic vocabulary? *Journal of Adolescent & Adult Literacy*, *54*, 4–12. https://doi.org/10.1598/JAAL.54.1.1

Bay-Williams, J. M., & Livers, S. D. (2009). Supporting MATH vocabulary acquisition. *Teaching Children Mathematics*, *16*, 238–245.

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction* (2nd ed.). New York, NY: Guilford Press.

Benedict, A., Holdheide, L. R., Brownell, M. T., & Foley, A. M. (2016). *Learning to teach: Practice-based preparation in teacher education* (Special issues brief). Center on Great Teachers and Leaders.

Berry, III, R. Q., Rimm-Kaufman, S. E., Ottmar, E. M., Walkowiak, T. A., Merritt, E., & Pinter, H. H. (2017). The Mathematics Scan (M-Scan): A Measure of Mathematics Instructional Quality. Unpublished measure, University of Virginia.

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in the primary grades. *Journal of Educational Psychology*, *98*, 44–62.

Boardman, A. G., Roberts, G., Vaughn, S., Wexler, J., Murray, C. S., & Kosanovich, M. (2008). *Effective instruction for adolescent struggling readers: A practice brief*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher*, *33*(8), 3–15.

Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and Teacher Education*, *24*, 417–436. https://doi.org/10.1016/j.tate.2006.11.012

Borko, H., Koellner, K., Jacobs, J., & Seago, N. (2011). Using video representations of teaching in practice-based professional development programs. *ZDM Mathematics Education*, *43*, 175–187. https://doi.org/10.1007/s11858-010-0302-5

Borko, H., Mayfield, V., Marion, S., Flexer, R., & Cumbo, K. (1997). Teachers' developing ideas and practices about mathematics performance assessment: Successes, stumbling blocks, and implications for professional development. *Teaching and Teacher Education*, *13*(3), 259–278.

Bos, C. S., & Anders, P. L. (1990). Effects of interactive vocabulary instruction on the vocabulary learning and reading comprehension of junior-high learning disabled students. *Learning Disability Quarterly*, *13*, 31–42. https://doi.org/10.2307/1510390

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42.

Bryant, D. P., Goodwin, M., Bryant, B. R., & Higgins, K. (2003). Vocabulary instruction for students with learning disabilities: A review of the research. *Learning Disability Quarterly*, *26*, 117–128.

Cannon, M. D., & Witherspoon, R. (2005). Actionable feedback: Unlocking the power of learning and performance improvement. *Academy of Management Executive*, *19*(2), 120–134. https://doi.org/10.5465/AME.2005.16965107

Capraro, M. M., & Joffrion, H. (2006). Algebraic equations: Can middle-school students meaningfully translate from words to mathematical symbols? *Reading Psychology*, *27*, 147–164. https://doi.org/10.1080/02702710600642467

Capraro, R. M., Capraro, M. M., & Rupley, W. H. (2010). Semantics and syntax: A theoretical model for how students may build mathematical misunderstandings. *Journal of Mathematics Education*, *3*(2), 58–66.

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

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

Cook, S. C., Collins, L. W., Morin, L. L., & Riccomini, P. J. (2019). Schema-based instruction for mathematical word problem solving: An evidence-based review for students with learning disabilities. Learning Disability Quarterly. Advance online publication. https://doi.org/ 10.1177/0731948718823080

Coyne, M. D., Simmons, D., Kame'enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality*, *12*, 145–162.

Cummins, J. (2000). Academic language learning, transformative pedagogy, and information technology: Towards a critical balance. *TESOL Quarterly*, *34*, 537–548.

Cunningham, A. E., Zibulsky, J., Stanovich, K. E., & Stanovich, P. J. (2009). How teachers would spend their time teaching language arts: The mismatch between self-reported and best practices. *Journal of Learning Disabilities*, *42*, 418–430. https://doi.org/10.1177/0022219409339063

Darling-Hammond, L. (2014). Strengthening clinical preparation: The holy grail of teacher education. *Peabody Journal of Education*, *89*, 547–561. https://doi.org/10.1080/0161956X.2014.939009

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective Teacher Professional Development*. Palo Alto, CA: Learning Policy Institute. Retrieved from https://learningpolicyinstitute.org/product/teacher-prof-dev

Darling-Hammond, L., Wei, R. C., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher* *development in the United States and abroad*. Palo Alto, CA: National Staff Development Council.

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.

Dennen, V. P., & Burner, K. J. (2008). The cognitive apprenticeship model in educational practice. In *Handbook of research on educational communications and technology* (pp. 425–439).

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*, 181-199. doi:10.3102/0013189x09331140

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

Dingle, M. P., Brownell, M. T., Leko, M. M., Boardman, A. G., & Haager, D. (2011). Developing effective special education reading teachers: The influence of professional development, context, and individual qualities. *Learning Disability Quarterly*, *34*, 87–103.

Doabler, C. T., Fien, H., Nelson-Walker, N. J., & Baker, S. K. (2012). Evaluating three elementary mathematics programs for presence of eight research-based instructional design principles. *Learning Disability Quarterly*, *35*, 200–211. https://doi.org/10.1177/0731948712438557

Duchaine, E. L., Jolivette, K., & Fredrick, L. D. (2011). The effect of teacher coaching with performance feedback on behavior-specific praise in inclusion classrooms. *Education and Treatment of Children*, *34*(2), 209–227.

Ebbers, S. M., & Denton, C. A. (2008). A root awakening: Vocabulary instruction for older students with reading difficulties. *Learning Disabilities Research & Practice*, *23*, 90–102. https://doi.org/10.1111/j.1540-5826.2008.00267.x

Ely, E., Kennedy, M. J., Pullen, P. C., 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*, 25–43.

Faggella-Luby, M. N., Graner, P. S., Deshler, D. D., & Drew, S. V. (2012). Building a house on sand: Why disciplinary literacy is not sufficient to replace general strategies for adolescent learners who struggle. *Topics in Language Disorders*, *32*, 69–84. https://doi.org/10.1097/TLD.0b013e318245618e

Faggella-Luby, M. N., Ware, S. M., & Capozzoli, A. (2009). Adolescent literacy - Reviewing adolescent literacy reports: Key components and critical questions. *Journal of Literacy Research*, *41*, 453–475. https://doi.org/10.1080/10862960903340199

Fallon, L. M., Collier-Meek, M. A., Maggin, D. M., Sanetti, L. M. H., & Johnson, A. H. (2015). Is performance feedback for educators an evidence-based practice? A systematic review and evaluation based on single-case research. *Exceptional Children*, *81*(2), 227–246. https://doi.org/10.1177/0014402914551738

Fisher, D., & Frey, N. (2008). *Word wise and content rich, grades 7-12: Five essential steps to teaching academic vocabulary*. Portsmouth, NH: Heinemann.

Ford-Connors, E., & Paratore, J. R. (2015). Vocabulary instruction in fifth grade and beyond: Sources of word learning and productive contexts for development. *Review of Educational Research*, *85*(1), 50–91.

Franklin, R. D., Gorman, B. S., Beasley, T. M., & Allison, D. B. (1996). Graphic display and visual analysis. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 119–158). Mahwah, NJ: Lawrence Earlbaum Associates.

Frayer, D., Frederick, W. C., and Klausmeier, H. J. (1969). *A Schema for Testing the Level of Cognitive Mastery*. Madison, WI: Wisconsin Center for Education Research.

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. S. (2009). *Assisting students struggling with mathematics: Response to Intervention (RtI) for elementary and middle schools* (No. NCEE 2009-4060). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.gov/ncee/wwc/publications/practiceguides

Graves, M. F. (2006). *The vocabulary book: Learning and instruction*. New York, NY: Teacher's College Press.

Greenwood, S. C. (2002). Making words matter: Vocabulary study in the content areas. *The Clearing House*, *75*, 258–263. https://doi.org/10.1080/00098650209603951

Grossman, P., Wineburg, S., & Woolworth, S. (2001). Toward a theory of teacher community. *Teachers College Record*, *103*, 942–1012.

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

Harris, M. L., Schumaker, J. B., & Deshler, D. D. (2011). The effects of strategis morphological analysis instruction on the vocabulary performance of secondary students with and without disabilities. *Learning Disability Quarterly*, *34*, 17–33.

Hattie, J., Fisher, D., & Frey, N. (2017). *Visible learning for mathematics*. Thousand Oaks, CA: Corwin.

Hebert, M. A., & Powell, S. R. (2016). Examining fourth-grade mathematics writing: Features of organization, mathematics vocabulary, and mathematical representations. *Reading and Writing*, *29*, 1511–1537. https://doi.org/10.1007/s11145-016-9649-5

Hill, H. C., Beisiegel, M., & Jacob, R. (2013). Professional development research: Consensus, crossroads, and challenges. *Educational Researcher*, *42*, 476–487. https://doi.org/10.3102/0013189X13512674

Hirsch, S. E. (2016). *Professional development in practice: Improving novice teachers' use of evidence-based classroom management practices* (Unpublished doctoral dissertation). University of Virginia, Charlottesville, VA. Retrieved from Libra.

Holcombe, A., Wolery, M., & Gast, D. L. (1994). Comparative single-subject research: Description of designs and discussion of problems. *Topics in Early Childhood Special Education*, *14*, 119–145.

Holdheide, L. R., & Reschly, D. J. (2008). *Teacher preparation to deliver inclusive services to students with disabilities* (No. TQ Connection Issue Paper). Washington, DC: National Comprehensive Center for Teacher Quality. Retrieved from http://www.tqsource.org/publications/TeacherPreparationtoDeliverInclusiveServices.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*(2), 165–179.

Hudson, P., Miller, S. P., & Butler, F. (2006). Adapting and merging explicit instruction within reform based mathematics classrooms. *American Secondary Education*, *35*, 19–32.

Hughes, C. A., Morris, J. A., Therrien, W. J., & Benson, S. K. (2017). Explicit instruction: Historical and contemporary contexts. *Learning Disabilities Research & Practice*, *32*, 140–148. https://doi.org/10.1111/ldrp.12142

Hughes, E. M., Powell, S. R., & Lee, J. (2018). Development and psychometric report of a middle-school mathematics vocabulary measure. *Assessment for Effective Intervention*. Advance online publication. https://doi.org/10.1177/1534508418820116

Jitendra, A. K., Edwards, L. L., Sacks, G., & Jacobson, L. A. (2004). What research says about vocabulary instruction for students with learning disabilities. *Exceptional Children*, *70*, 299–322. https://doi.org/10.1177/001440290407000303

Kame'enui, E. J., & Baumann, J. F. (2012). *Vocabulary instruction: Research to practice* (2nd ed.). New York, NY: Guilford Press.

Kazdin, A. E. (2011). Multiple-baseline designs. In *Single-case research designs: Methods for clinical and applied settings* (2nd ed., pp. 144–166). New York, NY: Oxford University Press.

Kennedy, M. J., Deshler, D. D., & Lloyd, J. W. (2015). Effects of multimedia vocabulary instruction on adolescents with learning disabilities. *Journal of Learning Disabilities*, *48*, 22–38. https://doi.org/10.1177/0022219413487406

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. https://doi.org/10.1016/j.tate.2016.12.009

Kennedy, M. J., Rodgers, W. J., Romig, J. E., Lloyd, J. W., & Brownell, M. T. (2017). Effects of a multimedia professional development package on inclusive science teachers' vocabulary instruction. *Journal of Teacher Education*, *68*(2), 213–230. https://doi.org/10.1177/0022487116687555

Kennedy, M. J., Rodgers, W. J., Romig, J. E., Mathews, H. M., & Peeples, K. N. (2018). Introducing the content acquisition podcasts for professional development (CAP-PD) process: Supporting vocabulary instruction for inclusive middle school science teachers. *Teacher Education and Special Education*, *41*, 140–157. https://doi.org/10.1177/0888406417745655

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. https://doi.org/10.106/j.compedu.2013.08.010

Klingner, J. K. (2004). The science of professional development. *Journal of Learning Disabilities*, *37*, 248–255.

Koellner, K., Jacobs, J., Borko, H., Schneider, C., Pittman, M. E., Eiteljorg, E., … Frykholm, J. (2007). The problem-solving cycle: A model to support the development of teachers' professional knowledge. *Mathematical Thinking and Learning*, *9*, 273–303. https://doi.org/10.1080/10986060701360944

Kostos, K., & Shin, E. (2010). Using math journals to enhance second graders' communication of mathematical thinking. *Early Childhood Education Journal*, *38*, 223–231. https://doi.org/10.1007/s10643-010-0390-4

Kraft, M. A., & Blazar, D. (2017). Individualized coaching to improve teacher practice across grades and subjects: New experimental evidence. *Educational Policy*, *31*(7), 1033–1068. https://doi.org/10.1177/0895904816631099

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

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education*, *34*(1), 26–38. https://doi.org/10.1177/0741932512452794

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45*, 1058-1072. https://doi.org/10.3758/s1342-013-0334-1

Lesaux, N. K., Kieffer, M. J., Faller, S. E., & Kelley, J. G. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading & Writing Quarterly*, *45*, 196–228. https://doi.org/10.1598/RRQ.45.2.3

Maheady, L., Smith, C., & Jabot, M. (2014). Field experiences and instructional pedagogies in teacher education: What we know, don't know, and must learn soon. In P. T. Sindelar, E. D. McCray, M. T. Brownell, & B. Lignugaris/Kraft (Eds.), *Handbook of research on special education teacher preparation* (pp. 161–177). New York, NY: Routledge.

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, *63*, 760–769.

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

McAdams, L. A. (2011). The effect of direct instruction with math and science content area vocabulary on student achievement. *The California Reader*, *45*(2), 17–21.

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. https://doi.org/10.1177/0731948714564576

McLaughlin, D.H., Bandeira de Mello, V., Blankenship, C., Chaney, K., Esra, P., Hikawa, H., Rojas, D., William, P., and Wolman, M. (2008). *Comparison Between NAEP and State Mathematics Assessment Results: 2003* (NCES 2008-

475). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Montague, M., Enders, C., & Dietz, S. (2011). Effects of cognitive strategy instruction on math problem solving of middle school students with learning disabilities. *Learning Disability Quarterly*, *34*, 262–272. https://doi.org/10.1177/0731948711421762

Morin, J. E., & Franks, D. J. (2009). Why do some children have difficulty learning mathematics? Looking at language for answers. *Preventing School Failure*, *54*, 111–118. https://doi.org/10.1080/10459880903217861

Myers, D. M., Simonsen, B., & Sugai, G. (2011). Increasing teachers' use of praise with a response-to-intervention approach. *Education and Treatment of Children*, *34*(1), 35–59.

Nagy, W. E., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly, 47*, 91-108.

Nash, H., & Snowling, M. (2006). Teaching new words to children with poor existing vocabulary knowledge: A controlled evaluation of the definition and context methods. *International Journal of Language and Communication Disorders*, *41*, 335–354.

National Center for Education Statistics. (2015). The nation's report card: Mathematics and reading assessments 2015 (NCES 2015-136). Washington, D.C.: Institute of

Education Sciences, U.S. Department of Education. Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2015136

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 Center for Education Statistics. (2018). The nation's report card: Mathematics and reading assessments 2017 (NCES 2018-037). Washington, D.C.: Institute of Education Sciences, U.S. Department of Education. Retrieved from https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018037

National Council of Teachers of Mathematics (2000). *Principles and standards for school mathematics*. Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics (2012). Closing the opportunity gap in mathematics education [Position statement]. Retrieved from: https://www.nctm.org/uploadedFiles/Standards_and_Positions/Position_Statements/Opportunity%20Gap.pdf

National Council of Teachers of Mathematics (2014). Access and equity in mathematics education [Position statement]. Retrieved from: https://www.nctm.org/uploadedFiles/Standards_and_Positions/Position_Statements/Access_and_Equity.pdf

National Governors Association Center for Best Practices, & Council of Chief State School Officers. (2010). *Common core state standards in mathematics*. Washington, DC: Authors.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the national mathematics advisory panel*. Washington, DC: U.S. Department of Education.

O'Connor, R. E., Beach, K. D., Sanchez, V. M., Bocian, K. M., & Flynn, L. J. (2015). Building BRIDGES: A design experiment to improve reaching and United States history knowledge of poor readers in eighth grade. *Exceptional Children*, *81*, 399–425. https://doi.org/10.1177/0014402914563706

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal*, *44*(4), 921–958. https://doi.org/10.3102/0002831207308221

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. https://doi.org/10.1007/s10972-015-9436-1

Pierce, M. E., & Fontaine, L. (2009). Designing vocabulary instruction in mathematics. *The Reading Teacher*, *63*, 239–243. https://doi.org/10.1598/RT.63.3.7

Powell, S. R., & Driver, M. K. (2015). The influence of mathematics vocabulary instruction embedded within addition tutoring for first-grade students with mathematics difficulty. *Learning Disability Quarterly*, *38*, 221–233.

Powell, S. R., Driver, M. K., Roberts, G., & Fall, A. M. (2017). An analysis of the mathematics vocabulary knowledge of third- and fifth-grade students: Connections to general vocabulary and mathematics computation. *Learning and Individual Differences*, *57*, 22–32. https://doi.org/10.1016/j.lindif.2017.05.011

Powell, S. R., & Nelson, G. (2017). An investigation of the mathematics-vocabulary knowledge of first-grade students. *The Elementary School Journal*, *117*, 664–686.

Putnam, R. T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, *29*(1), 4–15.

Reinke, W. M., Lewis-Palmer, T., & Martin, E. (2007). The effect of visual performance feedback on teacher use of behavior-specific praise. *Behavior Modification*, *31*(3), 247–263. https://doi.org/10.1177/0145445506288967

Riccomini, P. J., Smith, G. W., Hughes, E. M., & Fries, K. M. (2015). The language of mathematics: The importance of teaching and learning mathematical vocabulary. *Reading & Writing Quarterly*, *31*, 235–252. https://doi.org/10.1080/10573569.2015.1030995

Rock, M. L., Gregg, M., Thead, B. K., Acker, S. E., Gable, R. A., & Zigmond, N. P. (2009). Can you hear me now? Evaluation of an online wireless technology to provide real-time feedback to special education teachers-in-training. *Teacher*

Education and Special Education, *32*, 64–82. https://doi.org/10.1177/0888406408330872

Romig, J. E. (2018). *Improving writing instruction of in-service teachers of students with disabilities: A multimedia approach to professional development* (Unpublished doctoral dissertation). University of Virginia, Charlottesville, VA. Retrieved from Libra.

Romig, J. E., Sundeen, T., Thomas, C. N., Kennedy, M. J., Philips, J., Peeples, K. N., … Mathews, H. M. (2018). Using multimedia to teach self-regulated strategy development to pre-service teachers. *Journal of Special Education Technology*, *33*, 124–137. https://doi.org/10.1177/0162643417746373

Sanetti, L. M. H., Fallon, L. M., & Collier-Meek, M. A. (2012). Increasing teacher treatment integrity through performance feedback provided by school personnel. *Psychology in the Schools*, *50*(2), 134–150. https://doi.org/10.1002/pits.21664

Sawyer, R. K., & Greeno, J. G. (2009). Situativity and learning. In P. Robbins & M. Aydede (Eds.), *Cambridge handbook of situated cognition* (pp. 347–367). New York, NY: Cambridge University Press.

Scheeler, M. C., Congdon, M., & Stansbery, S. (2010). Providing immediate feedback to co-teachers through bug-in-ear technology: An effective method of peer coaching in inclusive classrooms. *Teacher Education and Special Education*, *33*(1), 83–96. https://doi.org/10.1177/0888406409357013

Scheeler, M. C., & Lee, D. L. (2002). Using technology to deliver immediate corrective feedback to preservice teachers. *Journal of Behavioral Education*, *11*, 231–241.

Scheeler, M. C., McAfee, J. K., Ruhl, K. L., & Lee, D. L. (2006). Effects of corrective feedback delivered via wireless technology on preservice teacher performance and student behavior. *Teacher Education and Special Education*, *29*(1), 12–25. https://doi.org/10.1177/088840640602900103

Scheeler, M. C., McKinnon, K., & Stout, J. (2012). Effects of immediate feedback delivered via webcam and bug-in-ear technology on preservice teacher performance. *Teacher Education and Special Education*, *35*(1), 77–90. https://doi.org/10.1177/0888406411401919

Scheeler, M. C., Morano, S., & Lee, D. L. (2018). Effects of immediate feedback using bug-in-ear with paraeducators working with students with autism. *Teacher Education and Special Education*, *41*(1), 24–38. https://doi.org/10.1177/0888406416666645

Scheeler, M. C., Ruhl, K. L., & McAfee, J. K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education*, *27*(4), 396–407.

Schleppegrell, M. J. (2007). The lingustic challenges of mathematics teaching and learning: A research review. *Reading & Writing Quarterly*, *23*, 139–159.

Seidel, T., Stürmer, K., Blomberg, G., Kobarg, M., & Schwindt, K. (2011). Teacher learning from analysis of videotaped classroom situations: Does it make a difference whether teachers observe their own teaching or that of others?

Teaching and Teacher Education, *27*, 259–267. https://doi.org/10.1016/j.tate.2010.08.009

Sherin, M. G., & Han, S. Y. (2002). Teacher learning in the context of a video club. *Teaching and Teacher Education*, *20*, 163–183.

Siuty, M. B., Leko, M. M., & Knackstedt, K. M. (2018). Unraveling the role of curriculum in teacher decision making. *Teacher Education and Special Education*, *41*, 39–57. https://doi.org/10.1177/0888406416683230

Solomon, B. G., Klein, S. A., & Politylo, B. C. (2012). The effect of performance feedback on teachers' treatment integrity: A meta-analysis of the single-case literature. *School Psychology Review*, *41*(2), 160–175.

Stahl, S. A. (1999). *Vocabulary development*. Cambridge, MA: Brookline Books.

Sutherland, K. S., Wehby, J. H., & Copeland, S. R. (2000). Effect of varying rates of behavior-specific praise on the on-task behavior of students with EBD. *Journal of Emotional and Behavioral Disorders*, *8*(1), 2–8.

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

Swanson, E. A., Solis, M., Ciullo, S., & McKenna, J. W. (2012). Special education teachers' perceptions and instructional practices in response to intervention implementation. *Learning Disability Quarterly, 35*, 115-126. doi:10.1177/0731948711432510

Sweigart, C. A., Landrum, T. J., & Pennington, R. C. (2015). The effect of real-time visual performance feedback on teacher feedback: A preliminary investigation. *Education and Treatment of Children*, *38*, 429–450.

Sweller, J., Van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.

Temple, C., & Doerr, H. M. (2012). Developing fluency in the mathematical register through conversation in a tenth-grade classroom. *Educational Studies in Mathematics*, *81*, 287–306. https://doi.org/10.1007/s10649-012-9398-6

Thompson, D. R., & Rubenstein, R. N. (2000). Learning mathematics vocabulary: Potential pitfalls and instructional strategies. *The Mathematics Teacher*, *93*, 568–574.

van der Walt, M., Maree, K., & Ellis, S. (2008). A mathematics vocabulary questionnaire for immediate use in the intermediate phase. *South African Journal of Education, 28*, 489–504.

van Garderen, D. (2008). Middle school special education teachers' instructional practices for solving mathematical word problems: An exploratory study. *Teacher Education and Special Education*, *31*, 132–144.

Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*, 147–177.

Virginia Department of Education (2017). Virginia state quality profile. Retrieved from http://schoolquality.virginia.gov/virginia-state-quality-profile#desktopTabs-2

Wei, R. C., Darling-Hammond, L., and Adamson, F. (2010). *Professional development in the United States: Trends and challenges*. Dallas, TX. National Staff Development Council.

Witzel, B. S., Mercer, C. D., & Miller, M. D. (2003). Teaching algebra to students with learning difficulties: An investigation of an explicit instruction model. *Learning Disabilities Research & Practice*, *18*, 121–131.

Xin, Y. P., Jitendra, A. K., & Deatline-Buchman, A. (2005). Effects of mathematical word problem-solving instruction on middle school students with learning problems. *Journal of Special Education*, *39*, 181–192.

Yoon, K. S., Duncan, T., Lee, S., Scarloss, B., & Shapley, K. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. Washington, DC: Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest, U.S. Department of Education.

Appendix A

Table A1

Teacher Demographic Survey

Survey Question	Response Options
1. What is your gender/gender identity?	Multiple choice (1 selection allowed) • Female • Male • X or non-binary • Prefer not to answer
2. What is your age?	Text/numerical response
3. Which category best describes you and/or your background?	Multiple choice (multiple selections allowed) • African-American • Asian/Asian-American • Latinx/Latin-American • White/non-Latinx • Additional/Other not listed [text response] • Prefer not to answer
4. What is the highest level of education you have completed?	Multiple choice (multiple selections allowed) • Bachelor’s degree [with text response] • Master’s degree [with text response] • Doctoral degree [with text response] • Other professional degree [with text response]
5. For how many total years have you taught mathematics, including any grade/curriculum?	Text/numerical response
6. Including the 2018-2019 school year, for how many years have you taught 6th grade mathematics/Math 6?	Text/numerical response
7. What other grades/courses in mathematics have you taught? Please also	Text/numerical response

indicate how many years you taught each course.	
8. Which concept or topic from the curriculum of your current course (Math 6) do you feel *most* prepared, confident, and/or excited to teach this year?	Text/numerical response
9. Which concept or topic from the curriculum of your current course (Math 6) do you feel *least* prepared, confident, and/or excited to teach this year?	Text/numerical response
10. Are you currently licensed in Virginia to teach Math 6?	Yes/No
11. What is your current licensure status?	Multiple choice (1 selection allowed) • Initial (includes Collegiate Professional) • Provisional (includes Career Switcher) • Professional (includes Postgraduate Professional) • Awaiting Virginia licensure transfer • Other [with text response]
12. If your current license covers Math 6, which grade band does your current licensure include?	Multiple choice (1 selection allowed) • Elementary (PK-6 or K-6) • Middle Grades (6-8) • Secondary (6-12) • Other [with text response]
13. How many classes (i.e., separate groups of students) do you teach each day this year?	Text/numerical response
14. Do you teach mathematics courses in more than one grade level this year? (e.g. Math 6 and Algebra I)	Yes [with text response]/No
15. How many total students do you teach this year, in all of your rosters combined?	Text/numerical response
16. Across all of your current rosters, how many students do you teach that receive some kind of special education services, including Section 504 Plans?	Text/numerical response

17. As of today, how many total students are in the class that is being observed for this project?	Text/numerical response
18. As of today, how many total students *who have current IEPs or 504 Plans* are in the class that is being observed for this project?	Text/numerical response

Table A2

Teacher Social Validity Survey

Survey Section/Question	Response Options
A. Time Use	
A1. Before this study began, about how much time did you spend teaching vocabulary?	Likert response • 1 = None at all • 2 = A little (only when questions come up) • 3 = About 5-7 minutes per class • 4 = About 7-10 minutes per class • 5 = More than 10 minutes per class
A2. Before this study began, about how much time did you spend planning lessons that included teaching vocabulary?	Likert response • 1 = None at all • 2 = Not a lot, but some • 3 = Sometimes a lot, sometimes less • 4 = A lot of time
A3. During the study, about how much time did you spend planning lessons that included teaching vocabulary?	Likert response • 1 = None at all • 2 = Not a lot, but some • 3 = Sometimes a lot, sometimes less • 4 = A lot of time
A4. During the study, about how much time did you spend reading the feedback that you received?	Likert response • 1 = None at all • 2 = Not a lot, but some • 3 = Sometimes a lot, sometimes less • 4 = A lot of time
A5. During the study, how often did you watch the CAP-TV videos?	Likert response • 1 = None at all • 2 = I watched some videos, but not all. • 3 = I watched each one only once. • 4 = I watched some more than others.

	• 5 = I watched each video multiple times.
A6. Now that the study is over, do you think you will watch the videos again?	Likert response • 1 = No • 2 = Maybe • 3 = Yes
A7. Now that the study is over, do you think you will include explicit vocabulary instruction in your lesson planning?	Likert response • 1 = No • 2 = Maybe • 3 = Yes
B. Impressions of Lesson Feedback	
B1. How often did you look at your CT Scan data output (the link included in your feedback document)?	Likert response • 1 = I didn't look at it very often. • 2 = Most of the time • 3 = Every time I received it (and sometimes more than that)
B2. Overall, to what extent did receiving the feedback influence your planning and/or instructional decisions?	Likert response • 1 = It wasn't relevant or helpful for my instructional planning. • 2 = It was sometimes helpful for my instructional planning. • 3 = It was very helpful for my instructional planning.
B3. Overall, to what extent did receiving the feedback influence your planning *specifically regarding your vocabulary instruction?*	Likert response • 1 = It wasn't relevant or helpful for my instructional planning. • 2 = It was sometimes helpful for my instructional planning. • 3 = It was very helpful for my instructional planning.

B4. Which components of the feedback did you find most helpful to your instructional planning? Rank each component, with 1 being the most helpful.	Ranked response • CT Scan data output • Number of opportunities to respond, feedback or redirect statements, etc. • Minutes spent on vocabulary instruction (line graph) • The Big Picture • One Big Thing
B5. In general, which feedback format did you think was easiest to understand quickly?	Multiple choice (1 selection allowed) • Written/narrative (such as Big Picture or One Big Thing) • Numbers and rates (such as OTRs, feedback or redirect statements, etc.) • Graphics (such as timeline, pie graphs, or line graph)
B6. In general, which feedback format did you think the most informative?	Multiple choice (1 selection allowed) • Written/narrative (such as Big Picture or One Big Thing) • Numbers and rates (such as OTRs, feedback or redirect statements, etc.) • Graphics (such as timeline, pie graphs, or line graph)
C. Impressions of CAP-TVs	
C1. What is your impression of the length of each of the CAP-TV videos (on average)?	Multiple choice (1 selection allowed) • They are just the right length. • Some could be shorter or were too repetitive. • Some could be longer or have more explanations and examples.
C2. What is your impression of the overall quality of the CAP-TV videos?	Likert response • 1 = I didn't think any of the videos were easy to follow or informative. • 2 = Most of the videos are not that informative. • 3 = They're mostly just OK. • 4 = Most of the videos are easy to follow and informative.

	• 5 = They are all easy to follow and informative.
C3. What is your impression of the overall relevance of the vocabulary practices you were shown in the CAP-TVs, *in terms of mathematics instruction?*	Likert response • 1 = I don't think vocabulary instruction is relevant to mathematics. • 2 = The practices are mostly relevant, but I had to make a lot of adjustments. • 3 = The practices are mostly relevant, and I felt confident adjusting them when needed. • 4 = The practices were relevant and easy to incorporate into my mathematics instruction.
C4. What is your impression of the overall relevance of the vocabulary practices you were shown in the CAP-TVs, *in terms of your students' academic needs?*	Likert response • 1 = I don't think vocabulary instruction would support my students' math learning and growth. • 2 = The practices might support my students' math learning and growth. • 3 = The practices did (or could) support my students' math learning and growth.
C5. How likely are you to use the materials from this PD again? (Note: "Materials" includes the videos, slide template, and your feedback)	Likert response • 1 = Extremely unlikely • 2 = Somewhat unlikely • 3 = Unsure • 4 = Somewhat likely • 5 = Extremely likely

Note: Adapted from Hirsch (2016) and Romig (2018).

Appendix B

Dear Mrs. #,

Thank you for being part of this project and welcoming me into your classroom. Immediately below is a table noting your strengths, targeted areas for improvement, and action steps for becoming even more awesome. Pay special attention to the **One Big Thing** (last row of the table). We think improving the One Big Thing will have a big payoff for your students.

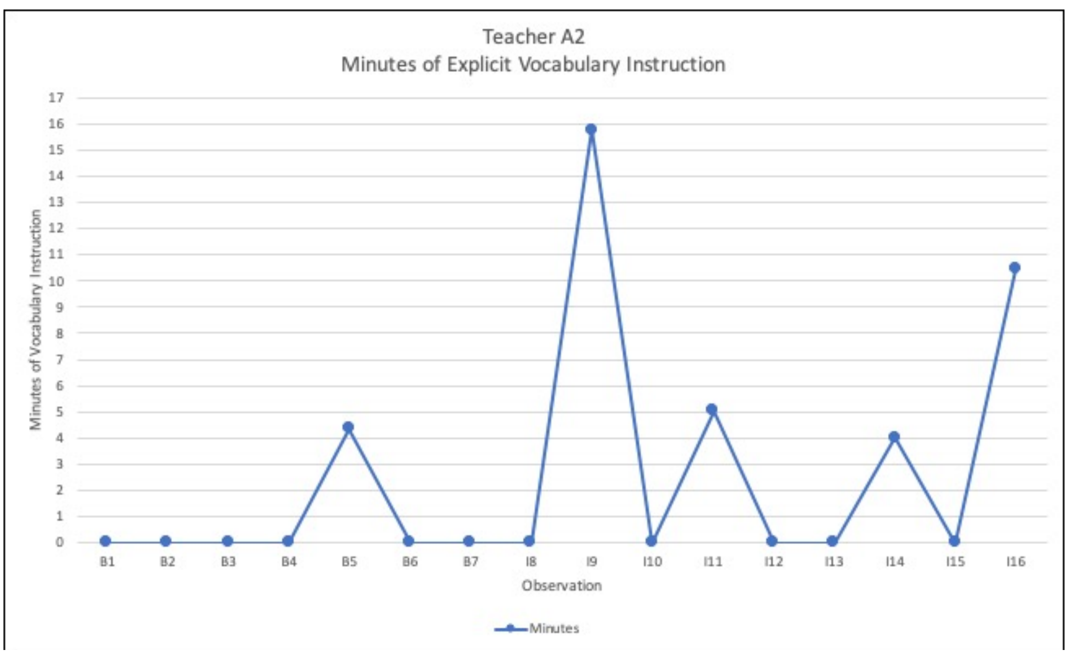
Click on the link below to view the performance feedback (descriptive data) from your lesson. We included a tutorial on how to interpret this data in Video #7 on www.mathvocabsupport.com (password is ######): https://vimeo.com/291025087
If you have any questions about the data or the feedback in this email, please email me (kpeeples@virginia.edu). Please do not share or discuss this feedback with other teachers until the conclusion of the study.

Link to CT Scan Data Output:

http://www.classroomteachingscan.com/ctscan/timeline.htm?menus.txt&782

Observation Date: 10/31/18

The graph below shows the number of minutes you have spent explicitly teaching vocabulary since the beginning of the study.
On average to date, you spend just over 2 ½ minutes per lesson on vocabulary instruction.



Practices	What You Did	Comments
General Vocabulary Instruction		N/A
Explicit Mathematics Vocabulary Instruction	**Student-Friendly Definition** *zero pair* *opposite* *absolute value* **Examples** *zero pair* *opposite* *absolute value* **Semantic Relationships:** *zero pair/opposite/absolute value*	Great job (again!) with these! I'm glad you're finding these practices useful for reviewing as well as for teaching new terms. This part of your lesson took just 11 minutes and was relevant for students today as well as being relevant again later on. You're rocking these vocabulary practices – keep up the good work!
Opportunities to Respond (OTRs)	34 questions 24 rote 10 choral 0 deep 0 student question	I liked that you provided so many opportunities for students to respond while you were doing the vocabulary review!
Feedback Statements (FB)	33 feedback statements 30 generic 3 academic 0 behavioral	You gave feedback statements 33 times, including 3 times when you expanded on student answers and built on their responses. You gave specific behavior redirect statements 3 times.
Comments		
The Big Picture	You seem really comfortable with using these practices, which is great to see! I watched the kids' reactions today and I think they respond (by staying engaged) to the sequence and the images. Very nice work! Keep doing what you're doing, and check out the videos when you need fresh ideas or a refresher on the implementation features. Most of your instructional time today was spent on facilitating independent practice, which happened at several points throughout the lesson (total of about 34 minutes, or 53.9% of the class period). Vocabulary instruction took up roughly 20% of the lesson.	

	Students split their time today fairly evenly, once again. I really like the consistency you have with this group – it makes sense to students and it means expectations are always clear when they know what to expect.
One Big Thing	You're doing great with these practices. Keep looking for ways to review/re-teach terms, teach new terms, and tackle multiple terms a day when it makes sense to do so.

Remember!

You can use the professional development video series and other resources at any time: www.mathvocabsupport.com. Individual links to the videos are below [NOTE: the password for all videos in this series is #####]:

https://vimeo.com/288598412 (Video 1: Welcome to the Project)
https://vimeo.com/288619218 (Video 2: Using Student Friendly Definitions)
https://vimeo.com/290174251 (Video 3: Using Examples & Non-Examples)
https://vimeo.com/291030400 (Video 4: Using a Morphological Approach)
https://vimeo.com/291026552 (Video 5: Demonstrating Semantic Relationships)
https://vimeo.com/291027887 (Video 6: Facilitating Academic Discussions)
https://vimeo.com/291025087 (Video 7: Tutorial on Interpreting Your Performance Feedback)