UNIVERSITY OF CALIFORNIA
RIVERSIDE

Effect of Performance Feedback on Increasing Quality Classroom Instruction in Middle School

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Education

by

Valerie M. Perry

June 2013

Dissertation Committee:
Dr. Michael Vanderwood, Chairperson
Dr. H. Lee Swanson
Dr. Gregory Palardy

The Dissertation of Valerie M. Perry is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE DISSERTATION

Effect of Performance Feedback on Increasing Quality Classroom Instruction in Middle School

by

Valerie M. Perry

Doctor of Philosophy, Graduate Program in Education
University of California, Riverside June 2013
Dr. Michael Vanderwood, Chairperson

Effective classroom instruction is a critical element to improving outcomes for low performing students and optimizing the effectiveness of multi-tiered systems. Middle school is a particularly important area to address in improving classroom literacy instruction as it becomes a place where students learn skills that allow them to transition from "learning to read" to "reading to learn." Professional development literature has suggested some effectiveness with the use of coaching, but study results have been mixed. One potential solution is the inclusion of a consultation with performance feedback model into the professional development process. This study proposed to examine to what extent instructional consultation with performance feedback increases teacher behavior in implementing targeted elements of quality instruction as well as student engagement and literacy skill acquisition. Teachers from 6th, 7th, and 8th grades at an urban middle school took part in instructional consultation with weekly performance feedback. Observations were conducted and single case design with combined concurrent and non-concurrent multiple baseline was utilized to determine

treatment effect. Results indicated an effect on use of random response in instruction and an overall increase in student engagement. Implications for these findings are discussed as they relate to the consultation literature and practical applications for school psychologists.

Table of Contents

List of Tables

List of Figures

Effect of Performance Feedback on Improving Fidelity to Professional Development for Classroom Instruction

While it is a generally accepted notion that all grade levels could benefit from high quality instructional practices, middle school instruction is in particular need of attention. By the time students reach middle school, literacy has become an integral part of the learning process in all subject areas resulting in a shift from "learning to read" to "reading to learn". Common teacher expectations at the middle school level are for students to read passages of text in order to extract meaning in a variety of classes, including and beyond English language arts (Kamil et al., 2008). However, if a student has not achieved these necessary literacy skills by this time in their academic career, their ability to perform to these expectations is greatly diminished (Juel, 1988; Stanovich, 1986).

Unfortunately, current research suggests that this is the case for the majority of middle school students in the United States. According to the National Assessment of Educational Progress (NAEP), in 2007, 69% of 8th graders performed below proficient in the area of grade level comprehension. This indicates that teacher expectations for middle school students and the students' actual skill level is incongruent. As a result middle school students face a unique challenge acquiring higher order literacy skills while simultaneously applying them in order to learn material across content courses.

This discrepancy between expectation and reality is starting to be addressed with the implementation of Common Core Standards. The Common Core Standards for literacy for adolescents revolve around teachers of English language arts, history/social

studies, science, and technical subjects using content areas to improve skills in reading, writing, speaking, and listening. Unfortunately, like their students, research indicates that this is not a change that teachers feel readily prepared for. Teacher efficacy research indicates that teachers of adolescents report feeling unprepared to teach literacy skills, particularly in content classes (Kamil et al., 2008). This is an indication that research is needed to determine how to best support teachers in applying quality literacy skill instruction in their classrooms.

Importance of Effective Classroom Instruction

Effective classroom instruction is crucial to student academic growth. A review of multi-level studies indicates that 20% of the variance in student achievement is attributable to classroom influences (Whitehurst, 2002). In other words, one fifth of student achievement can be accounted for by what goes on in the classroom. This is a significant contribution, made only more significant by the fact that this data is measured at a single time point. This means that the 60% contributed by the individual student's background also includes prior learning due to prior classroom experiences. This indicates that the potential influence for classroom impact could be even higher than the 20% indicated in the literature (Whitehurst, 2002).

In addition to multi-level studies, value added studies also indicate that the classroom has a significant impact on student achievement (Jordan, Mendro, & Weerasinghe, 1997; Sanders & Rivers, 1996; Whitehurst, 2002). In value added studies students are randomly assigned to classrooms with either the most effective or least effective teachers for three years. Teachers were considered effective, or "high," if they

were in the top 20% of average amount of annual growth on state tests in their school. The least effective, or "low," teachers were in the bottom 20% of growth. At the end of the third year their respective growth is examined. In 1996, Sanders and Rivers studied the influence of "high" and "low" teachers in two school districts in Tennessee. Results of the study showed that students randomly assigned to three "high" teachers in a row scored in the 85th percentile. Those assigned to three "low" teachers in a row scored in the 29th percentile. Adding to these results, Jordan and colleagues (1997) conducted a similar study which produced similar results. The students with three years in classrooms of "high" teachers scored in the 76th percentile while those with the "low" teachers scored in the 27th percentile. It should be noted that these studies are limited in that it is very unlikely that a student will be assigned to three "high" or "low" teachers; however, the vast discrepancy in academic performance between the groups of students is a strong indication of the impact a classroom teacher can have on achievement (Whitehurst, 2002).

In addition to its impact at the student level, core classroom instruction is a crucial element of school wide multi-tiered systems. This is because not only does it improve outcomes for students, but it also assists in a more accurate determination of which students are in need of individualized intensive services. In multi-tiered systems it is hypothesized that students performing below expected academic criterions may have an individual need that requires special attention with targeted services or potentially special education (Salvia, Ysseldyke, & Bolt, 2007; Vaughn & Fuchs, 2003). However, it is not possible for schools to determine if a student has a problem if the student has not

previously been exposed to quality core instruction (Ardoin, Witt, Connell, & Koenig, 2005; Fuchs & Fuchs, 1998). If a student begins their academic career with exposure to poor core instruction and continually falls behind, it is highly likely that he/she will continue to remain behind their peers (Stanovich, 1986). This phenomenon, known as the Matthew Effect, indicates that if students fall behind due to poor instruction they will likely remain behind their peers, possibly resulting in making them appear to be a potential candidate for special education. This can result in labeling a student with a disability, and a significant loss of resources for a school, both of which may have been preventable by ensuring quality core instruction. Therefore core instruction is a vital area for research to address as it has the potential to have a large impact on student achievement as well as improves efforts towards early identification of academically at risk students.

The need for quality instruction is also particularly salient for schools with populations of students that have a history of underachievement. One such population that is gaining increasing research attention are schools with high English language learner (ELL) populations (Echevarria, Short, & Powers, 2006). ELL populations are growing and it is estimated that by 2030, Ells will represent 40% of the school aged population (National Institute of Child Health and Human Development [NICHD] (2003). The largest proportion of Ells is Spanish speakers which represent approximately 80% of Ells in the United States (Zehler et al., 2003) and 85% of Ells in California (California Department of Education [CDOE], 2007). However, despite these large numbers, this population consistently achieves far below their native English speaking

(NES) peers. As mentioned previously, evidence supporting the Matthew Effect (Stanovich, 1986) suggests that students who begin their education behind will continue to stay behind and the achievement gap will continue to widen. This phenomenon is also demonstrated with the chronic low performance of Hispanic ELLs. In a 2004 report on the achievement of English language learners following a California based initiative to address the learning needs of students via sheltered instruction, Parrish and colleagues found that literacy achievement differed significantly between ELL and NES students between the years of 1997 and 2001. Results of the study showed that there was not only a significant gap between the two groups, but also that this gap grew over time with a 1-2 year, 2 year, and 4.5 year gap in 4th, 8th, and 11th grades, respectively.

Despite the potential for high quality core classroom instruction to improve student outcomes, and the potential impact quality core instruction can have in addressing the achievement gap between ELL and NES students, there are numerous reports of significant shortages of qualified teachers prepared to undertake this challenge (Echevarria, Short, & Powers, 2006). Because of this reality, as well as the research to support the dramatic impact that quality instruction can have on the academic trajectory of large groups of students (Ardoin, Witt, Connell, & Koenig, 2005; Fuchs & Fuchs, 1998; Hanusheck, 1992), the role of implementing and maintaining strong core instruction in the classroom is becoming an important area of research.

Quality Instruction and Middle School

Much of the current literature base has focused on increasing literacy outcomes for elementary grades and early interventions. In 2000 the National Reading Panel

released a report highlighting the importance of five areas of literacy instruction: phonological awareness, decoding, fluency, vocabulary, and comprehension. In addition to this it has been documented that the most effective literacy interventions take place in elementary school and have a focus of phonological awareness, decoding, and fluency (Torgeson et al., 2001; Wanzek & Vaughn, 2007). Teaching early literacy skills in this manner has gone beyond intervention work and can be readily seen in the classroom and in state standards for elementary grades.

The effect of these efforts can be seen in the improvement of reading scores, as reported by the National Assessment of Educational Progress's 2008 report. This report showed that, on average, nine year old students demonstrated a 12 point growth from 1971-2008. However, this same type of growth was not seen in the 13 and 17 year old groups. The 13 year old groups demonstrated only a 4 point gain and the 17 year old group did not demonstrate any significant growth at all. The growth seen with 9 year olds is an indication that implementing instruction based on research based practices addressing early literacy skills has been effective in improving outcomes for elementary grades. However the lack growth with 13 and 17 year olds indicates that attention to evidence based instructional practices for adolescents is needed. To address this need the Institute of Educational Sciences (IES) released a practice guide addressing effective instruction for improving literacy outcomes for middle and high school students (Kamil et al., 2008).

The following literature review will first define what comprises high quality instruction for adolescents using the IES document as a guide. It will then explore the

research base addressing how to change teacher behaviors in order to improve the quality classroom instruction.

Effective Classroom Instruction for Adolescent Students

As mentioned previously, in 2008, the Institute of Educational Sciences published a report on effective classroom practices with upper elementary, middle, and high school students. The report examined the literature base of instructional studies conducted with students in 4th through 12th grade settings, in which reading was the dependent variable. The study produced five recommendations. Four of these recommendations applied to classroom instructional practices and could be implemented in all core content classes to improve adolescent literacy across subject areas. One focused on intensive and individualized settings. The four classroom based recommendations included: (1) Provide explicit vocabulary instruction, (2) Provide direct and explicit comprehension instruction, (3) Provide opportunities for extended discussion of text meaning and interpretation, (4) Increase student motivation and engagement in literacy learning.

Common across these guidelines is the use of direct and explicit instruction, as well as utilizing practices that result in high active engagement. Therefore it can be conceptualized that educators utilizing quality instruction with adolescents will use these instructional practices in conjunction with the instructional components of vocabulary, comprehension, and discussion.

In the following discussion of quality instruction for adolescents, the instructional practices pervasive across instructional components (direct instruction and active engagement) will first be discussed. This will then be followed by a more specific review

of effective instructional practices in vocabulary, comprehension, and classroom discussion.

Instructional practices.

Direct instruction. A type of instruction commonly used to improve student outcomes is direct and explicit instruction. This type of instruction consists of systematically addressing and teaching the steps of the skills being taught, as well as providing strategies for learning and practicing the new skills via structured materials, teacher directed classrooms, monitoring of student performance, and immediate feedback (Burns, VanDerHeyden, & Boice, 2008). Instruction of this type has demonstrated positive results over other forms of instruction such as cognitive or affective models, as shown in one study called Project Follow Through.

Project Follow Through was a longitudinal study that began in 1967 and lasted through 1995. This study sought to examine three major models of instruction on early literacy skills (Adams, 1996). The first model was the Basic Skills Model, which included carefully designed lessons to teach desired educational behaviors. Within this model the following approaches were considered: (a) Direct Instruction Model (University of Oregon), which emphasized phonics and the above mentioned elements of direct instruction,(b) Behavioral Analysis Model (University of Kansas) which consisted of behavior reinforcement based on behavioral principles of positive reinforcement, and (c) the Language Development Model (Southwest Educational Development Laboratory) which focused on language development beyond specific reading skills.

The second model was the Cognitive/Conceptual Skills Model, which emphasized cognitive growth over specific content and self guided activity. Within this model the following approaches were considered: (a) Cognitively Oriented Curriculum Model (High Scope Foundation), which consisted of children scheduling their own activities and explaining causal relationships (b) Florida Parent Education Model (University of Florida), which emphasized parent education with a focus on teaching parents how to teach their children language, motor, and cognitive skills, and (c) the Tucson Early Education Model (University of Arizona), which focused on self directed literature materials that involved student choice of materials to encourage enjoyment of the activity with the theory that it would foster motivation. The third model was the Affective Skills Model, which emphasized quality interpersonal relationships, fostering social and emotional development of students, and students directed learning experiences. Within this model the following approaches were considered: (a) Bank Street College Model (Bank Street College of Education) which consisted learning centers where the student can select options in a structured activity, (b) Open Education Model (Education Development Center) which consisted of student exploration of learning in a semi-structured setting with no teacher involvement, and (c) the Responsive Education Model (Far West Laboratory), which focused self esteem building through dynamic individualized curriculum. Each of the programs within the models was implemented in 4-8 sites and compared to the control schools that were not implementing any of the specific models/programs.

Results of the study revealed that the basic skills model had the most optimal results over the other two models. The direct instruction format within the basic skills model had the highest scores in regards to improving basic skills (Adams, 1996). Surprisingly enough it was also the only instructional style within any model to show improvements in cognitive measures, in spite of the cognitive model having instructional programs attempting to directly influence this domain. Also, the instructional styles within the basic skills model had a positive impact on student affective skills, with direct instruction having the greatest impact, while the components within the affective style had no positive results. These results, together, form a strong empirical basis for the assertion that instruction organized around a basic skills model that emphasizes direct instruction has the potential for the largest impact on improving academic skills as well as improving social and emotional skills such as engagement and motivation.

The positive effects of direct instruction has also been demonstrated with teaching specific literacy skills such as vocabulary and comprehension. In regards to vocabulary, research supports the teaching of vocabulary, and vocabulary skills, in a rich instructional format with direct and explicit instruction (Beck, Perfetti, & McKeown; 1982; Bos & Anders, 1990; Jenkins, Matlock, Slocum, 1989; Townsend & Collins, 2009). This type of instruction consists of a combination of instructional practices that incorporates direct instruction, multiple exposures to a word in multiple contexts, active practice of the words, and making personal connections (Townsend & Collins, 2009). Very little evidence exists to support incidental learning of vocabulary words. A 1999 meta-analysis on the role of incidental word learning in vocabulary development found that

incidental word learning does occur, however the probability of it happening is only 15% (Swanborn & de Glopper). This means that out of 100 new words taught via incidental exposure a student can be expected to learn approximately 15. In addition to this, 66% of the new words learned could be explained by previous knowledge of a partial word and the student's age. In other words, incidental learning not only produces very little new word acquisition, but also is only effective for students with higher background knowledge and previous reading ability.

Direct instruction of vocabulary has also been found to have a larger effect than utilizing inferential techniques such as context clues. In a 1989 study 5th grade students were exposed to either direct instruction of vocabulary words or taught to use context clues to infer the meaning (Jenkins, Matlock, & Slocum). Instruction in this study consisted of either a direct instruction condition that consisted of either explicit introduction of the word, practice with definition repetition, exposure to and use of the word in sentence, and practice with replacing it with synonym, or, a context clue condition where instruction focused on how to use context clues to infer word meanings. Within these two groups, students were further broken down to low, medium, and high practice conditions. Results indicated that direct instruction, with any amount of practice, resulted in greater improvements than context clue instruction.

Consistent with vocabulary instruction, comprehension strategies are also best taught via direct and explicit methods. Bereiter and Bird (1985) conducted a study where levels of direct and explicit instruction were applied. Seventh and 8th grade students with average skills in oral and silent reading were taught comprehension strategies in one of

four treatment conditions. The first was "modeling with explanation" which consisted of explicit explanation of strategies, modeling of the strategies via teacher thinking aloud, and student practice and feedback while they thought aloud using the strategy. The second condition was "modeling only" which consisted of teacher modeling of the strategy via thinking aloud, however there was no explicit explanation or identification of the strategy. The third condition was "exercise" which consisted of oral and written exercises that required students to carry out operations taught by the strategies but no exposure to modeling or explicit explanation was received. Finally, the fourth condition was the "control" condition, in which students were exposed to their typical classroom instruction.

Of these four conditions only modeling with explanation had a significant effect on standardized measures of achievement. This result provides strong evidence that direct and explicit instruction is an important element in teaching strategies to improve comprehension Bereiter and Bird (1985). This study found that modeling alone was not sufficient to improve acquisition of strategies helpful in improving comprehension. Combining direct and explicit explanation of the strategies with modeling, however, did. This lead the authors to hypothesize that direct instruction is an important element that serves to direct student attention toward the relevant components of what is modeled.

Schunk and Rice (1992) also examined the influence of teaching explicit specific strategies on reading comprehension achievement with 4^{th} and 5^{th} grade students in remedial classes. In the study, students in the treatment condition were exposed to comprehension instruction on strategies on how to find main ideas and details. The

control group was also instructed to complete this task, but without being taught strategies on how to do so. Half way through the experiment (10 weeks) randomly selected students from the treatment group were exposed to a modified strategy on how to locate details. The strategy steps for locating details was the same as the steps for locating main ideas, except the words within the strategy specific to main idea was changed to words specific to details. Results via repeated measures ANCOVA indicated that students exposed to the modified strategy instruction group outperformed students in the static strategy instruction group and no strategy control group on measures of skill, self-efficacy, and reported use of the strategy steps. Interestingly, the study did not find a difference in skill gain between the static strategy instruction group and control group, as is counter to previous studies. However, this may be due to the low initial skill level of subjects in the study. This is in line with previous research that asserts that struggling students are in most need of direct and explicit instruction. The modified skill instruction condition in this study was essentially repeated direct instruction of the strategy with explicit attention across its uses. Therefore, this study supports the effectiveness of continuous direct and explicit instruction of comprehension strategies with struggling students that are easily implemented in a classroom setting.

High engagement. Academic engagement is one of the strongest factors that influence academic performance (Gettinger & Ball, 2008). Students cannot be expected to learn if they are not engaging in the instructional process. Academic engaged time is discussed in the literature as the amount of time a student is engaged in learning by doing things such as paying attention, working on assignments, and participating in class

discussion (Gettinger & Ball, 2008; Rathvon, 2008). These include answering questions, raising hands, writing, reading, and discussing content. According to Rathvon (2008), an effectively managed classroom will have a rate of 80% academic engaged time or higher.

One major classroom factor that can increase academic engagement is a teacher's instructional practices (DuPaul, Stoner, & O'Reilly, 2008; Gettinger & Ball, 2008). There are multiple ways to incorporate instructional practices that increase engagement. Techniques such as active student responding, frequent student-teacher interactions, providing performance feedback, and continuous progress monitoring have shown to increase academic engaged time. Active responding can be increased with frequent questions and optimizing the number of students called on to respond via random selection, coral responding, or peer tutoring (Gettinger & Ball, 2008; Greenwood, Maheady, & Delquadri, 2002; Rathvon, 2008). Active responding, such as writing, reading aloud, or answering a question, should be emphasized as opposed to passive responding, such as listening to a lecture.

High student/teacher interaction time can also increase academic engagement in students (Gettinger & Ball, 2008). Interactive instructional techniques include modeling, scaffolding, and explaining that requires students to respond to the teacher and subsequently engage in the activity. This is in opposition to low interactive activities such as unsupervised seatwork and sustained silent reading (Gettinger & Ball, 2008). Providing performance feedback is another way to increase student-teacher interactions. Teacher provided feedback that is frequent, specific, and relevant and provides specific

information on how to correct or rethink incorrect responses is associated with increased engaged time (Gettinger & Ball, 2008).

Finally progress monitoring is another way to increase engaged time. Formal monitoring with curriculum based measurement or other psychometrically sound screening tools has demonstrated effectiveness in identifying when classroom students need additional learning time with a particular skill and which individual students may need more focused individualized help in a more targeted setting (Ikeda, Neesen, & Witt, 2008). Informal progress monitoring can also be an effective "in the moment" instructional tool. With this type of monitoring a teacher quickly assesses student understanding of a topic at the start of a lesson via choral response or physical indication (ex. hold up fingers 1-5 to indicate how well you know this vocabulary word). This type of monitoring can provide instant feedback to a teacher about the needs of the class in acquiring a particular skill so that appropriate adjustments can be made in the lesson plan, as well as increases opportunities for engagement and feedback (Rathvon, 2008).

Another technique used to increase engagement is peer tutoring. Peer tutoring is an intervention technique that has strong empirical support for its effects on academic performance for students with and without disabilities (Greenwood, Maheady, & Delquadri, 2002; Rathvon, 2008). In their review of peer tutoring literature, Greenwood, Maheady, & Delquadri (2002) state that peer tutoring is effective because it provides increased opportunities to respond, receive and give encouragement, practice, receive immediate error correction with feedback, and receive individualized instruction.

To implement peer tutoring, first students are paired according to their instructional skill level, thus ensuring instructional match. Students then receive modeling from the tutor who is instructed on how to explicitly demonstrate methods to solve difficult reading problems such as sounding out difficult words. Pairing up in this way is hypothesized to be effective because the pair format allows for all students to have an increased opportunity to actively respond to material (Greenwood, Maheady, & Delquadri, 2002). In addition to this, because they are in one-on-one groups their academic engaged time is increased, as there are no other students to distract their tutor's attention. Peer tutoring also benefits average or above average performing students as they are allowed to strengthen their knowledge through extended exposure to the material and by teaching and articulating the process to others (Greenwood, Maheady, & Delquadri, 2002). Peer tutoring incidentally also has benefits beyond academic performance. In addition to increasing engagement and academic skills, Greenwood and colleagues (2002) also found that students demonstrated improved social status among peers, increased positive peer interactions, and improved relationships between minority and majority groups.

Peer tutoring has also shown promise in improving the academic performance of English language learners (Ells), as well. In a 2007 report, the Institute of Education Sciences named five recommendations for effective instructional strategies for Ells. One of the five recommendations was to consistently utilize peer tutoring for at least 90 minutes a week. In their review of the peer tutoring literature, the authors found strong

evidence for utilizing peer tutoring consisting of pairing students of differing skill levels and/or language proficiency levels to work on academic tasks in a structured fashion.

In a 2001 multiple baseline single case design study Greenwood and colleagues conducted a class wide peer tutoring program with five elementary school classrooms, 1st through 5th grade, in urban settings (117 ELL students). As part of the study, students engaged in peer tutoring to address vocabulary and spelling with teacher monitoring. Results indicated that students made and sustained gains in both areas. In addition, teachers implemented the program with high fidelity, indicating that peer tutoring can be easily incorporated into instruction in a classroom with high ELL populations and have beneficial results.

Overall increasing active engagement has shown to improve academic outcomes for students. Techniques such as active student responding, frequent student-teacher interactions, providing performance feedback, and continuous progress monitoring have shown to increase academic engaged time (Gettinger & Ball, 2008; Greenwood, Maheady, & Delquadri, 2002; Rathvon, 2008). Instructional practices such as active responding, scaffolding instruction, and peer tutoring are all specific elements of instruction that have shown to have positive outcomes for students at all academic performance levels (Gettinger & Ball, 2008; Greenwood, Maheady, & Delquadri, 2002; Rathvon, 2008).

Instructional components.

Vocabulary. Vocabulary knowledge has been cited as an important function in improving comprehension skills (Bauman et al., 2002; National Reading Panel, 2000). A

1986 meta-analysis found that improvements in vocabulary resulted in a small effect size of $d = 0.30$ for comprehension of standardized passages (Stahl & Fairbanks). In addition, vocabulary instruction utilizing instructional practices such as teaching a limited number of words, teaching high utility words, and balancing direct teaching of words with teaching word-learning strategies demonstrated improvements in comprehension on a standardized comprehension test (Gates-MacGintie) and was found equally effective with English language learners (Kelly, Lesaux, Kieffer, & Faller, 2010).

Given that vocabulary improvements are linked to improvements in comprehension, instructional practices that most effectively teach vocabulary are important in core classroom instruction. Not all attempts to teach vocabulary are equally effective. As mentioned previously, little evidence has been found with incidental or brief exposure while much evidence exists for direct and explicit instruction. Other ineffective vocabulary instruction include looking up definitions in the dictionary and utilizing context clues in a way that utilizes the same type of information for associative processing (Stahl & Fairbanks, 1986).

Looking up definitions in the dictionary is a common practice in vocabulary instruction, but has little evidence for effectiveness. Bos and Anders (1990) compared strategy instruction that included semantic mapping, semantic feature analysis, and semantic/syntactic analysis to definition instruction that consisted of exposure to a word list and definitions. Results indicated that strategy instruction had a larger effect than definition learning across both vocabulary and comprehension measures. This is in line

with the assertion that direct and explicit instruction with modeling is most effective with students over simple exposure and definition memorizing.

Instruction with context clues is another common instructional practice in schools that has limited demonstration of effectiveness. A research synthesis on vocabulary instruction with context clues found that while there are some studies that demonstrated improvement with context clue work, this effect was only seen in studies that did not include a practice control group (Kuhn & Stahl, 1998). This indicates that context clues may provide no more influence on word acquisition than simple practice. In addition to this there is little evidence that context clues provide strategies for determining meanings of novel words. A 1996 study found that brief explanation of words within a context revealed stronger gains than incidental learning, however there was little evidence to show that it helped students determine meanings of novel words not directly addressed by the teacher (Brett, Rothlein & Hurley).

Research that does exist showing the effectiveness of context clues often utilizes it with a combination of other instructional techniques such as frequent exposure to the words being taught, use of the new words in different contexts, and connecting them to previously learned content. In their 1982 study, Beck, Perfetti, & McKowan examined the effects of vocabulary instruction that utilized explicit instruction, synonyms, words in sentences, and words in different contexts on vocabulary and comprehension outcomes with 4th grade students. Students showed strong gains on standardized state tests of vocabulary and comprehension. In addition to this Nelson and Stage's (2007) study found that with vocabulary instruction that took place 2-3 days for 20-30 minutes a day,

related new words to known words (semantic), discussed related words (synonyms), practiced words with sentences and utilized word map/definition matching, students made much more substantial gains than a control group. Of note, the highest gains were made with students with lowest vocabulary at the pre-test. This is substantial as previous studies indicated that vocabulary improvement was highly dependent on previous achievement. This provides a potential solution to improving vocabulary for students with limited background knowledge or language exposure such as with students in poverty or those learning English.

Improving vocabulary skill, however, involves more than being able to retain words taught directly in class. Individuals also need skills to derive meaning from novel words. Morphological instruction in root words, prefixes, and suffixes is an area that has demonstrated effective growth in students' skills to determine meaning of novel words. A 2002 study conducted by Baumann and colleagues examined the effect of morphology instruction on instructed words and transfer words. Fifth grade students were exposed to morphemic instruction, instruction in morphemic and contextual analysis skills, or a control group with neither. Results supported use of morphology instruction with both treatment groups demonstrating similar gains in vocabulary over the control group. In addition to this, a 2010 meta-analysis on morphological instruction with students at risk for learning difficulties found moderate effects for vocabulary (d=.40) and small gains in comprehension (d= .24; Goodwin & Ahn, 2010).

Additionally, rich instruction that includes direct instruction with strategies for word meaning acquisition and novel word identification has also been found to be

beneficial for English language learner students (Carlo et al., 2004; Townsend & Collins, 2009). In a 2004 study, Carlo and colleagues provided rich vocabulary instruction to 5th grade English language learners (ELL). The instruction included teaching target words in contexts that relate to background knowledge of the students, reading passages with the target words, exposure to words in different contexts, synonym/antonym identification, semantic feature analysis, morphological study, and cognates. Results revealed a large effect for vocabulary outcomes (η^2=.34) and moderate effect for comprehension (η^2=.08) for both English only and ELL students. Despite the similar effects of these evidence-based practices, evidence for the need for further differentiating instruction for ELL students does exist. For instance ELL semantic knowledge has been found to be less developed than English only peers (Townsend & Collins, 2009). As a result including visual aids and providing increased opportunities for practice in various contexts may be beneficial to enhance instruction (August, Carlo, Dressler, & Snow, 2005; Townsend & Collins, 2009).

Comprehension. As mentioned previously, in middle and high school grades the focus of reading changes from learning to read to reading to learn. For adolescent students, this means strategy instruction in comprehension should focus on developing those skills that will help them acquire information and literacy skills that will be valuable in future academic and occupational needs (O'Connor & Goodwin, 2011). These skills include summarizing, drawing inferences, generating questions, organizing information, and predicting. It is important to note that there is a lack of evidence that supports one strategy over another. It is unknown whether, for example, summarizing

has better outcomes than inferring. However, there is evidence to suggest that instruction utilizing a combination of these strategies can have beneficial effects (Kamil et al., 2008).

Hansen and Pearson (1983) examined the effects of combining instruction with the goal of improving 4th grade student's skills in making connections across texts, making predictions, and drawing inferences. Students were assigned for five weeks to either a group with a combination of instructional strategies or a control group with typical instruction. Results indicated that improvements were demonstrated in the treatment group over the control group with poor readers on researcher made comprehension worksheets and tests of comprehension on reading passages at the student's reading level.

In a single case design study, Jitendra and colleagues (1998) examined the effects of direct instruction in main idea identification and summarization. Instruction was broken into 7 lessons that increased in difficulty, with the final lesson consisting of a review of the previous lessons and application of what was learned to passages in the students' readers. Student participants included four 6th graders who were receiving support for a specific learning disability in reading, judged by their teachers to be poor readers, scored two years below grade level on the Woodcock Reading Mastery comprehension test, and performed below 50% on a criterion test of comprehension. Two students were African-American and two were Hispanic. Results indicated that summarization and main idea identification instruction was effective in increasing reading comprehension performance on identifying and generating main ideas as

measured by researcher made questions based on probes from the student's curriculum reader.

Combining comprehension techniques by way of a graphic organizer has also received attention as a potential tool to benefit literacy skills. Graphic organizers are tools used to visually depict relationships between concepts with the theory that once relationships are made clear the learner will have easier access to the information in the text (DeCecco & Gleason, 2002). The impact of graphic organizers on comprehension outcomes is mixed. In a 2006 meta-analysis, Nesbit and Adescope reviewed experimental and quasi-experimental studies involving graphic organizers and found moderate effects on attaining, retaining, and transferring knowledge in various course subjects. The highest effect sizes were shown for intermediate schools (4^{th}-8^{th}), work with organizers in lecture/discussion settings, across subjects, and when students constructed the information in the map rather than studying a completed map. However, in comparison to other teaching techniques, such as writing summaries or outlines, the difference in effects was relatively small.

These results were consistent with an earlier meta-analysis by Moore and Readance (1984), which found small to moderate effect sizes for use of graphic organizers. Higher effect sizes in this study were associated with using organizers after reading text (as opposed to use in pre-reading activities). These mixed results suggest that positive effects found with graphic organizers may be due to the increased student engagement in comprehension activities (Moore & Readence, 1984; Nesbit & Adescope, 2006). However, the small improvement in effect size with graphic organizers over

reading text and writing summaries or outlines suggests that the role of graphic organizers in improving students' comprehension warrants further research into the exact nature of their influence (Nesbit & Adescope, 2006).

Another important area to address and combine with the previously noted strategies is teaching explicit strategies for different types of texts that students encounter. Typical texts encountered by students are narrative, expository, and documents (Gersten, Fuchs, Williams, & Baker, 2001; Hock & Mallard, 2005). Students typically struggle with expository writing due to its varied text structure, which can include description, cause and effect, and sequence formats (Gajira, Jitendra, Sood, & Sacks, 2007; O'Connor & Goodwin, 2011). Because of this, teachers should explicitly teach strategies for comprehending different structures. Smith and Friend's 1986 study found that in measures of text structure identification and main idea recall students taught strategies in recognizing and using text structure outperformed students that received general instruction on how to problem solve. This is worthy of note, as it has been typically seen that students who lack the knowledge of addressing different text structures often approach text without a comprehension plan (Meyer, Brandt, & Bluth, 1980). Based on this, specific strategies such as summarizing, drawing inferences, generating questions while reading, and predicting should be explicitly taught and applied across different types of text (narrative, expository, and documents).

In a study to combine the effects of summarizing, main idea identification and text format identification, Schumaker and colleagues (1982) taught teachers an instructional technique they termed the "multipass strategy". With this technique,

students were taught (via modeling, scaffolding and practice) to first pass through the chapter surveying information they could glean from parts of a chapter such as the title, introductory paragraph, subtitles, organization of the chapter and illustrations, and to note how the chapter relates to adjacent chapters. At the end of this pass students then summarized all information that had been presented. In the second pass students were taught how to glean information from a chapter without reading it from beginning to end. This was done by giving students questions and then teaching how to look for answer clues in the text, skim the surrounding text for the answer, and paraphrase the answer. After this the student re-paraphrased all the information they had acquired from the text. In the final pass students were given questions to answer, and were instructed to think of what section the answer might be found in, skim that section, think of another section it might be in if it was not in the section they thought it was in, skim that subsequent section, and so forth until they found the answer, if they were initially unsure. The technique resulted not only in increased use of the strategy in students, but also a generalization of the strategy to other courses.

Discussion. Discussion as a part of instruction has shown promising influence on vocabulary and comprehension skills. Due to a small research base and limitations in research designs less information is known about the direct impact discussion has on improving literacy skills. The IES 2008 practice guide noted that only moderate evidence is present for the assertion the discussion has a direct impact on instructional outcomes. However, they note that enough research exists that indicates aspects of discussion useful in maximizing impact on literacy skill improvement in students.

According to NAEP standards, an 8th grade student achieving proficient in reading should analyze beyond basic comprehension and demonstrate skill in summarization of main ideas, make arguments with support from text, and identify relationships between ideas in text. It is theorized that high quality discussion will allow students to develop these skills via guided thinking as well as having their thinking challenged by other students or teachers that can then be used as a model for future reading on their own (Kamil et al., 2008). In a 2003 study Applebee and colleagues used hierarchical linear modeling to examine the relationship of classroom discussion to literacy performance of students in middle school and high school English classrooms. Results indicated that when controlling for fall performance, discussion significantly predicted spring performance on a writing task designed to assess the degree to which a student provides information and elaborates beyond a given topic.

While the exact impact discussion has on academic outcomes is unknown in the current literature base, studies have found that factors do exist that moderate the effectiveness of discussion. There is evidence to suggest that specific discussion strategies that directly focus on improving summarizing, making predictions, and linking text to previous knowledge is a more effective approach than a less focused discussion (Kamil et al., 2008). In a 2007 meta-analysis, Murphy and colleagues explored the impact types of discussion had on comprehension outcomes. In the study three categories of discussion were studied and their effect on types of comprehension was analyzed. The three types of discussion included: 1) critical-analytical, which consisted of teacher encouragement of a student's subjective and critical response to a text, 2) efferent, which

consisted of discussion focused on acquiring and retrieving information from text, and 3) expressive, which consisted of discussion of a student's emotional or affective response to the text. Types of comprehension variables measured included: 1) text explicit comprehension, which consisted of comprehension of information explicitly stated within a single place in the text such as a sentence or paragraph, 2) text implicit comprehension, which consisted of comprehension requiring synthesizing information across the text such as in multiple paragraphs or chapters, 3) scriptally implicit comprehension, which consisted of utilizing prior knowledge and information in the text, and 4) general comprehension which consisted of comprehensions measures which were not clear enough to categorize into the prior three comprehension types.

Results from the study indicated that different discussion types had an impact on comprehension outcomes. The highest effect sizes on comprehension types was efferent discussion approach with text explicit comprehension ranging from g= 0.33-1.34 (compared to 0.49 with critical-analytical and no studies available with expressive), scriptially implicit comprehension ranging from g= 0.62-0.87 (compared with .42-.66 with critical-analytical and no studies available with expressive), and general comprehension ranging from g= -0.21-2.80 (compared to .26-.33 with critical-analytical and .43 with expressive). The only comprehension category that efferent discussion approach did not have the largest effect size was text implicit comprehension. In this case the expressive discussion approach had the highest effect with a range of g=.82-2.13. Efferent did second best however with a range of g=0.57- 1.12. These results suggest that discussions that focus on text content and teach students how to access text

information will have the largest influence in improving comprehension skills over discussions based on subjective impressions or emotional responses.

It has also been hypothesized that incorporating discussion is helpful with improving comprehension skills with English language learner students. Active discussions linking text to prior knowledge before, during, and after reading may assist in creating a knowledge base for future exposures to instruction rooted in the dominant culture. This can be done through the discussion of topics or new vocabulary words encountered in the text that may be similar to other experiences the student's may have had in their culture (Orosco & Klingner, 2010). In this way, through discussion of these cultural connections, text that is rooted in the dominant culture can become culturally relevant to the minority culture, which theoretically may improve students' comprehension of text.

Professional Development for Improving Classrooms

Knowing there is an evidence base for effective instructional techniques for improving adolescent literacy, the next step is to determine how to ensure core classroom teachers are equipped with the tools to implement this kind of instruction. As discussed previously, continuous professional development with an effort toward quality classroom instruction is critical for school success (Robertson, Woolsey, Seabrooks, & Williams, 2004), particularly with low performing cultural and linguistic minority students (Echevarria, Short, & Powers, 2006; Parrish et al., 2004). However, often there are reports of shortages of teachers in the field that feel prepared to implement quality instruction with these populations (Echevarria, Short, & Powers, 2006). In 2003 the U.S.

Secretary of Education reported that teacher education and certification programs demonstrated little relationship to teacher effectiveness. While the overall effectiveness of teacher education programs is debated in the literature (Darling-Hammond, 2000; Robertson, Woolsey, Seabrooks, & Williams, 2004), this report provides an indication that despite the quality of their teacher education program, very few teachers begin their teaching career with all of the tools necessary to be effective in teaching their students. It is an even larger indication that continuing professional development efforts at the school level are needed to empower teachers to acquire and retain skills necessary to provide quality instruction (Robertson, Woolsey, Seabrooks, & Williams, 2004).

Elements of effective professional development. Traditionally, professional development is one way schools attempt to improve instructional practices in a school. However, not all professional development approaches are equally effective. Traditional models that consist of a single meeting or information session, such as workshops and conferences for instance, have shown little to no effectiveness in changing teacher behaviors (Kinkead, 2007). This is because they are not designed to provide ongoing contextual support to improve learning (Kinkead, 2007). There is also weak empirical support to suggest that talking to teachers about change is sufficient to change behavior (Noell & Witt, 1996). Because traditional models of professional development and informal approaches have demonstrated little influence in changing teacher behavior, a strong push has been made in professional development research to determine what kind of supports teachers need to improve and sustain their instructional capacity (Carlisle, Cortina, & Katz, 2011).

Efforts to determine effective professional development components have found that effective professional development attempts need to be focused on increasing content knowledge, engage educators in active learning, and be part of a coherent model of school change (Desimone, Porter, Garet, Yoon, & Birman, 2002; Garet, Porter, Desimone, Birman, & Yoon, 2001; Stollar, et al., 2008). Effective procedural efforts include a combination of meetings, workshops, follow-up, and opportunities for hands-on practice (McIntyre, Kyle, Chen, Munoz, & Beldon, 2010).

One aspect of professional development the literature has shown may potentially improve the impact of professional development efforts is reducing professional isolation via interactive activities that connects teachers to other professionals such as with coaches or consultants (Guskey, 2002; McIntyre, Kyle, Chen, Munoz, & Beldon, 2010). Traditionally, during these interactions a teacher meets with another professional to receive feedback on current teaching practices and to learn research based best practices they can incorporate into the classroom. In regards to improving implementation of quality instruction models for students, coaching has had varied success in altering teacher behavior.

Coaching research to enhance professional development for quality instruction. In 2010, Carlisle and Berebitsky conducted a study to compare the effects of professional development in schools with and without a coaching component with first grade teachers. All teachers in the study attended a total of 9 seminars on reading and reading instruction, using Language Essentials for Teachers of Reading and Spelling (LETRS). Each session was three hours every 3-4 weeks and covered one module of LETRS (27 hours total).

Teachers in the professional development with coaching group met with coaches either individually or in groups on a weekly basis. The purpose of the coaching sessions were to support teacher understanding of the LETRS model and assist in their delivery of the content. Coaching sessions occurred once a week in weekly grade level meetings and consisted of modeling methods of instruction and review of research based literacy instruction. Coaches also provided one on one support via classroom visits. Results showed that compared to the no coach condition, teachers in the coaching group demonstrated higher patterns of reading instruction consistent with the LETRS model they were trained in. These patterns included increased use of small group instruction and a gradual diminishing instructional emphasis on phonological awareness as the year progressed. Students in coaching classrooms also showed an increase in word decoding. Overall, this study shows that coaching adds value to professional development efforts and improving student early literacy outcomes.

Additionally, Matsumura and colleagues (2010) conducted a randomized three-year, longitudinal study of schools engaging in a professional development program that highly emphasized literacy coaching of teachers. Participating classrooms consisted of 32 elementary schools serving the lowest achieving students in a Texas district. Literacy coaches worked to increase teacher use of discussions to support text understanding. Over the course of a year, coaches met with teachers once a week during grade level meetings and once a month for one on one sessions. In these sessions coaches assisted with planning and modeling lessons and discussed observations. Results of the study indicated that teachers reported increased use of classroom discussions related to text and

other activities related to quality comprehension instruction as taught to them by the literacy coach. Researcher observed sessions supported this report and found an increase in quality of reading comprehension instruction in schools with the coaching condition over those in the control condition; however this is difficult to interpret as the treatment group began with higher observer ratings than the control group. This significant difference over time then could potentially be a product of an unmeasured construct. Also confounding this study's results is the fact that the teachers used self report to indicate increases in evidence based instructional practices taught to them by the coaches, and as previously noted, teachers tend to over report implementation of interventions (Noell, 2008; Wickerstrom, Jones, LeFleur, & Witt, 1998). Finally, this study is limited in that there was no control group to assist in isolating the coaching variable. This study subsequently provides some evidence that professional development with coaching demonstrated change in teacher behavior, however the effect of coaching on its own is unknown.

Similar to the previous two studies, Batt (2010) also examined how coaching improves professional development efforts. Specifically Batt explored the influence of cognitive coaching on implementation of the SIOP model, which is meant to increase effective instruction for culturally and linguistically diverse students. The study first looked at the influence of an initial training in the SIOP model via a summer training session and district sponsored workshops over the course of a year. After this training, 80% of teachers reported a commitment to implementing the SIOP model in their classes, however only 50% reported using SIOP "often" or "very often" on a teacher survey.

Following this training, a coaching phase was implemented. In the coaching phase a cognitive coaching model was used which consists of a preconference, observation, and post-conference format.

Cognitive coaching differs from other types of coaching in that changing behaviors is not the primary focus. Rather, coaches serve as mediators to facilitate teacher self-reflection with the assumption that that will influence instructional behaviors. Post-conference coaching sessions took place one to two days after a classroom observation and consisted of rapport building, reflection on teaching episodes, questioning with the intent to affect cognition, and development of understanding of the SIOP model and how to apply it. Two to three days after this meeting teachers were provided with written feedback on the observed lesson that focused specifically on the SIOP components. The aim of these sessions was to increase level of SIOP implementation. Results indicated that after coaching 100% of teachers reported implementing SIOP to "a great extent." These teacher reports were corroborated with observations that indicated similar findings. For instance, prior to coaching 60% of teachers wrote and posted language objectives and by the end of the coaching phase 100% did. This study demonstrates that adding coaching to the professional development model influenced teacher behavior by increasing their implementation of the SIOP model.

However, not all studies have found coaching to add value to professional development efforts. Garet and colleagues (2008) conducted a study, published in a report by the Institute of Educational Sciences, that also examined the effect coaching has

on professional development efforts. In this study, 90 school districts were randomly assigned to a professional development, professional development and coaching, or a control condition. The control condition consisted of business as usual in regards to a given school district's typical professional development efforts. Similarly, in Carlisle and Berebitsky's (2010) study, the professional development only group attended seminars that utilized the LETRS model. Unlike their study however, the teachers in Garet and colleagues' study received 48 hours of these seminars; approximately 20 hours more than the subjects in the previous study. Teachers in the professional development and coaching group attended the 48 hours of seminars, plus engaged in coaching meant to increase their understanding of the LETRS content and provide support for ongoing practice and application of the new knowledge. Teachers and coaches met for approximately 2 hours a week with an expectation of a total of 60 hours over 30 weeks.

Coaches reported that their coaching time consisted of planning (14.8 hours), observing teachers and providing feedback (15.6 hours), working with the teacher in the classroom (20.3 hours), and conducting grade level meetings (10.6 hours). Observers collected data on three areas the professional development was intended to affect: explicit teaching methods, guided student practice, and differentiation of instruction. Results of the study indicated that the professional development only and professional development with coaching groups demonstrated increased explicit instruction (effect sizes .33 and .53). However, there was no difference between groups on the other two behaviors of measurement. Additionally, no statistically significant differences were seen between the coaching and non-coaching professional development groups in any of the three areas

measured, nor were significant differences found on measures of reading achievement from district assessments.

In addition to this study, McIntyre and colleagues (2010) conducted a study examining teacher implementation of sheltered instruction via professional development training on the SIOP model. In this study teachers were exposed to 50 hours of training over an 18 month period. Teachers also met with a coach to examine a lesson plan and get feedback from a classroom observation. Due to limitations in the study, the coaching session only occurred once. Outcomes of the study indicated that only a third of teachers became full implementers of the model. This study is significantly limited, however, in that no control group was present and unlike the other studies the coaching only consisted of one meeting. This finding suggests a need for significantly more time engaging in the coaching process. This is highlighted by the fact that teachers that did demonstrate high fidelity to the model were those that went above and beyond the professional development activities. Their increased involvement in the process could have served as a surrogate to the coaching process. In addition to this those teachers that grew the least were teachers that started the study with lowest implementation of the SIOP model and subscribed to a more traditional style of teaching. This indicates that perhaps teachers with lower initial practices targeted in professional development are in most need of supports and follow up. Therefore, while no significant effects were found on altering teacher behavior, this may be more of an indictment of the lack of coaching time versus a lack of effect coaching can provide.

Overall these studies provide evidence that coaching varies in its effectiveness. Three of the studies found increases in the targeted teacher behaviors (Carlsile & Berebitsky, 2010; Matsumura et al., 2010; Batt, 2010), while two did not (Garet et al., 2008; McIntyre et al., 2010). In addition, only two studies reported student outcomes (Carlsile & Berebitsky, 2010; Garet et al., 2008), and only one of these showing a positive effect (Carlsile & Berebitsky, 2010). This indicates that while there is some indication that coaching may be effective in influencing teacher behavior, more research is needed in order to discern what elements of coaching makes it effective and how to apply them effectively.

Consultation replacing coaching for improving professional development Efforts. Another area gaining increasing attention in the literature, due to its evidence of effectiveness in improving classroom instruction, is that of school based consultation. School based consultation also has practical value as it is a common existing practice of school psychologists. As mentioned previously core instruction is crucial to the success of multi-tiered systems. Not only is effective core instruction useful in ruling out potential problems due to exposure to poor instruction (Ardoin, Witt, Connell, & Koenig, 2005; Fuchs & Fuchs, 1998), but it also ensures that all students are receiving quality instruction that maximizes academic skill acquisition and prevents achievement gaps from occurring (Stanovich, 1986). Due to their relatively new role in developing and implementing multi-tiered systems, as well as their well-established role of engaging with teachers as a consultant, the school psychologist is in a potentially unique position to influence all tiers of the multi-tiered system.

Professional development literature highlights interactive and contextualized experiences as key components of effective professional development efforts (Guskey, 2002; McIntyre, Kyle, Chen, Munoz, & Beldon, 2010) and the coaching literature suggests that coaching can be an effective way to incorporate these experiences. However, the coaching literature is mixed in how much value coaching adds to changing teacher behavior in their classrooms and little is known on how to maximize its potential effectiveness (Garet et al, 2008). Research has found that the best outcomes are found when utilized coaching in an ongoing (more than once) and consistent manner. In the McIntyre and colleagues (2010) study, coaching only occurred once in 18 months and no effects were found. In the Garet and colleagues (2008) report, coaching was conducted over many more hours, however the authors note that despite the 60 hours on average, there was significant variability in that average with a range of 1.7 to 117 hours. Nine percent received less than 20 hours and 18% received between 20 and 39 hours. These studies, in conjunction with the other studies, do not necessarily indicate that more hours are needed, but perhaps that the consistency of coaching time is important. As seen in the studies that indicated positive effects (Carlisle & Berebitsky, 2010; Matsumura, 2010), coaching hours were kept consistent with either weekly or monthly meetings with teachers.

Another component that research has found leads to significant outcomes was use of a framework that the coaching was conducted that included sharing of data collected in classroom observations. In the Batt (2010) study, a cognitive model of coaching was conducted that had a clear process of conducting a preconference meeting to build

rapport and a common understanding of the process, a subsequent classroom observation, a post-conference meeting to discuss what was seen, and finally a written report was provided of what was discussed in the post-conference and what was seen in the observation. The use of a model with clear steps and components was not utilized in the other studies so it is hard to determine if a model is a necessary component of improving coaching efforts. However, research on consultation processes has generally shown support that utilizing a process model is effective in changing teacher behavior (Erchul & Sheridan, 2008). Thus, consultation research is an area that may provide similar influences as coaching efforts and address the areas that make coaching effective.

Consultation and Professional Development

School based consultation research frequently focuses on models and processes that optimize efforts toward student improvement via selection of appropriate interventions and maintaining proper implementation of that intervention. Traditionally, consultation efforts are applied to a single referral problem. The traditional application of consultation will be discussed followed by an analysis of how the process can be applied toward the broader concept of improving overall quality of classroom instruction. As seen in the previously reviewed coaching literature, adopting a consistent and clear model of a coaching process may boost the effects of professional development efforts.

Traditional consultation. Research has found that consultation can be effective in addressing a variety of learning and behavioral problems in school settings (Wilkinson, 2006). There has been some debate as to whether consultation is an intervention in itself (Frank & Kratochwill, 2008), but in general most researchers agree that it is better

conceived of as a way to boost components known to improve the effectiveness of evidence based interventions (EBI; Frank & Kratochwill, 2008). This means that the traditional goal of consultation is to (1) identify an appropriate evidence based intervention, and (2) ensure that the intervention is implemented appropriately in the classroom environment.

There have been broad investigations in the area of consultation models that optimize these goals. Behavioral Consultation (BC) is the most popular and empirically supported consultation model in schools in both the United States and the United Kingdom (Wilkinson, 2006). The BC model has four parts: (1) problem identification, (2) problem analysis, (3) treatment implementation, and (4) treatment evaluation (Kratochwill & Bergan, 1990; Wilkinson, 2006). During the problem identification stage the consultant and consultee (the individual implementing the intervention, typically the classroom teacher) work to operationally define problem behavior in observable and measurable terms and form a hypothesis on triggering antecedents and maintaining consequences occurring in the classroom (Kratochwill & Bergan, 1990). A plan is then made to observe and measure the frequency and magnitude of the problem behavior. Then during the problem analysis phase the data is analyzed to either confirm or reject the functional hypothesis and develop an intervention plan (Kratochwill & Bergan, 1990). During the next phase, treatment implementation, the consultee implements the agreed upon intervention from the problem analysis phase. The consultant's role in this phase is to optimize the consultee's ability to implement the plan. This includes teaching the consultee skills needed for the intervention, monitoring treatment implementation,

and revising the intervention as needed (Kratochwill & Bergan, 1990). Finally, in the treatment evaluation phase the consultant and consultee come together to evaluate if the goal for the student has been met, whether the intervention was effective, and engage in post-implementation planning for maintenance and generalization (Kratochwill & Bergan,1990).

Evolving from the BC model is the Instructional Consultation (IC) model (Rosenfield, Silva, & Gravois, 2008). Both models are based on an indirect model that involves the consultant and consultee working together in order to develop an intervention plan, with the responsibility for plan implementation primarily placed on the consultee with consultant support. Where the two models differ is primarily in regards to their philosophic approach and types of problems addressed. BC's philosophical focus is on both the consultee and client (i.e. the student) with the goal of changing client behavior through the consultee. IC's focus is primarily on changing the consultee's behavior with the assumption that the change will impact the client and potential future clients (Lopez & Nastasi, 2008). IC's explicit emphasis on improving teachers' professional capacity for effective instruction distinguishes it from other forms of consultation (Gravois & Rosenfield, 2006).

IC has a similar format to the BC problem solving process, however the content of each phase does differ slightly between the two (Rosenfield, 2008; Rosenfield, Silva, & Gravois, 2008). The first phase of IC is "entry and contracting." In this phase a consultant from the IC team and a consultee (teacher) meet to discuss what the academic concern is and how the IC process will work. The next phase is a combination of the

problem identification and problem analysis phase in BC. In this phase the academic concern is defined in measurable terms, baseline data of the concern is collected, and short and long-term goals are established for the class. The subsequent three phases (intervention design, intervention implementation, and resolution/termination) are the same as the third and fourth in BC. An intervention that is feasible, research based, and acceptable to all parties is decided upon and administered, data is collected to monitor progress, and goals are evaluated to determine if the intervention was successful and/or if further interventions are necessary (Rosenfield, 2008; Rosenfield, Silva, & Gravois, 2008).

Research on IC has indicated positive results for students and teachers. In 2002, Gravois and Rosenfield analyzed a series of studies implementing IC and how this affected special education placement patterns. Data from three studies across three years in 37 schools in seven districts were analyzed. In the first study, pre and post IC implementation special education referral rates were compared. In the first year, special education referrals decreased by 27% with a subsequent decrease of placement by 25%. In the second year of the study four new schools were analyzed. Placement rates from the first year cohort served as pre-IC implementation data and were compared with placement rates from the new four schools. In this phase a 55% reduction in special education placement was recorded. The author notes that these findings are a substantial decrease considering that over the two years school enrollment increased by approximately 5%.

In the second study analyzed by Gravois and Rosenfield (2002), thirteen elementary schools implementing IC were compared to twenty schools not implementing the model across a 5-year period. Results showed that schools implementing IC has a 2.2% reduction in special education referrals, translating to 125 fewer students being placed in special education across a five year period in these 13 schools. Control schools, in contrast, resulted in a decrease of only 0.37%.

Finally, in the third study, a quasi-experimental study was conducted with 20 schools across six districts with the aim of comparing patterns of referrals and placements within schools implementing either IC or utilizing an already existing pre-referral system. Teachers in these schools chose whether to utilize IC or the current system. Results, again, showed that IC demonstrated lower rates of referral and placement than traditional pre-referral systems. In addition to this IC teams also demonstrated a lower referral and placement of African American students. Taken together these studies provide preliminary data to suggest the consistency of IC in reducing unnecessary referrals for special education evaluation and subsequently less potential inappropriate placements. Following the purpose of IC then, conclusions can be drawn, that students received the benefit of remaining in their general education classrooms and their teachers learned instructional techniques that met their academic needs.

In a 2006 study, Gravois and Rosenfield further explored the impact of IC teams on decreasing the disproportional rate of referral and subsequent placement of minority students in special education settings. In the study 13 schools across five districts participated in IC team training to address over-identification of minority students into

special education. Nine comparison schools that did not participate in the IC team process were also selected from the same districts to serve as controls. Results of the study indicated that IC schools showed a drop in minority student placement in special education compared to control schools, with the odds of a minority student in IC schools being referred and placed decreasing by 50%.

In regards to teacher outcomes, Kaiser, Rosenfield, and Gravois (2009) explored the impact IC had on improving teacher perceptions of skill development and implementation. In this study, 274 teachers, representing kindergarten through 5th grade, from 27 schools completed a "Satisfaction with Instructional Consultation Survey." From this sample a further 79 teachers participated in detailed interviews about their experience with IC. Results indicated that teachers reported high satisfaction with the IC process. Interviews revealed that teachers commonly reported increased ability to handle similar problems in the future, development of skills as a result of IC participation, and increased use of these strategies with other students (or plans for future use). Overall, the study indicates that IC is a process that teachers find valuable and may have the potential to improve their teaching practices. It should be mentioned, however, that this study is significantly limited, as it only assessed teacher perception and not objective measures of teacher skill growth.

Consultation for improving treatment integrity. The benefit of consultation in schools is that it can be an efficient and effective tool for determining classroom interventions and improving outcomes for students. Because the premise of consultation is based on an indirect model, the effectiveness of the consultation process depends on

the consultant influencing the consultee to implement the treatment plan as intended (Noell, 2008). Treatment integrity is highly related to positive treatment outcomes and assists in determining if the intervention caused change (Wilkinson, 2006). Ensuring integrity is an important but relatively overlooked component of the consultation process as success of an intervention depends on more than simply identifying what intervention components are effective (Wilkinson, 2006).

One of the "most consistently identifiable" variables with a "clear and definite link to student or systems level outcomes," is treatment integrity (Frank & Kratochwill, 2008). There is evidence to suggest that, "degree of integrity is directly related to the degree of outcome," however, the magnitude of this relationship is less certain" (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000, p. 198). The most common conceptualization of this relationship is a probabilistic one where as integrity decreases the probability of treatment failure increases (Noell, 2008). However, any decrease in integrity does not necessarily dictate a decrease in intervention effectiveness. This is most readily observed when the deviation from treatment is small or minor and a large effect is still observed (Noell, 2008). In addition to its relationship to intervention outcomes, integrity also plays a large role in interpreting intervention results. Change, or lack of change, in student outcomes cannot be attributed to intervention effects if we are unsure if it was implemented correctly (Wilkinson, 2006). This is particularly important as multi-tiered models become more prominent in schools.

Treatment fidelity research is typically done with intervention research, however there is some evidence that it can play a role in predicting student outcomes in

implementation of classroom instructional models. In their 2011 study, Echevarria, Richards-Tutor, Chinn, and Ratleff examined the degree of implementation of the SIOP model and its relationship to student achievement. In the study teacher implementation of the SIOP model was assessed via observation using the SIOP measurement tool. Results indicated that teachers that implemented the model with the highest fidelity also had the students who made the highest gains in measures of vocabulary. Likewise, the lower fidelity, the smaller the student gains became.

In summary, fidelity to evidence based components plays a large part in positive effects being seen in students both in individual and classroom wide applications. However, the research indicates that when left to implement professional development elements on their own, teachers demonstrate low integrity to these components (Kinkead, 2007). Consultation research indicates that some sort of formal follow up or check in procedure is needed to ensure consistent and high levels of treatment implementation (Noell et al., 2000). This is similar to the findings in the PD literature for effective professional development programs to provide opportunities to engage in active learning (Guskey, 2002; McIntyre, Kyle, Chen, Munoz, & Beldon, 2010). Therefore, consultation efforts to increase treatment integrity of interventions could be more broadly conceptualized as also having a role in increasing integrity of implementation of professional development goals. Because of this, there is much need for research to determine the most effective and efficient way for consultants to ensure treatment fidelity to quality instruction in classroom environments (Erchul, Grissom, & Getty, 2008).

Consultation with performance feedback for treatment integrity. As mentioned previously, traditional consultation models focus primarily on development and implementation of interventions. An explicit model for monitoring and improving integrity to interventions or broader classroom instructional models is not explicitly provided (Wilkinson, 2006). One potential solution to this, found in the literature, is the addition of performance feedback into the consultation process. Performance feedback is a step added to the treatment implementation phase and consists of providing systematic feedback with objective data demonstrating treatment effectiveness and level of treatment fidelity (Witt, Noel, LaFleur, & Mortenson, 1997). During the performance feedback process, the consultee is observed by the consultant who collects objective data on the targeted student behavior and teacher treatment fidelity. Later the consultant and consultee meet and the data are presented. A link between the two is discussed, positive teacher actions are praised, and problem solving is done to determine any hindrances to treatment fidelity. The process continues until the teacher reaches a desired level of treatment implementation.

The use of performance feedback has roots in the theory of behaviorism, more specifically in the concept of operant conditioning. In operant conditioning it is believed that behavior change is a result of responding to environmental stimulus and subsequent learning of that behavior is a result of an individual making associations between a behavior and consequences for that behavior (Alberto & Troutman, 2008). In other words, the theory goes that stimuli in an individual's environment trigger a behavior and

then consequences that occur as a result of the behavior maintain the targeted behavior if the consequence reinforcing.

In the case of using performance feedback to increase a teacher's treatment fidelity, the presentation of the observation data may act as the trigger and reinforcer. Presentation of the baseline observation data may trigger increases in the desired behavior via an objective demonstration of the discrepancy between the demonstrated and desired levels of teacher adherence. The subsequent data presentation at each performance feedback meeting then acts as positive reinforcement of behavior change. When graphs demonstrate an increase in the targeted behaviors, coupled with praise from consultant, continued change and/or maintenance of the level of the behavior is reinforced.

Previous research has demonstrated the effectiveness of performance feedback in increasing intervention integrity with individual students. Witt and colleagues (1997) conducted a study where performance feedback was conducted with four general education teachers implementing an academic performance intervention. Utilizing a multiple baseline single case design, each teacher implemented an academic intervention with a student with performance deficits (as opposed to skill deficit). Teachers were then trained in how to implement a reinforcement-based intervention. Training consisted of didactic training with consultant and teacher, student training with consultant supervision, and, "in vivo training on the first day of implementation to ensure accurate implementation" (Noell, Witt, Gilbertson, Rainer, & Freeland, 1997, p.79). After training, baseline data were collected on treatment integrity via permanent products

during teacher independent implementation of the intervention. After baseline, performance feedback was provided with the teacher and consultant meeting daily. After the performance feedback phase, a maintenance phase was conducted where the teacher again independently implemented the intervention.

Results of the study indicated that on the training day teachers implemented the intervention with 100% fidelity, however during baseline implementation integrity quickly began to trend downward (2-3 days after training). Implementation trends increased and maintained high levels during the performance feedback phase with this pattern continuing during the maintenance phase for 3 of 4 teachers. Overall, this study provides preliminary evidence for the influence of performance feedback to not only increase intervention integrity, but to also maintain integrity after feedback is removed. However, it is difficult to say for certainty in this study that performance feedback was the causal influence. One particular criticism is that it is possible that the rigorous training methods during the training period may have influenced the teachers' ability to maintain integrity.

Due to this potential confounding variable, a replication of the Witt et al. (1997) study was conducted by Noell and colleagues (1997). This study aimed to determine if similar results could be found when the training phase was less intensive and more similar to practical consultation practices in schools. Sampling procedures in study were similar to the Witt et al. (1997) study with 3 teachers each implementing a reinforcement-based intervention with one student demonstrating an academic performance deficit. Procedures were the same as the previous study using a concurrent multiple baseline

single case design to collect baseline, performance feedback, and maintenance phases and data collection also consisted of permanent product data. Training, however, was conducted in a more traditional consultation format with the consultant explaining how to implement the intervention during a problem analysis meeting. The teacher then implemented the intervention independently and no further contact was made with the consultant until the performance feedback phase. Performance feedback was conducted in the same way as the Witt et al. (1997) study.

Results of this study were similar to those found in the Witt et al. (1997) study. Teachers demonstrated high levels of fidelity for 2-3 days after training and then began a declining trend during baseline phase. All teachers increased treatment fidelity in the performance feedback phase with consistent integrity of 80% or more. This trend continued into the maintenance phase, however results were less consistent for one teacher. Both of these studies provide strong preliminary evidence for the assumption that performance feedback is a promising tool to increase intervention implementation integrity.

Subsequent studies have examined variables that may or may not impact the effectiveness of performance feedback. For example, Mortensen and Witt (1998) explored the extent to which frequency of feedback meetings influenced treatment fidelity of a reinforcement based intervention to address academic performance deficits with three student-teacher dyads. Procedures were similar to those in the previous studies, with teacher training similar to Witt and colleagues' (1997) study, however performance feedback sessions were conducted weekly instead of daily. As with the

previous studies, a decrease in treatment integrity was seen during independent baseline implementation after training. Weekly performance feedback was effective in increasing integrity during the feedback phase with the maintenance phase continuing these effects for two of three teachers. Results of this study suggest that weekly performance feedback is effective for improving integrity, however, when compared to the effects in the Witt et al. (1997) study the magnitude of improvement was not as high. This suggests that higher frequency meetings may result in a greater effect, however, weekly feedback can produce a desired effect.

Studies have also examined the effect visual analysis of student outcomes and treatment integrity data has on the effectiveness of performance feedback (Noell et al., 2000; Noell et al., 2002). In their 2000 study Noell and colleagues examined the effect of brief feedback meetings with and without the use of student and teacher graphs. As with previous studies treatment integrity decreased and/or became inconsistent during baseline. Results comparing feedback conditions showed that brief meetings without graphed data were effective for two of five student/teacher dyads. The subsequent phase of feedback with graphed data was effective for four of five. A second study (Noell et al., 2002) followed similar procedures and found similar results where brief meetings without data was beneficial for one of four dyads but meetings with data driven feedback improved fidelity for all four. These studies provide some support for the inclusion of data review in the feedback process however some significant limitations exist in that results were somewhat mixed. Three of 9 dyads (across both studies) demonstrated elevated integrity with brief meetings without graphs. Also, due to the nature of the

procedure (providing performance feedback with graphs following a period of providing feedback without) it is difficult to say if the graphs alone incited the increase in implementation integrity or if the prior exposure to brief feedback sessions had an influence as well.

Consultation with performance feedback for professional development. Overall, there is support for performance feedback conducted on a structured, semi-frequent basis organized around graphed student and teacher data in consultation literature. The majority of the literature in this area focuses on a single student within classroom-based interventions. However, there is also some evidence that it can be effective with improving teacher instructional practices at the classroom level as well. The professional development literature indicates that effective professional development needs to be focused on increasing content knowledge, engage educators in active learning, and be part of a coherent model of school change (Desimone, Porter, Garet, Yoon, & Birman, 2002; Garet, Porter, Desimone, Birman, & Yoon, 2001; Stollar, et al., 2008). As mentioned previously instructional coaching is a technique often cited as an effective tool to address these needs (Kinkead, 2007; Duchane, Jolivet, Fredrick, 2011). However, the literature is mixed with regard to practices that predict a significant influence from coaching efforts.

When analyzing the coaching literature that focused on changing teacher instructional behaviors, two potential components could be frequent consistent meetings conducted via a structured process model. Consultation with performance feedback is a potential solution as it provides these hypothesized elements via regular meetings to

discuss specific instructional practices, with the use of data, with the goal to increase teacher behaviors in research based methods. Essentially consultation with performance feedback could theoretically be applied to the coaching process to increase teacher fidelity to the goals of professional development.

The use of applying the behavioral consultation model combined with performance feedback in increasing components of instruction has been examined. For instance, in a study by Duchaine and colleagues (2011) the authors examined the effectiveness of teacher coaching with feedback in increasing teacher use of behavior specific praise statements (BSPS) in high school inclusion math classrooms where students either did or did non exhibit high incidence disabilities such as emotional disturbance, learning disabilities, or other health impairments. During baseline teachers were observed teaching class as usual and the frequency of BSPS was recorded. During treatment teachers were exposed to a 45 minute training session on how to use BSPS, baseline data was shared with individual teachers, and a goal was set for the number of BSPS to make in a given class period. After every third observation session the teacher and coach would meet briefly to discuss the goal for BSPS goal, answer questions about BSPS, and give examples of BSPS. After every session the teacher was given a folder with data on the frequency of BSPS recorded. Results indicated that frequency of BSPS increased significantly and immediately during treatment and maintained after treatment. While this study was focused on a single component that makes up effective teaching, it provides some evidence that a performance feedback model could be applied to the broader construct of classroom instruction.

These results, again, show promise in the role of performance feedback in influencing teacher instructional behavior and taken with the coaching literature, it provides an indication that ongoing and consistent interactions with teachers based around objective data analysis and behavior specific feedback can increase teaching practices targeted in professional development. Based on these studies it is a plausible theory that instructional consultation with performance feedback may be worth exploring as an effective addition to coaching in professional development programs for teachers.

Limitations with performance feedback literature. While there is some evidence to suggest that instructional consultation with performance feedback may be a potential solution to increasing professional development efforts, some limitations do exist in the research that restrict the certainty of this argument. The first limitation is that there is a very small research base examining the application of instructional consultation and its outcomes (Lopez & Nastasi, 2008). The studies that do exist focus primarily on its effectiveness in decreasing special education referrals, inappropriate placements, and lowering the disproportionality rate of referrals and placements of ethnic minority students. An extended assumption from these findings could be that the consultation procedure increased the teachers' instructional effectiveness resulting in lowering the need continued assessment or placement in special education. However, this effect has not been directly studied and, therefore, is only a speculation.

In regards to performance feedback, gaps in the literature limit the application of it to increasing instructional quality. For instance, no instructional consultation studies have been conducted utilizing performance feedback. The majority of PF studies have

been conducted with behavioral consultation models. Therefore, the application of performance feedback beyond the BC model is unknown. In addition to this, the studies utilizing PF within the BC model primarily focus on addressing student behaviors rather than instructional practices. Thirdly, PF with BC has been primarily applied to targeted students rather than whole classrooms. The Duchaine and colleagues (2011) study did look at applying feedback to classroom coaching, however it, too, is limited in that the PF process differed from the process typically seen in the consultation literature. It also did not address classroom instruction on a broad level. Thus, while there is some indication that consultation can have an impact on teacher and student outcomes and improve fidelity to evidence based intervention at the individual and classroom level, the exact impact of extending instructional consultation with PF to improving overall instructional quality and student academic outcomes is not known for certain.

Purpose of This Study

The role of core instruction is vital in improving outcomes for students, as well as for providing a foundation for an effective multi-tiered system. In regards to middle school literacy, effective components of instruction include direct and explicit instruction with high engagement in the areas of vocabulary, comprehension, and discussion (Kamil et al., 2008). Fidelity to quality instruction has been linked to positive student outcomes (Echevarria et al., 2011; Frank & Kratochwill, 2008), however many teachers report feeling ill prepared when entering the workforce to use these components, particularly with historically low achieving populations such as high poverty and language minority students (Robertson et al., 2004; U.S. Secretary of Education, 2003). The professional

development literature calls for efforts beyond traditional workshops or seminars in order to increase contextualized experiences (Kinkead, 2007). Coaching has been cited as one area of promise, however the literature is mixed on its effectiveness (Batt, 2010; Carlisle and Berebitsky, 2010; Garet et al., 2008; McIntyre et al., 2010; Matsumura and colleagues, 2010). It is possible that efforts that provide consistency and are structured around objective data could be linked to influencing the behavioral changes in teachers in order to deliver effective instruction. School based consultation could be an alternative area that could address these needs. As demonstrated in the consultation literature, consultation with performance feedback demonstrated effectiveness in increasing fidelity of intervention implementation at the student level after a decline following a brief training period (Mortensen & Witt, 1998; Noel et al., 1997; Witt, Noel, LaFleur, & Mortenson, 1997). In addition, school psychologists are charged with the task of improving all levels of multi-tiered systems and are already employed and trained to deliver consultation in schools.

Consequently, it is possible that instructional consultation with performance feedback may be an effective model to include in professional development efforts at the classroom level. However, currently no study exists examining this application of consultation and performance feedback to improve overall instructional quality. The purpose of this study, then, was to examine how adding performance feedback to a middle school's ongoing professional development program can increase and stabilize the use of direct instruction practices in the classroom. The study also examined how

increasing these practices impacted student engagement and reading comprehension skills.

Research Questions

The research questions in this study addressed the efficacy of performance feedback consultation in increasing teacher behavior in the areas of utilizing direct instruction and instructional tools with evidence to increase engagement. Research questions also addressed the influence of performance feedback consultation on student outcomes of engagement and reading comprehension.

- Research Question 1: To what extent does performance feedback increase teacher use of targeted instructional practices?
- Research Question 2: To what extent does performance feedback with a teacher increase student engagement?
- Research Question 3: To what extent does performance feedback with a teacher increase student comprehension skills as measured by the general outcome measure MAZE?

Methods

Participants

School. Participant teachers and their respective classes were sampled from an intermediate school in southern California. The school consisted of 999 students in 6th, 7th, and 8th grades. Approximately 99.6% of students were on free and reduced lunch and 69% of students were classified as ELLs. The school was taking part in a year long school wide effort to improve classroom engagement and student literacy skills. At the

start of the school year all teachers from all content areas participated in a one day (8 hour) professional development seminar on the importance of increasing engagement and directly teaching literacy skills in all content areas. Specific evidence based practices in these areas were demonstrated during the seminar and applied to real examples of classroom situations that took place at the school over the previous year. The teachers also received hand outs describing these practices, as well as a copy of the Institute of Educational Sciences Adolescent Literacy Practice Guide (Kamil et al., 2008). Throughout the year a consultant was made available to all teachers and administrators as a resource to assist in implementing these practices in the classroom. The consultant primarily engaged in consultation with the English language arts (ELA) teachers.

Teachers. The teachers were sampled from all available general education core ELA teachers at the school. Eight teachers met this requirement and six teachers were randomly selected. The teachers were randomly assigned to two groups (three in each group). Treatment was applied to group 1 and upon completion, was then applied to group 2. The period that was observed, and subsequently addressed during consultation, was also randomly determined.

Teacher A was a 6^{th} grade teacher and the period observed was 4^{th} period. He had a total of 10 years teaching experience and was approximately 35 years old. There were 27 students in the selected period, 7 were male and 20 were female. Twelve were classified as ELL, 1 was initially fully English proficient (IFEP), 13 were reclassified fully English proficient (RFEP), and one was classified as English only.

Teacher B was an 8th grade teacher and the period observed was 5th period. She had a total of 14 years teaching experience and was approximately 35 years old. There were 37 students in the selected period, 15 were male and 22 were female. Seven students were classified as English only and the remainders were classified as English only.

Teacher C was a 7th grade teacher and the period observed was 1st period. She had a total of 6 years teaching experience and was approximately 30 years old. There were 30 students in the selected period, 18 were male and 12 were female. Fifteen were classified as ELL, 10 were IFEP, three were RFEP, and two were classified as English only. In addition to participating in the professional development activities at the school, Teacher C also attended a one day seminar on vocabulary instruction activities.

Teacher D was a 6th grade teacher and the period observed was 3rd period. She had a total of 10 years teaching experience and was approximately 30 years old. There were 16 students in the selected period, 10 were male and 6 were female. Fifteen students were classified as ELL, and 1 was RFEP. This teacher was away on sick leave for the first part of the school year. She attended the initial seminar but did not teach her classes until approximately 6 months later. When she began working with the consultant a copy of the IES practice guide was given to her again with a brief review of what was covered during the seminar.

Teacher E was a 7th grade teacher and the period observed was 4th period. She had a total of 2 years teaching experience, and was approximately 28 years old. There were 35 students in the selected period, 12 were male and 23 were female. Twelve students

were classified as ELL, one was IFEP, 13 were RFEP, and one was classified as English only.

Teacher F was a 6th grade teacher and the period observed was 5th period. She had a total of 34 years teaching experience and was approximately 50 years old. There were 32 students in the selected period, 12 were male and 20 were female. Three students were classified as ELL, 3 were IFEP, 26 were RFEP, and none were classified as English only.

Measures

Classroom instruction and student engagement. Classroom instruction and student engagement data was collected via direct observation utilizing an instrument created by the researcher. The development of the measure was based on the structure of the Ecobehavioral Assessment System Software [EBASS] (Greenwood, Carta, Kamps, & Delquadri, 1995). The EBASS is an ecobehavioral measure designed to allow for simultaneous, systematic, and objective collection of teacher behaviors, student behaviors, and classroom ecology. Simultaneous collection of these areas allows for data analysis that reveals relationships between these areas and their potential relationship to gains in student academic performance. Simultaneous documentation of these events was utilized for both providing feedback to teachers as well as exploring links between teacher instructional changes and student engagement and academic performance.

The measure consisted of 30 second intervals subdivided in half into two 15 second intervals. The first half of the 30 second interval measured classroom instructional behaviors and the 2nd half measured student engagement behaviors. Every

15 seconds alternates between instruction and engagement variables with both combined to be recorded as one 30 second interval.

Classroom instruction items. Teacher instructional behavior data consisted of items that are part of instructional components that have demonstrated effectiveness with improving middle school literacy. The content for these items were constructed based on the components of the Improving Adolescent Literacy practice guide from the Institute of Educational Sciences (Kamil et al., 2008). The IES guide is a synthesis of the research found to be effective in addressing the literacy needs of middle school students. As mentioned previously these components consist of direct and explicit instruction and practices to increase engagement while teaching vocabulary and comprehension skills and engaging in discussion around text.

Based on this document, ten objectively observable teacher and student behaviors were coded into instructional behaviors to be targeted in the consultation process. In order to target direct instruction, items were included that focus on components of scaffolded instruction and providing explicit feedback. Specifically, scaffolding was broken down into three items: (a) explanation/ modeling of strategy use, (b) guided practice, and (c) independent or partnered practice with feedback. At the start it was conceptualized that these scaffolding elements would be recorded when a teacher engaged in these practices to address solely vocabulary or comprehension skills or strategies, however, due to school administrator insistence that other areas be addressed in ELA classes (grammar, study skills, and test taking strategies) scaffolding was recorded anytime the scaffolding technique was used to address these areas as well.

Providing explicit feedback was coded whenever a teacher provided immediate feedback to a student following their response to a question or when asking a question.

In order to target instructional practices meant to increase engagement the following items were created: (a) elicit responses via random selection, (b) check for understanding, (c) written response, and (d) choral response (verbal or physical). Finally the following items were included to address effective discussion practices: (a) personal connections or building background knowledge, and (b) teacher or student explaining reasoning/thinking. Objective definitions, examples, and non-examples of all ten of these items can be seen in Table 1.

Student engagement items. Student engagement was coded as active on task, passive on task, or off task. Active on task behavior was defined as instances where a student is actively engaged in the classroom instruction. These behaviors included raising a hand to answer a question, answering a question out loud, reading aloud, writing about the content, and talking to a peer about the content. It was not considered active on task if the student was talking to a peer, writing, or generally talking about something unrelated to the class content. Passive on task was defined as instances where a student was passively engaged in the classroom instruction. These behaviors included reading silently, looking at a worksheet, listening to teacher instruction, and listening to a peer talk about class content. It was not considered passive on task if the student was reading material not related to the instruction content or looking at other objects in the room other than those that are part of instruction. Off task behavior was defined as any activity not associated with classroom instruction. These included motor behaviors such as drawing,

flipping pages, or walking around the room, verbal behaviors such as talking to another student about non-instruction content, and passive behaviors such as looking around the room, looking at materials unrelated to instruction content, and listening to peers talk about content unrelated to instruction.

Classroom instruction and student engagement observation procedure. Observations took place in a time sampling interval recording format. The observer wore head phones and listened to a recording where a tone played every 15 seconds indicating the start of each interval. Interval recording data provides a percentage of the observed time that a behavior occurred. Due to the varying nature of the behaviors under observation two types of interval recording took place. Scaffolding instruction, active engagement, and passive engagement utilized momentary time sampling (MTS) procedures. During momentary time sampling, at the tone the observer recorded if one of the scaffolding behaviors, active engagement, or passive engagement was taking place at that moment. This was due to the fact that these behaviors do not have a discrete start and stop period and have a potential to occur at a steady state (Hintze, Volpe, & Shapiro, 2008).

All other instructional behaviors, and student off task behavior, consisted of partial interval recording (PIR). Like the behaviors measured by MTS, these behaviors lack a clear beginning and end, however because they are of a more inconsistent duration and occur at a relatively smaller rate they lend themselves better to be measured by PIR (Hintze, Volpe, & Shapiro, 2008). During PIR, one or more behaviors were recorded if they occurred during any portion of the observed interval. For example, if choral

response took place during an interval where teacher behavior is being recorded then that instance will be coded. If it occurred again during that interval however the instance was still recorded only the one time.

***Interrater reliability*.** Interrater reliability was assessed with the classroom observation tool after treatment ended with all six teachers. Ten percent of observations were conducted with another trained observer. Inter-rater agreement was calculated by dividing the sum of total number of observations coded from observer A by the sum total of observer B and multiplying it by 100.

Reliability of Classroom Instruction Items. Interrater reliability of modeling ranged from .33 to 1.00, with an average of .71. Guided Practice ranged from .81 to 1.00, with an average of .91. Practice ranged from .86 to 1.00, with an average of .95. Total Scaffolding ranged from .84 to 1.00, with an average of .91. Due to high average reliability, all areas of scaffolding were retained in the analysis (See table 3 for summary of all interrater coefficients).

Random Response ranged from .46 to 1.00, with an average of .75. Feedback ranged from .50 to 1.00, with an average of .79. Check for Understanding ranged from .5 to 1.00, with an average of .71. Because of the limited range of this area (recorded instances in each observation session was either 0 or 1), this area was removed from analysis. Personalized Connections ranged from .12 to .71, with an average of .55. Due to this low reliability coefficient this area was removed from analysis. Choral Response ranged from .10 to 1.00, with an average of .72. Written Response ranged from .06 to 1.00, with an average of .72. Explaining Thinking ranged from .06 to 1.00, with an

average of .45. Due to the low reliability coefficient this area was removed from analysis. After interrater reliability was calculated the following areas remained in the analysis: random response, feedback, choral response, and written response.

Reliability of Engagement Items. Interrater reliability of total on task behavior ranged from .87 to .99, with an average of .94. Active engagement ranged from .06 to .98, with an average of .80. Passive engagement ranged from .71 to .97, with an average of .90. Off task behavior ranged from .37-.90, with an average of 0.76.

Student Reading Skills. Student reading skills were assessed via the AimsWeb MAZE measure. With this measure a student reads a passage for three minutes. Within the passage every seventh word is presented as a choice of three words. The student circles the word they think is most appropriate in the space. Scores were calculated as the total number of words selected correctly within the three minutes. MAZE scores were obtained weekly and averaged to give a class wide comprehension score. According to the National Center on Response to Intervention (2011), alternate form reliability for MAZE ranged from 0.66-0.91 depending on age and disability status and its criterion validity with the SAT-10 is 0.79.

Procedure

Teacher and student observation data were collected by the author. Observations took place three times a week during a full period of a core English language arts (ELA) class. Distribution of the MAZE occurred at the end of each week and was administered by the teacher. MAZE scoring was be done by the author. After the baseline phase the classroom observations continued three times a week. Once each week the teacher met

for a performance feedback session with the author during the teacher's prep period. This process continued for 4-5 weeks ending either at the 5 week mark or if stable level and trend of instructional practices was observed by the 4th week.

Consultation Process

In traditional instructional consultation the consultant and consultee (typically the teacher) meet to determine the potential need to intervene with a student. This study differed in that the basis for engaging in consultation was not instigated by the teacher, but rather was pre-determined by the school when determining the components of the school's professional development agenda. Another significant way this study also differed from traditional instructional consultation is that the problem addressed was not a specific student or classroom problem. Instead it was a school wide problem that involved the more general objective to improving literacy instruction at the universal classroom level. This is important to note as it is an attempt to take an instructional consultation model, originally designed to address individual student or classroom problem, and apply it to a school wide problem. Because of these differences, a slightly modified approach to the instructional consultation model was employed.

Entry and contracting. In traditional instructional consultation need for consultation is sought out by the consultee (teacher) via request to the Instructional Consultation team. In this study, the school pre-determined that the school wide problem to focus on via professional development is increasing student engagement and literacy skills. Therefore, traditional entry and contracting between the teacher and consultant was not able to be conducted. Instead all core ELA teachers were selected by

school administration to engage in the consultation process. When each identified teacher began the consultation phase the consultant and teacher met briefly to discuss the process. After the meeting, baseline collection of instructional quality, student engagement, and comprehension skills began.

Problem identification and analysis. The problem identification and analysis phase took place in the same manner as done in traditional IC (Rosenfield, 2008; Rosenfield, Silva, & Gravois, 2008). The baseline data collected for instructional quality, student engagement, and comprehension skills were presented to the teacher. The teacher and consultant then analyzed where increases could be made in the targeted instructional practices and set goals.

Intervention planning. After analyzing baseline and determining goals, the teacher and consultant discussed ways to increase these behaviors in the classroom based on the research presented in the 2008 IES Improving Adolescent Literacy instructional guide. Teachers were encouraged to identify support needs such as increases in resources, informational guides, or classroom management techniques. Where possible the consultant assisted the teacher in obtaining that support to meet his/her needs. For instance, if a teacher requested information on what strategies to implement they were provided with a copy of the IES practice guide and given the opportunity to ask specific questions of the consultant.

Treatment implementation with performance feedback. In traditional consultation, teachers implement the intervention decided upon with supports from the consultant to maintain fidelity. In this study, however, performance feedback was added

to this phase with the intention to increase the desired behavior with consistency and maintenance. Performance feedback between the teacher and consultant occurred once a week. During these sessions the consultant shared teacher instruction and student engagement data with the teacher and discussed the connection between the two. Strengths observed were also discussed followed by collaborative problem solving to address any hindrances to performing the targeted behaviors as had been decided on in the problem analysis meeting.

Resolution/Termination. In traditional IC the consultant and consultee meet after the intervention is over to determine its effectiveness. Because classroom instruction is not a treatment with a discrete end time this phase was modified. In this study treatment evaluation took place when the performance feedback phase ended. During this phase the teacher and consultant met to discuss effectiveness of the consultation practice and identify ways to ensure maintenance of the effects.

Analysis

To address all three research questions, a combined concurrent and non-concurrent multiple baseline single case design (SCD) procedure was conducted. Utilizing a combined concurrent and non-concurrent baseline is beneficial in that it addresses the potential threats to internal validity that the individual designs are subject to when conducted on their own, particularly history effects. History effects occur when events occurring concurrently with the treatment influence the change in behavior, leading to the mistaken assumption that it was a treatment effect (Kratochwill, 2010). In this study history effects could influence instructional behavior or student engagement

due to co-occurring external influences such as preparation for state or district testing or proximity to school breaks. For instance teachers may naturally teach more comprehension skills in a direct manner when preparing students for a test than they would after the test has been completed. Likewise a teacher may not teach new skills when a school break is approaching in the expectation that students might not retain the information over the break.

Non-concurrent designs when conducted alone are subject to history effects as they lack comparisons to other subjects' responses to treatment at the same time point (Harvey, May, & Kennedy2004; Kennedy, 2005). Concurrent baseline designs are arguably more stringent in protecting against history effects impacting individuals (Riley-Tillman & Burns, 2009) however because this design takes place at the same time point it is also potentially subject to history effects that may affect all of the individuals in a study period (such as school policies that impact all teachers). Combining the two designs allows for examination of potential history effects. Replicating the concurrent portion of the study at a non-concurrent time point buffers against history effects as it makes it more difficult to mistakenly assume that the change in behavior is due to an extraneous event. This is because it allows for comparison of responses to treatment at multiple time points (Harvey, May, & Kennedy, 2004). This limits the potential for history effects as it would require the confounding event to occur with multiple subjects over multiple points in time (Kratochwill et al., 2010).

In the concurrent section of the multiple baseline design 3-4 subjects began baseline data collection at staggered starting points. Then treatment (consultation with

performance feedback) was implemented and data was collected three times a week for four weeks. After treatment concluded, 3-4 more teachers began the process again in the non-concurrent combined portion of the design. In this study baseline data were collected for 3-5 data points or until a stable baseline trend was observed for student engagement data (Horner et al., 2005; Riley-Tillman & Burns, 2009). This was a priori determined to be the area for stable baseline due to practical limitations of the school setting's time constraints. Treatment phase data were attempted to be collected for 12-15 data points for teacher instruction and student engagement, however due to practical limitations the resulting number of treatment data points ranged from 9-11. For student comprehension data 3-5 data points was the goal but with practical limitations resulted in 2-5. For classrooms with only 2 data points that data was not analyzed as it did not allow for adequate analysis.

Design

Single case design has been established as a methodological approach for evaluating evidence based practices in school psychology and special education research (Gresham & Vanderwood, 2008; Horner et al., 2005). The purpose of SCD is to establish a functional relation between the independent and dependent variables (Gresham & Vanderwood, 2008; Horner et al., 2005; Riley-Tillman & Burns, 2009). In multiple baseline SCD this is demonstrated when data collected during treatment demonstrates a pattern that differs significantly from data collected at baseline and is then replicated across subjects, behaviors, settings, or other objects one wishes to generalize across (Christ, 2007; Horner et al., 2005; Riley-Tillman & Burns, 2009). SCD allows one to

focus on individuals, which is particularly useful when a small sample size is present (such as with classrooms in a particular school). This focus on individual variance allows for detailed analysis of responders and non-responders to a particular treatment. It is also a non-intrusive and feasible method in practical settings as it is not dependent on random assignment as is the case for group design (Horner et al., 2005).

There are some advantages of visual analysis SCD designs over group designs. One is that in SCD intra-subject variability becomes a source of information from which to draw conclusions while with group designs intra-subject variability is considered error (Gresham & Vanderwood, 2008). This allows for more detailed analysis of how an individual may respond to a treatment (Riley-Tillman & Burns, 2009). In addition to this SCD allows for replication of results in the design (Riley-Tillman & Burns, 2009; Gresham & Vanderwood, 2008). With group designs a complete study must be repeated to allow for replication analysis. In SCD replication can be easily conducted by repeating the treatment with a single subject or across multiple subjects. It can also be applied across different types of subjects (such as teachers teaching different subjects) or settings (such as general education and special education classrooms). This can quickly demonstrate a case for generalizability of the treatment.

Visual Analysis in SCD

The traditional approach to examining the existence of a functional relationship is via visual examination of between and across phase data (Horner et al., 2005; Riley-Tillman & Burns, 2009). With visual examination one analyses the degree to which the behavior of measurement varies within and between baseline and treatment conditions

(Gresham & Vanderwood, 2008; Riley-Tillman & Burns, 2009). Visual components analyzed include level, trend, immediacy, and variability (Horner et al., 2005; Riley-Tillman & Burns, 2009).

Change in level is evaluated by seeing if the data points in the treatment phase are significantly different than the level at baseline. In addition, to level, trend of the data must also be evaluated to determine if the trend of data in treatment condition is sufficiently different from the trend at baseline. This rules out the possibility that the trend in baseline could not have predicted the level attained during treatment. Evaluating for immediacy of change is another component for visual analysis. The more immediate the change in level and trend after implementation of the treatment phase the more secure one can be in attributing the change in the subject's behavior to the treatment. Finally, change in variability in the data between baseline and treatment phases can further assist in visual analysis of the relationship between dependent and independent variables. A decrease in variability during the treatment phase may indicate a steady influence on a subject's behavior. These four indicators all provide evidence for a functional relationship.

It has been argued that with SCD visual analysis is the preferred analysis method, over statistical analysis (Brossart, Parker, Olson, & Mahadevan, 2006). It is thought that any effect large enough for substantial attention will be visible resulting in lower error rates and conservative assumptions. However, disadvantages do exist with visual analysis. One is that subtle effects are more difficult to detect. Thus, this often results in an increased probability of Type II error (Gresham & Vanderwood, 2008). In a 2006

study Brossart and colleagues analyzed various research studies on the reliability of visual analysis when the data is not clear in level, trend, immediacy, and variability. The authors found low to moderate interrater reliability ranging from .40-.60. Trend lines have been cited as a way to help clarify ambiguous results, however occasionally trend lines can create discrepancies as well, resulting in overemphasis on trend and neglect of other visual components (Brossart, Parker, Olson, & Mahadevan, 2006). Because of the potential ambiguity visual analysis can produce statistical analysis via effect sizes has been considered in the research as an additional analysis method. It has been suggested that statistical analysis via effect sizes may be necessary when there is no stable baseline, changes in behavior due to treatment is not predictable, and when sharing ambiguous results with other professionals (Brossart, Parker, Olson, & Mahadevan, 2006; Parker, Cryer, & Byrns, 2006). Based on this, effect size analysis was also included in the analysis of research questions 1-3.

Effect Size Analysis

Calculating effect sizes is a statistical method of examining the extent of the functional relationship between independent and dependent variables. The use of effect sizes instead of traditional statistical analysis, however, is a theoretical shift in emphasis from statistical significance to practical significance (Brossart, Parker, Olson, & Mahadevan, 2006). An advantage of this is that it allows a researcher to infer the strength of the relationship between the independent and dependent variable. Multiple options exist for generating effect sizes in single case design. One option is Percent of all Non-Overlapping Data (PAND). PAND is calculated by dividing the total number of

intervention data points that overlapped baseline by the number of all data points (intervention and baseline combined). For example, if a study included 10 baseline data points and 10 intervention data points, 2 of which overlapped with the highest baseline data point, then the PAND would equal 2 divided by 20, which equals .10. That number is then subtracted from 1.0 and multiplied by 100 to compute the percentage of all data that do not overlap. In this example then PAND would equal 90%.

PAND is particularly beneficial as it can be converted to the more commonly known effect size of Phi (Parker, Hagan-Burke & Vannest, 2007; Riley-Tillman & Burns, 2009). This allows for analysis of practical effect. It also fairly simple in its data requirements as only 20 data points are necessary and it is not subject to the parametric assumptions of normality and equal variance (Parker, Hagan-Burke & Vannest, 2007, p. 196). Sufficient statistical power is also relatively easy to achieve. In a 2007 study, Parker, Hagan-Burke, and Vannest analyzed 75 multiple baseline SCD studies. Results indicated that studies with 45-96 data points resulted in acceptable power when moderate effects are expected. Phi is calculated by creating 2x2 table with the data:

Cell A: % of baseline (baseline/total)	Cell B: % overlapping data/2
Cell C: % overlapping data/2	Cell D: % of tx (tx/total)

Then using the table the following equation is applied: $\Phi = [a/(a+c)-[b/(b=d)]$ (Parker et al, 2007; Riley-Tillman & Burns, 2009). Cohen (1988) suggested that phi coefficients of .10 were small, .30 were medium, and .50 were large. Continuing from the previous

example then phi would be .8, a very large effect as corroborated by the fact that only two data points overlapped in the hypothetical 45-96 data points.

Other common effect sizes considered for this study included R^2 and Percentage of Overlapping Data (PND). However noted limitations exist with these effect sizes. R^2, which is most commonly associated with regression analysis, is highly dependent on the assumptions of linear regression. This is a problem as most SCD research does not meet these assumptions, particularly that of autocorrelation of data (Riley-Tillman & Burns, 2009). It is also unclear how the R^2 statistic should be interpreted as wide variations have been found in effect size values and practical significance classification. For instance, Brossart and colleagues (2006) found that very effective interventions had R^2 effect sizes that ranged from 0.035-0.895 while not effective interventions had a range of -0.022-0.649. This wide range and significant overlap makes it difficult to securely infer a categorical effect size or apply pre-determined effect size cut offs (such as Cohen's) to SCD research. PAND/Phi overcomes these shortcomings as the phi coefficient is a common effect size and is not subject to the assumption of autocorrelation (Riley-Tillman & Burns, 2009).

The effect size PND is similar to PAND in that it involves visual analysis of the data to calculate an effect size based on percentage of overlap between phases (Parker, Hagan-Burke & Vannest, 2007). However, one significant difference between the two methods is that PND only takes into account one baseline phase data point, the highest point, into its calculation. Utilizing one data point can be unreliable as there is high potential for outlier effects (Parker, Hagan-Burke & Vannest, 2007; Riley-Tillman &

Burns, 2009). PND is also limited in that it is not associated with a currently accepted effect size or sampling distribution resulting in the need for its own interpretation guidelines and the inability to compute a confidence interval (Parker, Hagan-Burke & Vannest, 2007; Riley-Tillman & Burns, 2009). PAND/Phi overcomes this as PAND takes into account all baseline data and when converted to Phi becomes associated with an effect size that has potential for common interpretation and sampling distributions for reliability estimation (Parker, Hagan-Burke & Vannest, 2007; Riley-Tillman & Burns, 2009). Because of this PAND/Phi was utilized for calculating effect sizes in this study.

However, limitations do exist with PAND that is common to all effect sizes calculated with SCD research. One is lack of sensitivity with higher effects (Parker, Hagan-Burke & Vannest, 2007; Riley-Tillman & Burns, 2009). For instance, when there is no overlap between baseline and treatment an effect size will result in the largest possible score, regardless of the distance between the levels of baseline and treatment. Therefore, an intervention with a change in level of two points would be considered just as effective as an intervention with a level change of 40. In addition to this effect sizes do not take baseline trend into account (Parker, Hagan-Burke & Vannest, 2007; Riley-Tillman & Burns, 2009). Because of this, it is impossible to know by effect size alone if the change in behavior was due to the intervention or if it would have occurred based on how the behavior was naturally increasing/ decreasing at baseline. These shortcomings can be addressed however when combined with visual analysis as one can visually assess and confirm if effect sizes are a true reflection of the data and rule out these potential confounding influences. Because of this, both visual analysis and effect sizes were taken

into consideration in the analysis. Primary emphasis was placed on visual analysis with effect size analysis as a supplementary analysis (see table 2 for results analysis summary).

Results

Research Question 1

Research question 1 asked: To what extent does performance feedback increase teacher use of targeted instructional practices? The targeted instructional practices included scaffolding, feedback, written response, random response, and choral response. Each practice was evaluated first via visual analysis then with effect size analysis.

Scaffolding. Figure 1a and Figure 1b show that a positive change in level of scaffolding was clearly present with Teacher E and slightly present with teacher D. Of the components that make up total scaffolding a significant increase in level was seen in guided practice. Modeling and Practice levels remained the same across baseline and treatment phases. Teacher D demonstrated a delayed increase (after two treatment sessions) in level of total scaffolding. A positive change in trend was seen in Teacher D. Immediacy of change was only seen in Teacher E. Decrease in variability was seen in Teacher F.

Taken together it can be interpreted that significant change in scaffolding behavior was only seen in teachers E and D. Teacher E had a clear increase in level that occurred immediately after the first application of performance feedback. A change in variability and trend was not seen in this teacher, however this can be explained by ceiling effects and relative lack of variability seen in baseline. Teacher D showed some

increase in level and trend. However the effects of this are questionable as the change was not immediate and had high variability above and below baseline.

Teacher A showed no change in level, trend, or variability. Teacher B did show an increase in scaffolding trend, however this increase was consistent with the trend at baseline, making it more likely that change was due to natural behaviors instead of application of performance feedback. Teacher C demonstrated a decrease in level of scaffolding during treatment which was seen after first application of performance feedback, a slight increase in variability, and no change in trend. Finally Teacher F showed a decrease in variability after first application of performance feedback but no change in trend or level.

Effect size analysis of total scaffolding indicated a medium effect (Cohen, 1988). Percent of all non-overlapping data (PAND) points was 54.88%. This resulted in a Phi of .34 (See table 4 for all effect size statistics). For modeling PAND was 37.80% resulting in a Phi of .19, indicating a small effect. For guided practice PAND was 57.31% resulting in a Phi of .36, indicating a medium effect size. Finally for practice PAND was 53.65% resulting in a Phi of .33, indicating a medium effect size.

Feedback. Figures 2a and 2b show that level of feedback increased with only one teacher (Teacher E). Immediacy was seen with this teacher in regards to level increase. No observable change in trend or variability was seen for any teachers. Effect size analysis indicated a medium effect. PAND for feedback was 56.01%. This resulted in a Phi of .35, indicating a medium effect.

Written Response. Figures 3a and 3b show that level of written response increased for two teachers. Level increased markedly in one teacher (Teacher D) and slightly for another (Teacher B). Change in trend was also seen in both teachers. Immediacy of change in level and trend was seen in Teacher B with change occurring two data points later for Teacher D. Change in variability was seen in one teacher (Teacher F). Effect size analysis indicated a medium effect. PAND for written response was 61%. This resulted in a Phi of .40, indicating a medium effect.

Random Response. Figures 4a and 4b show that level of random response increased for four teachers, three from the second concurrent baseline phase (Teachers A, D, E, and F). Immediacy of this change in level was seen in all four of these teachers as well. Neither an increase in trend nor a decrease in variability was seen in any teachers. Effect size analysis indicated a medium effect. PAND for random response was 62.2%. This resulted in a Phi of .42, indicating a medium effect size.

Choral Response. Figures 5a and 5b show that level of choral response did not change for any teachers. Change in trend was seen in teacher E. Change in variability was seen in teacher D. Immediacy of these changes was seen in both teachers. Effect size analysis indicated a small effect. PAND for choral response was 37.8%. This resulted in a Phi of .20, indicating small effect.

Research Question 2

Research question 2 asked: To what extent does performance feedback with a teacher increase student engagement? Elements of engagement included total engagement, active engagement, passive engagement, and off task behaviors. As with the

instructional practices, change in engagement was evaluated first via visual analysis then with effect size analysis.

Figure 6a and 6b show that level of total engagement increased with teachers A, B, D, and F. An increase in trend was not, however lack of positive change could be attributed to a ceiling effect. Immediacy of these effects was seen in all four teachers. Decrease in variability was seen in Teacher B, however lack of change in variability could be attributed to limited variability seen at baseline in teachers A, D, and F.

Increase in level of active engagement was seen in teachers A, D, and E. Of these three teachers an increase in trend was seen in teachers A and E. Immediacy of these changes was seen in only with Teacher A. Decrease in variability was not seen. Increase in level of passive engagement was seen in teachers A and F. Increase in trend was seen in Teacher A. Decrease in variability was seen in Teacher F. Immediacy of these effects was seen in both teachers. Off task behavior had limited variability and therefore it was unable to determine if any significant changes in behavior in this domain occurred.

Effect size analysis for total engagement indicated a large effect for total engagement. PAND for total on task engagement was 72%. This resulted in a Phi of .53. PAND for active engagement was 62.19%, resulting in a Phi of .42, indicating a medium effect size. PAND for passive engagement was 52.43%, resulting in a Phi of .32, indicating a medium effect size. Finally, PAND for off task behavior was 58.53, resulting in a Phi of .38, indicating a medium effect size.

Research Question 3

Research question 3 asked: To what extent does performance feedback with a teacher increase student comprehension skills as measured by the general outcome measure MAZE? As with the previous two variables, change in comprehension was evaluated first via visual analysis then with effect size analysis.

Figures 7a and 7b show that average level of classroom comprehension skills, as measured by AimsWeb Maze, increased for only one teacher, Teacher E. A change in trend was seen in teacher F. A change in trend could also potentially be seen in Teacher E, however due to there being only two data points in the treatment phase, it can only be said for certain that change was seen for teacher F. This change in trend was immediate. Change in variability was also not seen in any of the teachers. Effect size analysis indicated a medium effect. PAND for comprehension was 65.83%. This resulted in a Phi of .49.

Discussion

The purpose of this study was to investigate the potential effect of consultation with performance feedback on literacy instruction, classroom engagement, and student comprehension skills. Significant positive relationships were found between performance feedback and the variables of random response and overall student engagement. These significant results are consistent with the current literature as it indicates that performance feedback is an effective addition to the consultation process.

In addition to supporting the current literature, this study sought to expand on it by applying the consultation with performance feedback process to core classroom

instruction for all students. Previous studies have examined traditional consultation processes in limited settings with individuals or small groups of students with targeted problems. This study also expanded on the literature by utilizing consultation as a method of preventing of academic problems. Traditional consultation models take a reactionary approach in which consultation is applied after a problem arises (Kratochwil & Bergan, 1990). The significant results from this study indicate that consultation with performance feedback is a promising mechanism in the areas of whole classroom instruction and preventative action that warrants further study.

Although consultation with performance feedback yielded promising results, not all variables measured had significant positive results. Many of the instructional variables measured did not demonstrate any relationship with performance feedback. The same was true for specific areas of student engagement and classroom comprehension skills. The results for all variables and their significance will be discussed in further detail in the following sections. Theories for non-significant results will also be discussed, followed by study limitations and implications for practice.

Performance Feedback and Literacy Instruction

Research question one explored the influence of performance feedback on instructional behaviors. Results indicated that the only behavior that appeared to be influenced by performance feedback was random response. Visual analysis results yielded no relationship between performance feedback and scaffolding, feedback, written response, and choral response. Effect size analysis indicated small to medium effects for these various areas. However, due to limitations in effect size analysis with single case

design, particularly lack of control for baseline trend, only the effect sizes for areas meeting visual analysis standards were considered for further interpretation (Parker, Hagan-Burke & Vannest, 2007; Riley-Tillman & Burns, 2009). In considering that a positive result was found in one of the targeted instructional behaviors, it is worthwhile to analyze the findings in regards to random response and later follow with a discussion of what may have contributed to the lack of significant effects in the other targeted instructional behaviors.

Before discussing the positive effects, it is important to review the criteria for a significant effect in visual analysis of single case design studies. The presence of an effect using visual analysis is determined by a change in level, change in trend, diminishing variability of data, and/or the relative immediacy of these changes from baseline to treatment phases (Harvey, May, & Kennedy, 2004; Kratochwill et al., 2010). This study utilized a mixed concurrent and non-concurrent multiple baseline design with six subjects total and three in each concurrent phase. Harvey and colleagues (2004) suggest that in order to determine the presence of an effect within this kind of design, the effect should be seen for two subjects within each concurrent phase and for at least one subject in the subsequent non-concurrent phase.

In regards to random response, a change in level and immediacy was seen in four of the six teachers with three teachers represented from the second concurrent phase and one from the first. A change in trend and/or decrease in variability was not seen; however, the nature of the variable being measured can account for this lack of change. Specifically, the variability at baseline was already very low in most teachers, which

makes it difficult or impossible to demonstrate a decrease during treatment. In addition, some teachers' variability was found to increase during treatment phase. Anecdotal data from conversations with teachers during consultation meetings suggest that this was due to trying to balance introduction of random response along with the other elements targeted, indicating that the interrelationship of these instructional variables should be taken into account.

There are some theoretical explanations for the increase of random response as a result of performance feedback. First, it appears that random response was the easiest to understand and comprehend among all the targeted instructional behaviors. This was substantiated by informal teacher report during consultation meetings. Utilizing random response only required changing the questioning format from asking for student volunteers to selecting students at random. Also, unlike some of the other instructional behaviors (e.g.: scaffolding), random response was a simple behavior that was easily explained and comprehended during the consultation sessions. Therefore, it is logical to conclude that if any instructional behavior were to be remembered and adopted, it would be random response (Gresham, 1989; Rathvon, 2008; Roach & Elliott, 2008).

The nature of the random response being simple, easy to understand, and easy to implement may explain why it was the only instructional behavior that had a significant effect. This explanation may also account for why the other instructional behaviors did not. Simplicity may be a factor that influences a change in teacher instructional practices. Support for simplicity can be found in the intervention literature. Intervention research suggests that one element of best practices for choosing an intervention is to choose one

that is capable of being taught in consultation format (Rathvon, 2008). In addition, intervention fidelity research has found that an intervention is less likely to be implemented if it is complex (Gresham, 1989; Roach & Elliott, 2008). While the majority of this research is based on individual or small group interventions, it is possible that this concept can apply to classroom instruction as well. Random response is an instructional behavior that can be quickly taught in short consultation meetings and is simple to introduce into instructional routines. More complex instructional behaviors (e.g.: scaffolding, providing feedback, providing opportunities for written or choral response) may require more time for development and practice than can be addressed in the short weekly consultation settings utilized in this study. Future research is needed to determine which instructional behaviors can be effectively influenced in a consultation format and which may require more intensive training.

It should be noted here that other explanations for the lack of significant results for the majority of instructional behaviors may exist due to specific elements encountered during the course of the study. Additional explanations for these non-significant effects will be analyzed in more detail following the discussion of effects found for the remaining research questions.

Performance Feedback and Student Engagement

Research question two explored the influence of performance feedback on student engagement behaviors. In regards to the area of total engagement, four of the six classrooms showed change in one or more necessary components in visual analysis, with two classrooms within each concurrent phase. In addition to this, effect size analysis

indicated that there was a large effect (Φ=0.53). Therefore, results indicate a potential relationship between performance feedback and increasing overall student engagement.

Despite these results, the data are less clear in regards to the specific areas of engagement. In the area of active engagement, three teachers demonstrated classrooms with increased levels. However, only two demonstrated an increase in trend, both of which were very small. In addition, only one showed immediacy of these effects. None demonstrated effects on variability. Some of these findings may be related to the ceiling effects found in total engagement. Overall, these results suggest that performance feedback may have some influence on active engagement. However, the influence of performance feedback on engagement overall is much more clear. In regards to passive engagement, performance feedback did not appear to have an effect. Specifically, only two teachers demonstrated a change in level and only one demonstrated a change in trend. Therefore, while there is some indication that the increase in overall engagement may be due to an increase in active engagement, conclusive statements cannot be made regarding the effects of performance feedback on specific areas of engagement.

The findings of increased overall engagement and increased random response are interesting to note. The presence of these results may be a result of the interrelated nature of these two variables. In the literature, random response is one technique that can be used to increase engagement via active responding (Gettinger & Ball, 2008; Rathvon, 2008). Therefore, it is possible that increasing just this one instructional variable contributed to an increase in engagement. Also, since random response is a tool utilized to increase the number of students that actively respond to a prompt, this finding supports

the possibility that active engagement was the area within overall engagement that increased. Overall, this finding is interesting to explore as it provides some indication that the increase in engagement was possibly due to the increase in random response. However, it must be noted that the type of analysis utilized in this study only allows for a hypothesis of a potential connection rather than support for causality.

Performance Feedback and Student Comprehension Skills

The third research question of this study explored the relationship between performance feedback with a teacher and its impact on classroom comprehension skills. No effect was found for this variable as evidenced by a change in level in only one teacher and change in trend in only one other. There are a few possible explanations for the lack of effects in this area. One explanation could lie in the lack of results seen in the majority of the teaching behaviors. Due to the fact that student comprehension was a variable hypothesized to change as a result of teacher implementation of evidence based teaching practices, a lack of implementation of these practices would theoretically lead to a lack of change in classroom comprehension. Another possible explanation for the lack of change in classroom comprehension is that other factors known to influence learning were not addressed. The instructional literature indicates that variables such as environment and curriculum are other important areas that can be altered to improve student outcomes (Hunley, 2008). Due to the fact that these areas were not addressed in this study, the impact of the target instructional variables on student outcomes may have been limited.

A third notable reason for the lack of effect on classroom comprehension could lie in the nature of the assessment tool used to measure comprehension skill growth. The MAZE is a general outcome measure with much research establishing its validity and reliability in use with students in sixth grade and above (Jenkins & Jewell, 1993). A drawback to the measure, however, is that it is not very sensitive to skill growth (Fuchs & Fuchs, 1992; Tincha et al., 2009; Tolar et al., 2012). In their 1982 study, Fuchs and Fuchs found that the weekly rate of improvement across grades 1 through 6 over 15 weeks was 0.39 correct words per week. Tincha and colleagues (2009) conducted a similar study with eighth-grade students. Over 10 weeks, growth rates ranged from 0.41 to 1.29 correct words per week with lower achieving students improving at the lower rate. Taken together, these studies indicate that the MAZE measure has low sensitivity to growth especially for students who are lower achieving.

In this study, only three to four weeks were assessed, which means that the expected average growth from baseline would range between 1.17 to 5.16 words. Taking testing error into account, this leaves very little room for growth to be observed. In addition, the school in which the classrooms were nested showed that approximately 60% of students performed in the at-risk range and 30% performed in the some risk range on the MAZE. Based on the work by Tincha and colleagues (2009), this indicates that the rate of growth would be expected to be at the lower end of the expected range. Therefore, the lack of effect found for comprehension in this study could be due to the low sensitivity of the measure, which may have been even more exaggerated by the overall low literacy achievement of the students in the school.

Summary of Significant Effects

Overall, the results of this study support a relationship between performance feedback and two of the measured behaviors: random response and overall classroom engagement. Specially, there is evidence for a medium relationship between performance feedback and teacher use of random response as well as for a strong relationship between performance feedback and overall engagement. Possible explanations for the increase in these elements included the relative simplicity of implementing random response and the theoretical relationship between random response and engagement. In addition to these effects, the lack of effects for certain outcomes is also worth analyzing. As mentioned earlier, elements unique to the setting of this study may have contributed to the lack of effects. These elements are worth exploring as it may shed light on potential moderators to the effectiveness of performance feedback in the classroom and school-wide setting.

Discussion of Non-Significant Effects

It should be noted that this part of the discussion is analysis conducted after the conclusion of the consultation process in the study. However, it is worth discussing as it draws attention to areas of future research in regards to possible moderating circumstances to the effectiveness of performance feedback as part of school-wide improvement efforts.

One of the primary influences that could have impacted the lack of effects in certain areas was teacher resistance to school improvement efforts and the consultation

process. Support of leadership and effective power relationships during consultation are two areas that have been cited as potential variables that affect resistance to school improvement and consultation efforts (Datnow & Castellano, 2001; Erchul & Raven, 1997; Erchul & Martens, 2002). In regards to this study, these two important areas were lacking in the consultation process.

Impact of Leadership. The school reform literature has found that one crucial element to effective school improvement the presence of consensus and commitment from all necessary stakeholders (Grimes, Kurns, & Tilly, 2006; McGlinchey & Goodman, 2008). Much of the school reform literature indicates that support from leadership is crucial for change to occur (Datnow & Castellano, 2001). At the beginning of the study, consensus building took place over the summer by creating a plan for improvement with all stakeholders (principal, department heads, and instructional leadership team). However, once the school year started, changes were made to the school's leadership team resulting in a new principal and operations administrator who were unfamiliar with the improvement plan. As a result, the consensus built at the beginning of the process and the subsequent implementation efforts were affected.

The lack of consensus may have impacted other elements crucial to school-wide improvement planning. Improvement efforts depend on implementation with enthusiasm, commitment, collaboration, and consistency as well as with fidelity to the program (Knoff, 2008). The new administration, unfamiliar with and not committed to the previously established plan, began the school year by implementing their own plan for reform. This lead to a lack of consistency in the enforcement of the improvement plan,

which subsequently resulted in very low fidelity to the administrator's roles and responsibilities as indicated in the improvement plan previously established over the summer. For example, a key component of the improvement plan was to organize and implement inter-departmental learning walks once a month. However, the learning walks only took place once over the entire instructional year.

Lack of implementation support from the administration in these crucial areas likely affected teacher participation in the consultation process. The missing element of consensus building led to a lack of consistency and fidelity regarding the improvement plan by administration and staff. As a result, many teachers reported a significant drop in enthusiasm and commitment to the agreed upon improvement components, which included the consultation component.

Impact of Power Relationships. The lack of strong leadership supporting the improvement plan and the resulting inconsistent support from teachers in plan implementation are not the only factors that may have led to teacher resistance to the consultation process. The interaction between teachers and the consultant conducting performance feedback is also a potential factor to consider, especially in regards to the issue of power. Power, as defined as the ability of the consultant to influence the behavior of the consultee (Erchul & Raven, 1997), is a commonly cited element necessary to minimize resistance. When certain elements of power are not established, teacher resistance often increases (Gonzalez, Nelson, Gutkin, & Shwery, 2004). There are three proposed models that are often utilized by consultants in order to improve power in consultation (Erchul & Martens, 2002). The first is the power-coercive approach. This

approach assumes that a consultee will change when presented with political, social, or economic punishments or rewards (Erchul & Martens, 2002; Erchul & Raven, 1997). The power-coercive approach is often limited because it is dependent on the consultant either having a position of power over the consultee (which is uncommon when a consultant is in the position of staff rather than administrator) or having a third party to which the consultant can turn, which effectively removes power from the consultant to that third party (Erchul & Martens, 2002).

The second model is the empirical-rational approach. This approach assumes that the consultee will change when given information that influences them to believe that change is rational and logical (Erchul & Martens, 2002). This model is utilized is by the consultant distributing information to the consultee and forming a relationship to the consultant as a kind of expert (Erchul & Martens, 2002, Erchul & Raven, 1997). This model has some effectiveness; however, there is some belief that it is naïve in its assumption that people will be motivated to change based solely on the acquisition of new knowledge (Erchul & Martens, 2002).

Due to the weaknesses present in the previous models, a third model has been proposed which is called the normative-reactive approach. This model assumes that people depend on both knowledge and social factors in order to decide whether to change (Erchul & Martens, 2002). Consultants who use this model approach the consultation process by attempting to provide information as well as influence a consultee's feelings and values by appealing to personal norms and personal relationships (Erchul & Martens, 2002). This model is implemented by applying both referent power, which is influence

that is based on a consultee's identification with a consultant and/or their desire for such identification, and expert power, which is influence based on the perception that the consultant possesses knowledge or expertise in the area (Erchul & Raven, 1997; French & Raven, 1959).

Based on this information, a normative-reactive approach was chosen to be utilized in this study. Specifically, attempts were made to provide information and build personal relationships with the teachers. To build expert and informational power, at the start of the consultation process teachers were provided with copies of the IES adolescent literacy practice guide. Then, throughout the process, articles were provided and informal information was shared to underscore why particular behaviors were being monitored and how to increase them. This attempt was often successful as teachers reportedly consumed the information and often asked relevant follow-up questions related to the information provided. To build referent power, the consultant made efforts to engage in social interactions with the teachers outside the classroom, engage in personal conversations at the start and end of consultation sessions, and utilize foot-in-the-door techniques (i.e. making small requests and building to larger ones) when significant changes in classroom teaching behaviors was necessary (Martens, Kelly, & Diskin, 1996). This attempt for referent power was less successful.

The overall climate of the school provided limited social interactions that could be had with the teachers. Teachers rarely left their classrooms for leisure time, which made it difficult for the consultant to socialize and build relationships with the teachers outside of the consultation setting. This inability to fully make a social connection with the

teachers caused the power model utilized in the study to resemble the empirical-rational approach. As previously mentioned, this approach is less likely to create influence and overcome resistance (Erchul & Martens, 2002) when compared to the normative-reactive approach. Therefore, it is likely that the influence of performance feedback could not overcome the climate of resistance present at the school.

Limitations

Similar to all studies, limitations were present in this study that limited the analysis and interpretation of effects. One major limitation was the lack of a steady baseline for some of the variables observed. Due to practical time constraints, it was not possible to wait until stable baseline was present in all areas. In order to address practical constraints, but still strive for the stable baseline standard, it was determined a priori that treatment would begin once total engagement became stable. As a result for some teachers, examining the treatment effect in all variables was difficult or, in some circumstances, impossible.

Another notable limitation in this study is the reliability of the observation instrument. Due to the fact that the observation instrument was created specifically for the study, it did not go through rigorous process of establishing reliability and validity. In an attempt to address this issue, the measure was piloted with two teachers prior to the beginning of the study. During the pilot process, changes were made to the instrument in order to better operationally define variables. In order to address reliability, interrater reliability was calculated. However, not all areas demonstrated adequate reliability coefficients. The areas of checking for understanding, personalized connections, and

explaining thinking all yielded reliability coefficients below .70. As a result, these areas were removed from the analysis. Further studies on the reliability of this instrument in different settings are necessary to determine if there is adequate reliability for these areas.

In regards to validity, the instrument utilized reputable research to establish content validity. The Institute of Educational Sciences Adolescent Literacy Practice Guide (Kamil et al., 2008) consists of a review of research that meet rigorous research standards. The instructional variables measured by the observation tool were selected from the recommendations in this document. When studied independently, these areas have shown positive effects on literacy outcomes, which creates a strong case for content validity. However, predictive validity of the instrument cannot be assumed because it is unknown if particular levels of these areas predict improved achievement. Future research with this instrument would be beneficial to not only improve the validity of the instrument, but also to help establish crucial levels of these instructional variables to improve adolescent literacy.

Conclusion and Implications

This study explored one potential way through consultation with performance feedback. Results of the study indicated that this method had an influence on teacher use of random response and overall student engagement. However, due to procedural and methodological limitations, it is unclear if this form of consultation can have an influence on other areas crucial to adolescent literacy development. The current study extended the performance feedback literature by applying it to core instruction in a whole classroom setting rather than to a targeted problem among individual or groups. It also contributed

to the literature by utilizing the process with a goal of preventing problems rather than addressing problems after they arise. Future research with this method is needed to help determine the optimal way to increase evidence-based adolescent literacy instruction and student literacy skill development.

Due to the focus of schools on utilizing multi-tiered systems as well as the relatively recent focus on the implementation of common core standards, effective core literacy instruction is becoming a crucial piece for educators to evaluate and improve. The results of this study have important implications for school psychologists. School psychologists are in a unique position because they are often highly involved in the multi-tier system process and provide consultation to address learning needs at all levels. As schools continue to progress into a multi-tier focus, and for those schools where weaknesses in core instruction are present, school psychologists will likely find themselves increasingly involved in procedures to intervene in classroom instructional practices. Research is needed to determine the optimal ways for psychologists to assist teachers in implementing evidence-based classroom literacy instruction.

References

Adams, G. (1996). Focus: What was project follow through? *Effective School Practices, 15*(1), http://darkwing.uoregon.edu/~adiep/ft/adams.htm

Alberto, P. A., & Troutman, A. C. (2008). Applied behavioral analysis for teachers. Columbus, OH: Pearson.

Ardoin, S.P., Witt, J.C., Connell, J.E., & Koenig, J.L. (2005). Application of a three tiered response to intervention model for instructional planning, decision making, and the identification of children in need of services. *Journal of Psychoeducational Assessment, 23*(4), 362-380.

August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for english language learners. *Learning Disabilities Research & Practice, 20*(1), 50-57.

Batt, E.G. (2010). Cognitive Coaching: A critical phase in professional development to implement sheltered instruction. *Teacher and Teacher Education, 26*, 997-1005.

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), 150–76.

Beck, I. L., Perfetti, C. A., & McKeown, M. G. (1982). Effects of long-term vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, *74*(4), 506–21.

Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction, 2*(2), 91–130.

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*(1), 31–42.

Brett, A., Rothlein, L., & Hurley, M. (1996). Vocabulary acquisition from listening to stories and explanations of target words. *Elementary School Journal*, *96*(4), 415–22.

Brossart, D.F., Parker, R.I., Olson, E.A., & Mahadevan, L. (2006). The relationship between visual analysis and five statistical analyses in a simple AB single-case research design. *Behavior Modification, 30*(5), 531-563.

Burns, M.K., VanDerHeyden, A.M., & Boice, C.H. (2008). Best practices in delivery of intensive academic interventions. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (pp. 1151-1162), Bethesda: The National Association of School Psychologists.

California Department of Education. (2007). *Technical Report for the California English Language Development Test (CELDT).* Monterey: California Department of Education by CTB/McGraw-Hill LLC.

Carlisle, J.F. & Berebitsky, D. (2010). Literacy coaching as a component of professional development. *Reading and Writing, 24*, 773-800.

Carlisle, J.F., Cortina, K.S., & Katz, L.A. (2011). First grade teachers' response to three models of professional development in reading. *Reading and Writing Quarterly, 27*(3), 212-238.

Carlo, M.S., August, D., McLaughlin, B., Snow, C.E., Dressler, C., Lippman, D.N., Lively, T.J., & White, C.E.(2004). Closing the gap: Addressing the needs of english-language learners in bilingual mainstream classrooms. *Reading Research Quarterly, 39*(2), 188-215.

Christ, T.J. (2007). Experimental control and threats to internal validity of concurrent and nonconcurrent multiple baseline designs. *Psychology in the Schools, 44*(5), 451-459.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ Erlbaum.

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Educational Policy Analysis Archives, 8. Available at http://epaa.asu.edu/epaa/v8n1*.

Datnow, A. & Castellano, M.E. (2001). Managing and guiding school reform: Leadership in success for all schools. *Educational Administration Quarterly, 37*(2), 219-249.

DeCecco, V.M. & Gleason, M.M. (2002). Using graphic organizers to attain relational knowledge from expository text. *Journal of Learning Disabilities, 35*(4), 306-320.

Desimone, L.M., Porter, A.C., Garet, M.S., Yoon, K.S., & Birman, B.F. (2002). Effects of professional development on teachers' instruction: Results from a three year longitudinal study. *Educational Evaluation and Policy Analysis, 24* (2), 81-112.

Duchane, E.L., Jolivet, 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.

DuPaul, G.J., Stoner, G., & Reilly, M.J. (2008). Best practices in classroom interventions for attention problems. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (pp. 1421-1438), Bethesda: The National Association of School Psychologists.

Echevarria, J., Richards-Tutor, C., Chinn, V.P., & Ratleff, P.A. (2011). Did they get it? The role of fidelity in teaching english learners. *Journal of Adolescent and Adult Literacy, 54*(6), 425-434.

Echevarria, J., Short, D., & Powers, K. (2006). School reform and standards based education: A model for english language learners. *The Journal of Educational Research, 99*(4), 195- 211.

Erchul, W.P., Grissom, P.F., & Getty, K.C. (2008). Studying interpersonal influence within school consultation. In W. Erchul & S. Sheridan (Eds.), *Handbook of Research in School Consultation* (pp.293-322), New York: Lawrence Erlbaum Associates.

Erchul, WP, Martens, BK. (2002). *School consultation: Conceptual and empirical bases of practice- 2nd edition*. Boston and New York: Kluwer Academic/Plenum Publishers; 2002.

Erchul, W.P. & Raven, B.H. (1997). Social power in school consultation: A contemporary view of french and raven's bases of power model. *Journal of School Psychology, 35*(2), 137-171.

Erchul, W. & Sheridan, S. (2008) *Handbook of Research in School Consultation*. New York: Lawrence Erlbaum Associates.

Frank, J.L. & Kratochwill, T.R. (2008). School based problem-solving consultation: Plotting a new course for evidence based research and practice in consultation. In W. Erchul & S. Sheridan (Eds.), *Handbook of Research in School Consultation* (pp.13-30), New York: Lawrence Erlbaum Associates.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review*, *21,* 45–58.

Fuchs, L.S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research & Practice, 13*(4), 204-219.

Gajira, M., Jitendra, A.K., Sood, S. & Sacks, G. (2007). Improving comprehension of expository text in students with LD: a research synthesis. *Journal of Learning Disabilities,40*(3), 210-225.

Garet, M.S., Cronen, S., Eaton, M., Kurki, A., Ludwig, M., Jones, W., Uekawa, K., Falk, A., Bloom, H., Doolittle, F., Zhu, P., & Sztejnberg, L. The Impact of Two Professional Development Interventions on Early Reading Instruction and Achievement (NCEE 2008-4030). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Garet, M.S., Porter, A.C., Desimone, L., Birman, B.F., & Yoon, K.S. (2001). What makes professional development effective?: Results from a national sample of teachers. *American Educational Research Journal, 38*(4), 915-945,

Gersten, R., Baker, S.K., Shanahan, T., Linan-Thompson, S., Collins, P., & Scarcella, R. (2007). *Effective Literacy and English Language Instruction for English Learners in the Elementary Grades: A Practice Guide* (NCEE 2007-4011). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides

Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of the research. *Review of Educational Research*, *71*(2), 279–320.

Gettinger, M. & Ball, C. (2008). Best practices in increasing academic engaged time. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (pp. 1043-1058), Bethesda: The National Association of School Psychologists.

Gonzalez, J.E., Nelson, J.R., Gutkin, T.B., & Shwery, C.S. (2004). Teacher resistance to school-based consultation with school psychologists: A survey of teacher perceptions. *Journal of emotional and behavioral disorders, 12*(1), 30-37.

Gravois, T.A. & Rosenfield, S.A. (2002). A multi-dimensional framework for evaluation of instructional consultation teams. *Journal of Applied School Psychology, 19*(1), 5-29.

Gravois, T.A. & Rosenfield, S.A. (2006). Impact of instructional consultation teams on the disproportional referral and placement of minority students in special education. *Remedial and Special Education, 27*(1), 42-52.

Greenwood, C.R., Arreaga-Mayer, C., Utley, C.A., Gavin, K.M., & Terry, B.J. (2001). Classwide peer tutoring learning management system: Applications with elementary-level english language learners. *Remedial and Special Education, 22*(1), 34-47.

Greenwood, C.R., Carta, J.J., Kamps, D., & Delquadri, J. (1995). *Ecobehavioral Assessment System Software.* Kansas City, KS: Juniper Gardens Children's Center.

Greenwood, C.R., Maheady, L. & Delquadri, J. (2002). Classwide peer tutoring programs. In M.R. Shinn, H.M. Walker & G. Stoner (Eds.) *Interventions for Academicand Behavior Problems II: Preventatative and Remedial Approaches* (611-650). Bethesda, MD: NASP.

Gresham, F.M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review, 18,* 37-50

Gresham, F.M., MacMillan, D.L., Beebe-Frankenberger, M.E., & Bocian, K.M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented?. *Learning Disabilities Research & Practice, 15* (4), 198-205.

Gresham, F.L. & Vanderwood, M. (2008). Quantitative research methods and designs in consultation. W. Erchul & S. Sheridan (Eds.), *Handbook of Research in School Consultation* (pp.63-88), New York: Lawrence Erlbaum Associates.

Grimes, J., Kurns, S., & Tilly, W.D. (2006). Sustainability: An enduring commitment to success. *School Psychology Review, 35*(2), 224-244.

Guskey, T.R. (2002). Professional development and teacher change. *Teachers and Teaching:Theory and Practice, 8*(3), 381-391.

Hansen, J. & Pearson, P.D. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth grade readers. *Journal of Educational Psychology, 75*(6), 821-829.

Harvey, M.T., May, M.E., & Kennedy, C.H. (2004). Nonconcurrent multiple baseline designs and the evaluation of educational systems. *Journal of Behavioral Education, 13*(4), 267-276.

Hintze, J,M., Volpe, R.J., & Shapiro, E.S. (2008). Best practices in the systematic direct observation of student behavior. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (pp. 319-336), Bethesda: The National Association of School Psychologists.

Hock, M. & Mallard, D. (2005). Reading comprehension strategies for adult literacy outcomes. *Journal of Adolescent & Adult Literacy, 49*(3), 192-200.

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.

Hunley, S. (2008). Best practices for preparing learning space to increase engagement. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (pp. 813-826), Bethesda: The National Association of School Psychologists.

Ikeda, M.J., Neessen, E., Witt, J.C. (2008). Best practices in universal screening. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (pp. 103-114), Bethesda: The National Association of School Psychologists.

Jenkins, J.R. & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children, 59* (5), 421-432.

Jenkins, J. R., Matlock, B., & Slocum, T. A. (1989). Two approaches to vocabulary instruction: The teaching of individual word meanings and practice in deriving word meaning from context. *Reading Research Quarterly*, *24*(2), 215–35.

Jitendra, A., Cole, C., Hoppes, M., & Wilson, B. (1998). Effects of a direct instruction main idea summarization program and self-monitoring on reading comprehension of middle school students with learning disabilities. *Reading and Writing Quarterly*, *14*(4), 379–96.

Jordan, H.R., Mendro, R., & Weerasinghe, D. (1997). *Teacher effects on longitudinal student achievement: A preliminary report on research on teacher effectiveness*. Paper presented at the National Evaluation Institute, Indianapolis, IN.

Juel, C. (1988). Learning to read and write: A longitudinal study of fifty-four children from first through fourth grades. *Journal of Educational Psychology*, *80*, 437-447.

Kaiser, L., Rosenfield, S.A. & Gravois, T.A. (2009). Teachers' perception of satisfaction, skill development, and skill application after instructional consultation services. *Journal of Learning Disabilities, 42*(5), 444-457.

Kamil, M. L., Borman, G. D., Dole, J., Kral, C. C., Salinger, T., and Torgesen, J. (2008). *Improving adolescent literacy: Effective classroom and intervention practices: A*

Practice Guide (NCEE #2008-4027). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc.

Kelly, J.G., Lesaux, N.K., Kieffer, M.J., & Faller, E.S. (2010). The effectiveness and ease of implementation of an academic vocabulary intervention for linguistically diverse students in urban middle schools. *Reading Research Quarterly, 45*(2), 196-228.

Kinkead, S. (2007). *Improving instruction through coaching.* Olympia, WA: The Center For Strengthening the Teacher Profession, Superintendent of Public Instruction, Washington.

Knoff, H.M. (2008). Best practices in implementing statewide positive behavioral support systems. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (pp. 749-764), Bethesda: The National Association of School Psychologists.

Kratochwill, T.R., & Bergan, J.R. (1990). *Behavioral consultation in applied settings: An individual guide.* New York: Plenum.

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

Kuhn, M.R. & Stahl, S.A. (1998). Teaching children to learn word meanings from context: A synthesis and some questions. *Journal of Literacy Research, 30*(1), 119-138.

Lopez, E.C. & Nastasi, B.K. (2008). Section commentary: An integrative view of process/outcome research from selected models of consultation. In W. Erchul & S. Sheridan (Eds.), *Handbook of Research in School Consultation* (pp.247-168), New York: Lawrence Erlbaum Associates.

Martens, B.K., Kelly, S.Q., & Diskin, M.T.(1996). The effects of two sequential request strategies on teachers' acceptability and use of a classroom intervention. *Journal of Educational and Psychological Consultation, 7*(3), 211-221.

Matsumura, L.C., Ganier, H.E., Correnti, R., Junker, B., & Bickel, D.D. (2010). Investigating the effectiveness of a comprehensive literacy coaching program in schools with high teacher mobility. *The Elementary School Journal, 111*(1), 35-62.

McGlinchey, M.T. & Goodman, S. (2008). Best practices in implementing school reform. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (pp. 893-994), Bethesda: The National Association of School Psychologists.

McIntyre, E., Kyle, D., Chen, C., Munoz, M., & Beldon, S. (2010). Teacher learning and ell reading achievement in sheltered instruction classrooms: Linking professional development to student development. *Literacy Research and Instruction,* 49(4), 334-351.

Meyer, B.J.F., Brandt, D.M. & Bluth, G.J. (1980). Use of top-level structure in text: key for reading comprehension of ninth-grade students. *Reading Research Quarterly, 16*(1), 72-103.

Moore, D. W., & Readence, J. E. (1984). A quantitative and qualitative review of graphic organizer research. *Journal of Educational Research, 78*(1), 11–17.

Mortenson, B.P. & Witt, J.C. (1998). The use of weekly performance feedback to increase teacher implementation of a prereferral academic intervention. *School Psychology, Review, 27*(4), 613-627.

Murphy, K.P., Wilkinson, I.A.G., Soter, A.O., Hennessy, M.N., & and Alexander, J.F. (2007). Examining the effects of discussion on students' comprehension of text: A meta analysis. *Journal of Educational Psychology, 101*(3), 740-764.

National Commission on Teaching & America's Future. (1996). *What matters most: Teaching for America's future. Report of the National Commission on Teaching & America's future* (report ED395931). New York: Author.

National Institute of Child Health and Human Development (2003). *National symposium on learning disabilities in English language learners.* Retrieved October 2, 2011 from http://www.nichd.nih.gov/

National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Rockville, MD: National Institute of Child Health and Human Development.

Nelson, J. R., & Stage, S. A. (2007). Fostering the development of vocabulary knowledge and reading comprehension though contextually-based multiple meaning vocabulary instruction. *Education &Treatment of Children, 30*(1), 1–22.

Nesbit, J.C. & Adescope, O.O. (2006). Learning with concept and knowledge maps: A meta-analysis. *Review of Educational Research, 76*(3), 413-448.

Noell, G.H. (2008). Research examining the relationships among consultation process, treatment integrity, and outcomes. In W. Erchul & S. Sheridan (Eds.), *Handbook of Research in School Consultation* (pp.323-342), New York: Lawrence Erlbaum Associates.

Noell, G.H. & Witt, J.C. (1996). A critical re-evaluation of five fundamental assumptions underlying behavioral consultation. *School Psychology Quarterly, 11*(3), 189-203.

Noell, G.H., Duhon, G.J., Gatti, S.L., & Connell, J.E. (2002). Consultation, follow-up, and implementation of behavior management interventions in general education. *School Psychology Review, 31*(2), 217-234.

Noell, G.H., Witt, J.C., LaFleur, L.H., Mortenson, B.P., Rainer, D.D., & LaVelle, J. (2000). Increasing intervention implementation in general education following consultation: A comparison of two follow-up strategies. *Journal of Applied Behavioral Analysis, 33*, 271-284.

Noell, G.H., Witt, J.C., Gilbertson, D.N., Rainer, D.D., & Freeland, J.T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly, 12*(1), 77-88.

O'Connor, R. E. & Goodwin, V. (2011) Teaching older students to read. In R.E O'Connor & P. Vadasy (Eds.) Handbook of Reading Interventions. New York: Guilford.

Orosco, M.J. & Klingner, J. (2010). One school's implementation of RTI with english language learners: Referring into RTI. *Journal of Learning Disabilities, 43*(3), 269-288.

Parker, R.I., Hagan-Burke, S., & Vannest, K. (2007). Percentage of all non-overlapping data (PAND): An alternative to PND. *Journal of Special Education, 40*(4), 194-204.

Parker, R.I., Cryer, J., Byrns, G. (2006). Controlling baseline trend in single case research. *School Psychology Quarterly, 21*(4), 418-443.

Parrish, T B., Linquanti, R., Merickel, A., Quick, H. E., Laird, J., & Esra, P. (2002). *Effects of the implementation of Proposition 227 on the education of English learners, K-12: Year 2 Report.* Palo Alto, CA: U.S. American Institutes for Research. Retrieved September 5, 2011, from www.wested.org/online_pubs/year2finalrpt.pdf

Progress Monitoring Tools. *National Center on Response to Intervention*. US Department of Education Office of Special Education, Apr. 2011. Web. 02 Mar. 2012, from http://www.rti4success.org/progressMonitoringTools>.

Rathvon, N. (2008). *Effective school interventions: Evidence-based strategies for improving student outcomes.* New York: Guilford.

Riley-Tillman, T. C. & Burns, M. K. (2009). *Single case design for measuring response to educational intervention.* New York: Guilford.

Roach, A.T. & Elliott, S.N. (2008). Best practices in facilitating and evaluating treatment integrity. In A. Thomas & J. Grimes (Eds.), *Best Practices in School*

Psychology (pp. 195-208), Bethesda: The National Association of School Psychologists.

Robertson, L., Woolsey, M.L., Seabrooks, J., & Williams, G. (2004). An ecobehavioral assessment of the teaching behaviors of teacher candidates during their special education internship experiences. *Teacher Education and Special Education: The Journal for the Teacher Education Division of the Council for Exceptional Children, 27,* 264-273.

Rosenfield, S.A. (2008). Best practices in instructional consultation and instructional consultation teams. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (pp. 1645-1660), Bethesda: The National Association of School Psychologists.

Rosenfield, S.A. Silva, A., & Gravois, T.A. (2008). Bringing instructional consultation to scale: Research and development of IC and IC teams. In W. Erchul & S. Sheridan (Eds.), *Handbook of Research in School Consultation* (pp.203-224), New York: Lawrence Erlbaum Associates.

Salvia, J. & Ysseldyke, J.E, & Bolt, S. (2007). *Assessment in special and inclusive education.* Boston: Houghton Mifflin Company.

Sanders, W. and Rivers, J. (November 1996). *Cumulative and Residual Effects of Teachers on Future Student Academic Achievement*. Knoxville, TN: University of Tennessee Value-Added Research and Assessment Center.

Schumaker, J., Deshler, D., Alley, G., Warner, M., & Denton, P. (1982). Multipass: A learning strategy for improving reading comprehension. *Learning Disability Quarterly, 5*, 295-304.

Schunk, D.H. & Rice, J.M. (1992). Influence of reading comprehension strategy information on children's achievement outcomes. *Learning Disability Quarterly, 15*(1), 51-64.

Smith, P.L. & Friend, M.(1986). Training learning disabled adolescents in a strategy for using text structure to aid recall of instructional prose. *Learning Disability Research, 2,* 38-44.

Stahl, S.A., & Fairbanks, M. M.(1986). The effects of vocabulary instruction: A mode-based meta-analysis. Review of Educational Research, 56, 72-110.

Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407.

Stollar, S.A., Schaeffer, K.R., Skelton, S.M., Stine, K.C., Latter-Huhn, A., & Poth, R.L. (2008). Best practices in professional development: An integrated three-tier model of academic and behavior supports. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology* (pp. 875-886), Bethesda: The National Association of School Psychologists.

Swanborn M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research, 69*(3), 261–85.

Ticha, R., Espin, C., & Wayman, M. W. (2009). Reading progress monitoring for secondary-school students: Reliability, validity, and sensitivity to growth of reading aloud and Maze selection measures. *Learning Disabilities Research & Practice, 24*, 132–142.

Tolar, T.D., Barth, A.E., Francis, D.J., Fletcher, J.M., Stuebing, K.K., & Vaughn, S. (2012). Psychometric properties of maze tasks in middle school students. *Assessment for Effective Intervention, 37* (3), 131-146.

Torgesen, J., Alexander, A., Wagner, R., Rashotte, C., Voeller, K., & Conway, T. (2001). Intensive Remedial Instruction for Children with Severe Reading Disabilities: Immediate and Long-term Outcomes From Two Instructional Approaches. *Journal of Learning Disabilities, 34* (1), 33-58.

Townsend, D. & Collins, P. (2009). Academic vocabulary and middle school english learners: An intervention study. *Reading and Writing, 22*(9), 993-1019.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) Reading Assessments. (2008). *Trends in Academic Progress.* U.S. Department of Education.

U.S. Department of Education, Office of Policy Planning and Innovation, *Meeting the Highly Qualified Teachers Challenge: The Secretary's Second Annual Report on Teacher Quality*, Washington, D.C., 2003

Vaughn, S. & Fuchs, L.S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research &Practice, 18*(3), 137-146.

Wanzek, J., & Vaughn, S. (2007). Research-Based Implications From Extensive Early Reading Interventions. *School Psychology Review , 36* (4), 541-561.

Whitehurst, G.J. (2002). Remarks delivered at the White House Conference on Preparing Tomorrow's Teachers, March 2002. Available at http://www.ed.gov

Wickerstrom, K.F., Jones, K.M., LaFleur, L.H., & Witt, J.C. (1998). An analysis of treatment integrity in school-based behavioral consultation. *School Psychology Quarterly, 13*(2), 141-154.

Wilkinson, L.A. (2006). Monitoring treatment integrity: An alternative to the 'consult and hope' strategy in school-based behavioural consultation. *School Psychology International, 29*(4), 426-438.

Witt, J. C, Noeil. G. H.. LaFleur, L, H,, & Monenson, B.P. (1997). Teacher usage of interventions in general education: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30,* 693-696,

Zehler, A. M., Fleischman, H. L., Hopstock, P. J., Stephenson, T. G., Pendzick, M. L., & Sapru, S. (2003). *Descriptive study of services to LEP students and LEP students with disabilities. Volume I: Research report.* Arlington, VA: Development Associates.

Table 1
Behavioral Definitions of Instructional Observation Items

	Definition	Example	Non-Example
Explanation/ Modeling of strategy use	Teacher introduces and/or models strategy to be learned	Teacher explains how to look for information in expository text Teacher shows how to put word parts together to gain understanding of what a word means	Teacher puts a strategy on the projector and the class copies it down Students copy definitions from the board
Guided practice	Teacher and students practice the strategy together	Teacher goes through the steps of a strategy asking for student responses Student responds to teacher's questions to complete the strategy Students fill in graphic organizer worksheets while the teacher guides them through the strategy	Teacher goes through a strategy without eliciting responses from students
Practice with monitoring or feedback	Partner or independent practice using the skill presented in the lesson	Students work on a comprehension strategy in pairs Students independently use morphology skills to work on a vocabulary worksheet	Students practice a strategy in pairs or independently but the teacher is not monitoring the integrity of the practice
Explicit feedback provided	Teacher provides immediate feedback to a student's response.	Teacher gives information on why a response is correct or incorrect. Teacher follows up a student answer with more information. Teacher gives clear step by step directions of behavioral expectations	Teacher responds, "yes", "no", or "ok", with no explanation as to why

Eliciting Response with Random Selection	Teacher eliciting response or student giving a response as a result of being chosen at random	Teacher randomly selects students to respond. Student responds to a teacher's question as a result of being randomly chosen	Teacher calls on students that have their hands raised Students respond after being called for volunteering
Check for understanding	Teacher briefly evaluates if class has the level of understanding expected for the lesson	Teacher asks class to raise indicate with fingers how well they understood the content. Teacher asks follow up questions to determine depth of understanding. Students responding to these kinds of requests	Teacher asks students to raise their hands if they understood the lesson with no follow up to determine accuracy.
Choral response	Class responds to a question verbally or physically	Class gives answer together. Students indicate an answer holding up a number of fingers Teacher cues class to respond chorally	Students shouting answers without being asked to.
Written response	Students briefly respond to a question via writing	Students take notes on important details Students write a brief response to a question	Students write as part of practice of a skill (coded in #10)
Personal connection	Teacher or student shares or makes connection to background knowledge	Connection made to cultural background or personal experiences Connection made to previously learned material in class or other subject.	Student shares a story unrelated to the content.

Explaining reasoning/thinking	Teacher or student explains why they gave an answer or how they came up with the answer	Student explains how they determined the meaning of a word or how a comprehension strategy was chosen. Student explains why one answer is correct while another is not.	Student responds to a question without explanation of how they determined that answer.

Table 2
Analysis of Results Summary

Visual Analysis	
Type	Definition
Level	Data points in treatment phase are higher or lower than those at baseline.
Trend	Trend (slope) of data points in treatment is different from those at baseline. Treatment data point trend should not be predicted from baseline trend.
Immediacy	Change in level and trend occurs immediately or very soon after treatment is implemented.
Variability	Variability of the behavior at treatment is lower than at baseline. This indicates that behavior is more stable and predictable during treatment than without treatment.

Effect Size Analysis		
	Calculation	Interpretation
Percent of All Non-overlapping Data (PAND)	100- [(Total Intervention Data Points Overlapping Baseline / Total Data Points)*100]	Percent of behavior measured above/below baseline behavior
Phi	Step 1 Cell A: % of baseline (baseline/total) / Cell B: % overlapping data/2 Cell C: % overlapping data/2 / Cell D: % of tx (tx/total) Step 2 Φ = [a/(a+c)-[b/(b=d)]	Small=.10 Medium= .30 Large = .50

Table 3
Interrater Reliability

	Average	Range
Total Scaffolding	0.91	0.84 – 1.00
Modeling	0.71	0.33 – 1.00
Guided Practice	0.91	0.81 – 1.00
Practice	0.95	0.85 – 1.00
Feedback	0.79	0.50 – 1.00
Written Response	0.72	0.06 – 1.00
Random Response	0.75	0.46 – 1.00
Choral Response	0.72	0.10 – 1.00
Explaining Thinking	0.45	0.06 – 1.00
Check for Understanding	0.71	0.75 – 1.00
Personalized Connections	0.55	0.12 – 0.71
Total Engagement	0.94	0.87 – 0.99
Active	0.80	0.60 – 0.98
Passive	0.90	0.71 – 0.97
Off Task	0.76	0.37 – 0.90

Table 4
Effect Size Statistics

	PAND	Phi	Interpretation (Cohen, 1988)
Total Scaffolding	54.88%	0.34	Medium
Modeling	37.80%	0.19	Small
Guided Practice	57.31%	0.36	Medium
Practice	53.65%	0.33	Medium
Feedback	56.01%	0.35	Medium
Written Response	61.00%	0.40	Medium
Random Response	63.20%	0.42	Medium
Choral Response	37.80%	0.20	Small
Total Engagement	72.00%	0.53	Large
Active	62.19%	0.42	Medium
Passive	52.43%	0.32	Medium
Off Task	58.53%	0.38	Medium
Comprehension	65.83%	0.49	Medium

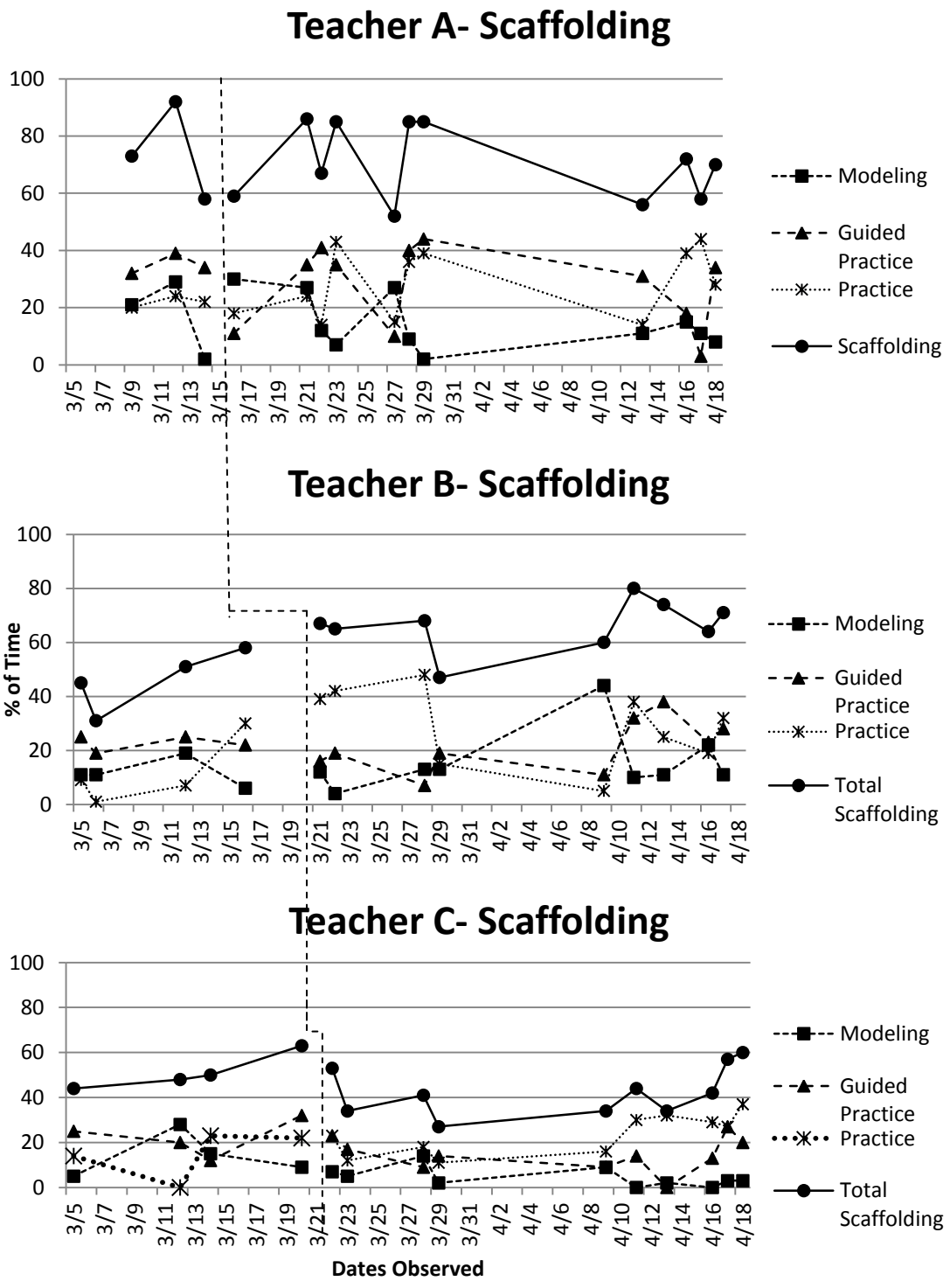

Figure 1a. Percent of intervals with scaffolding during time observed in concurrent phase 1.

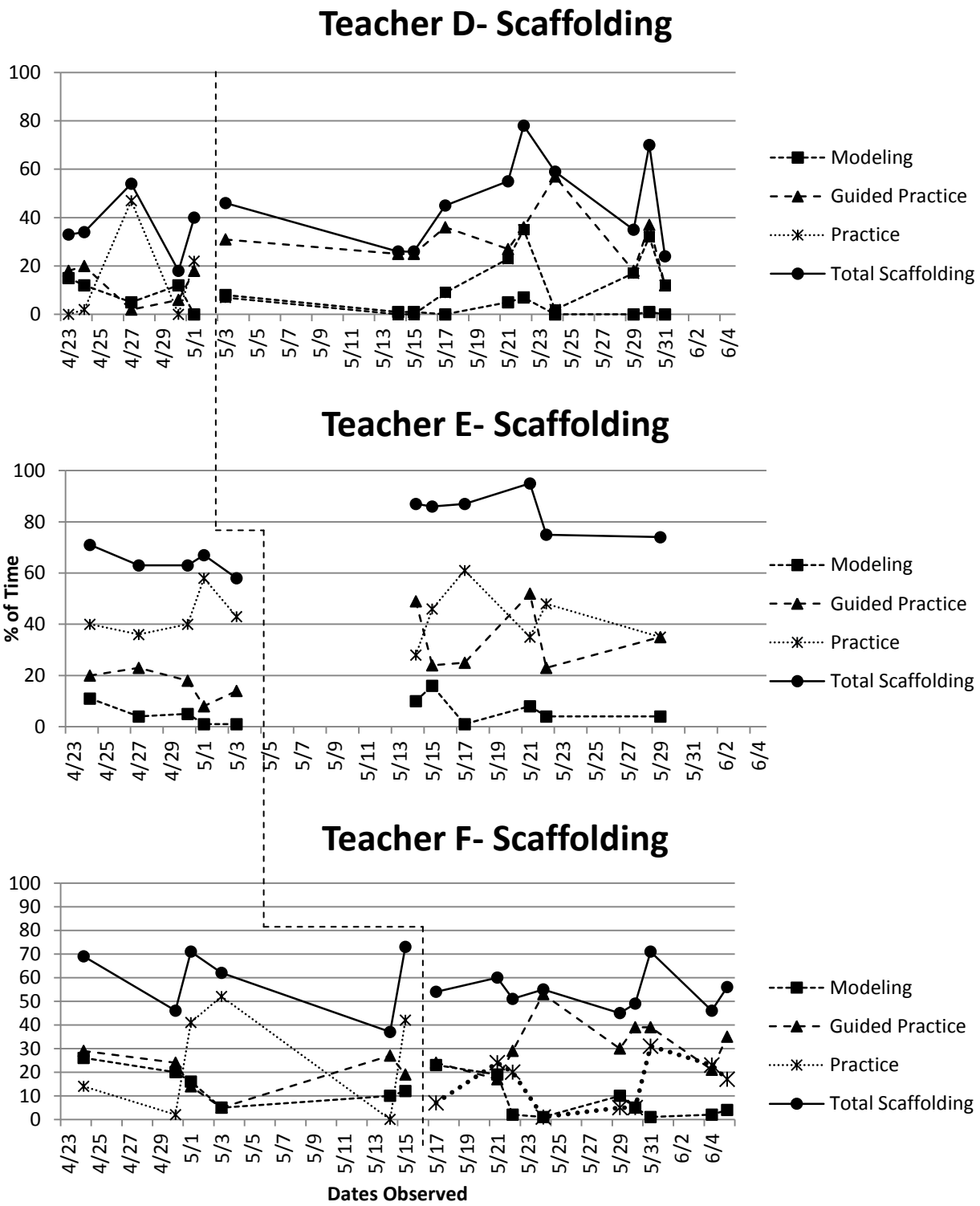

Figure 1b. Percent of intervals with scaffolding during time observed in concurrent phase 2.

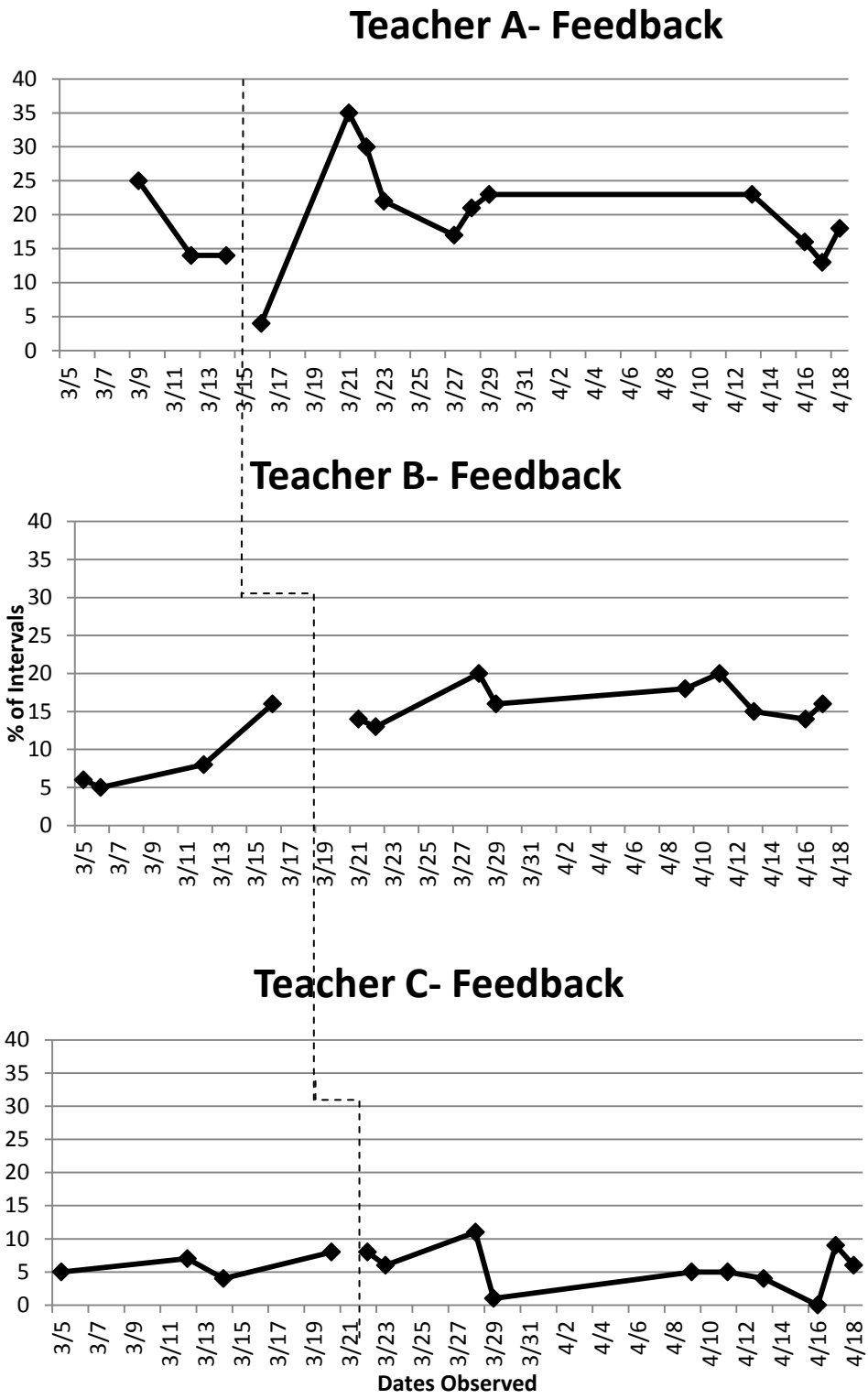

Figure 2a. Percent of intervals with feedback during time observed in concurrent phase 1.

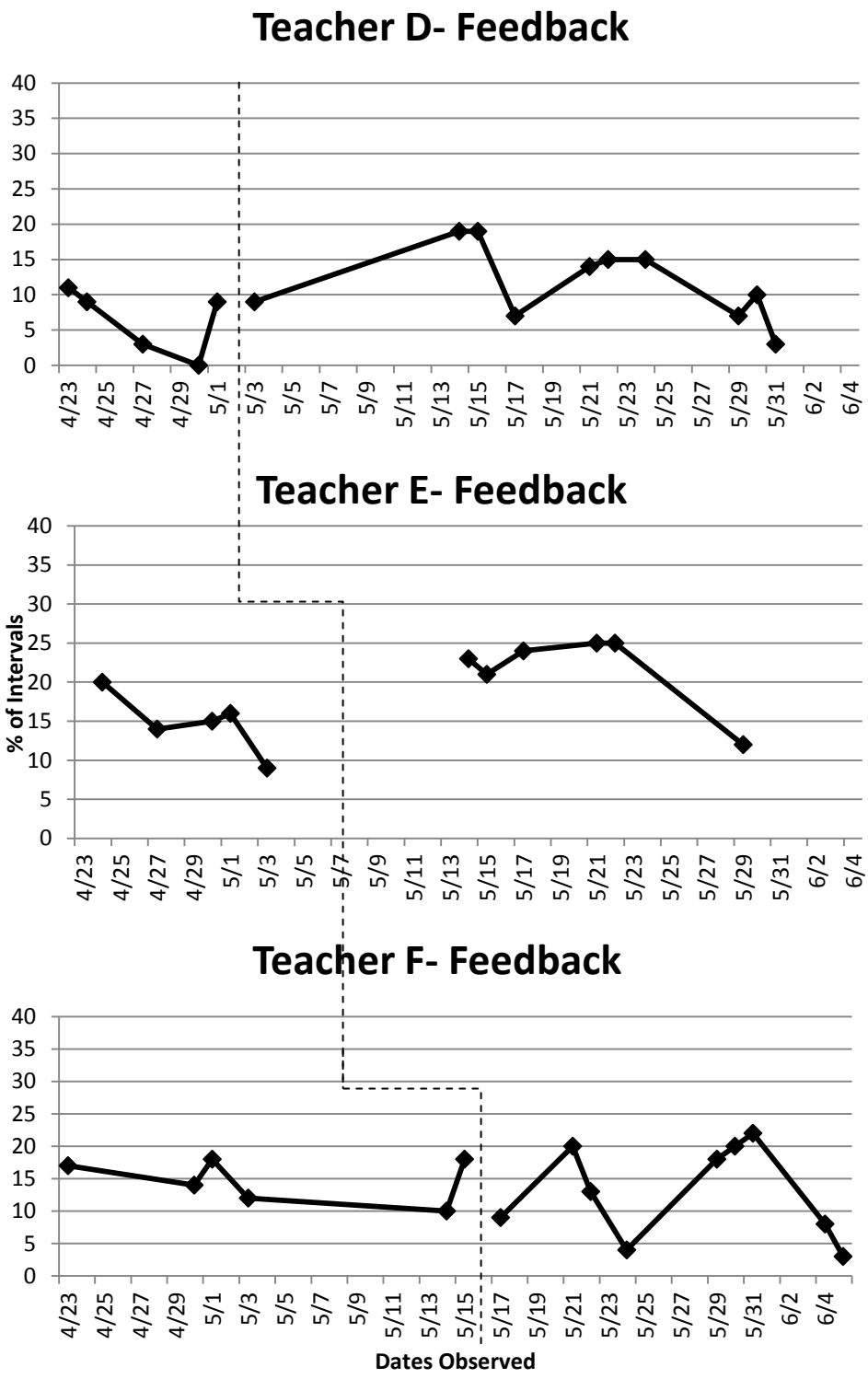

Figure 2b. Percent of intervals with feedback during time observed in concurrent phase 2.

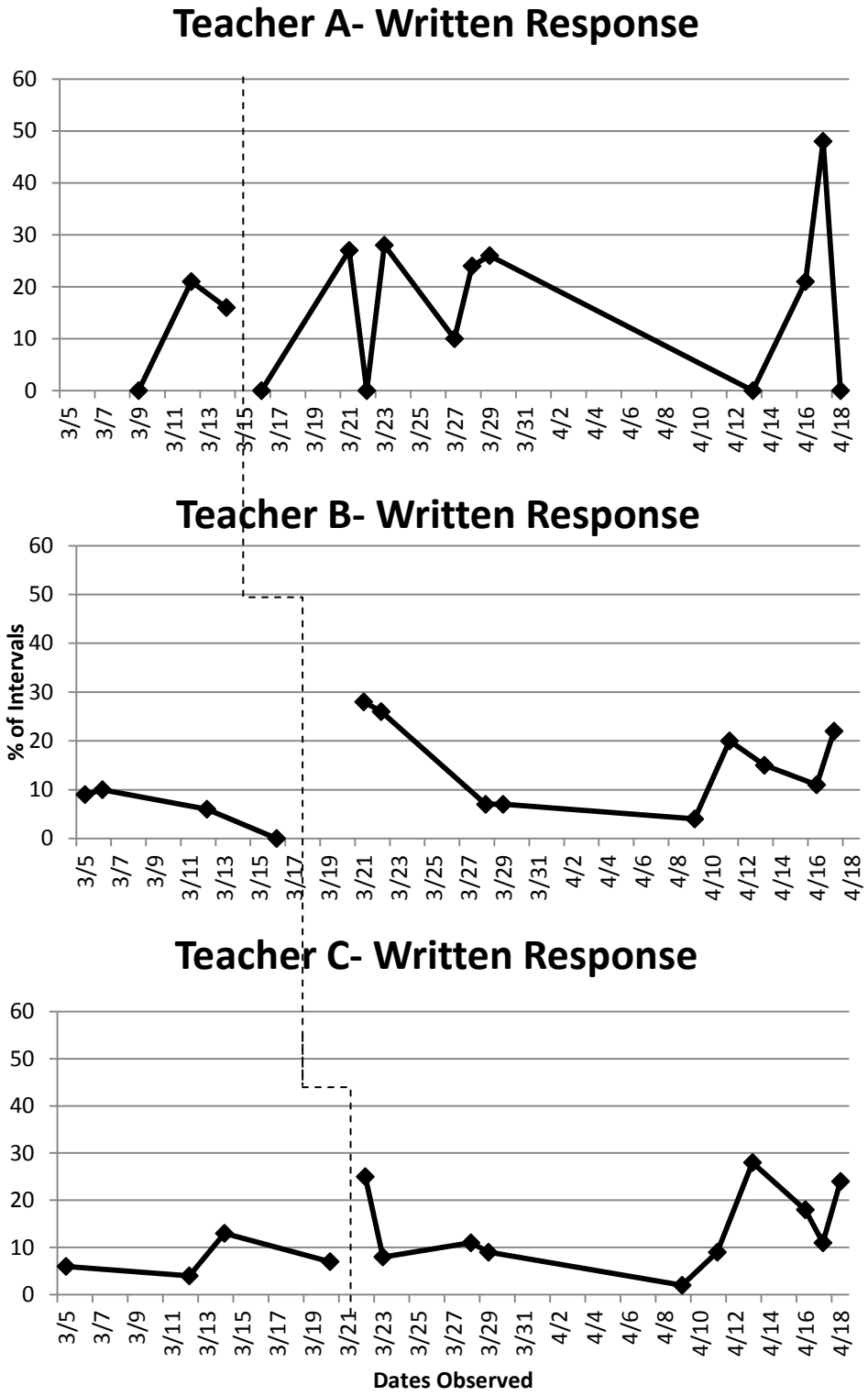

Figure 3a. Percent of intervals with written response during time observed in concurrent phase 1.

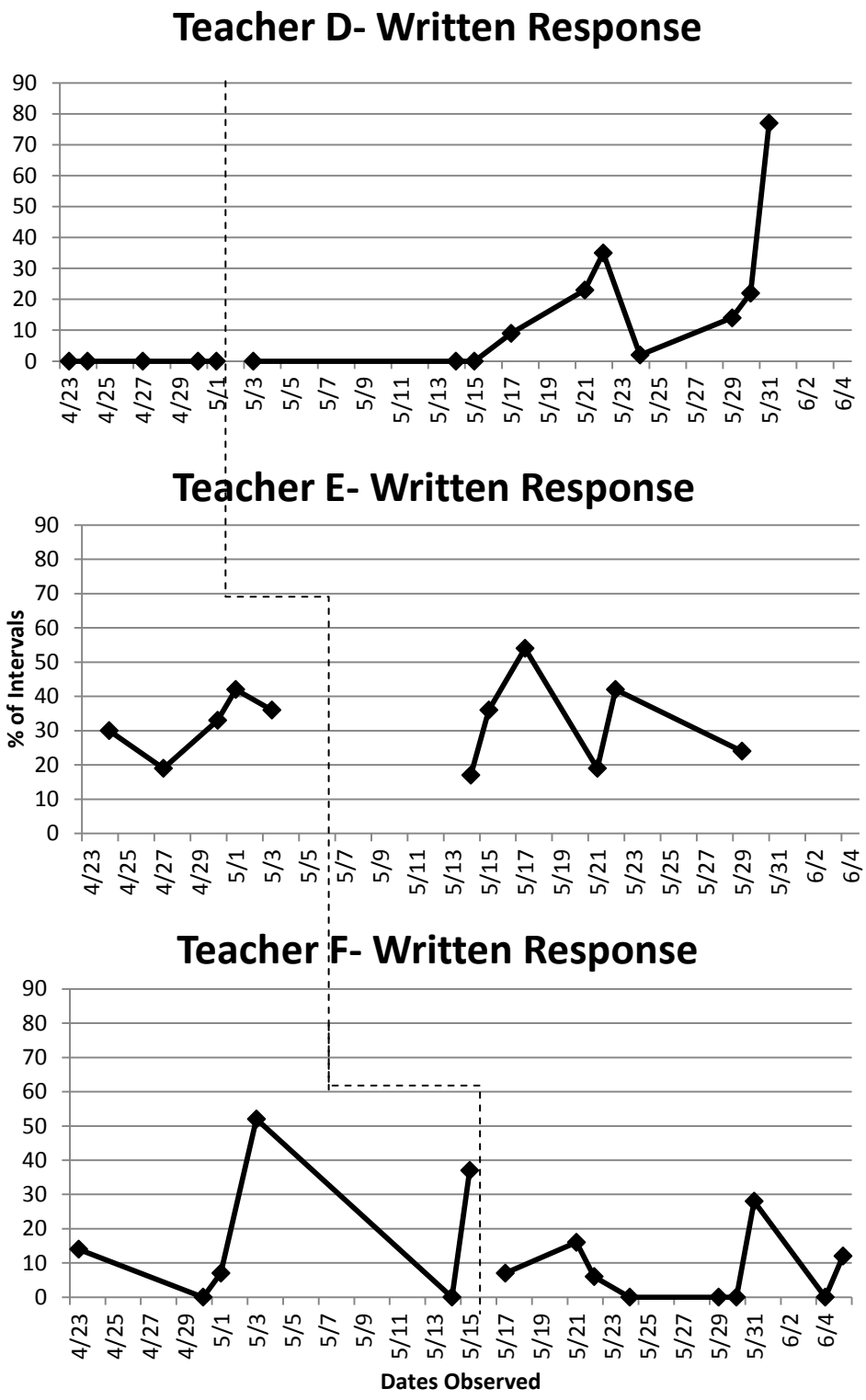

Figure 3b. Percent of intervals with written response during time observed in concurrent phase 2.

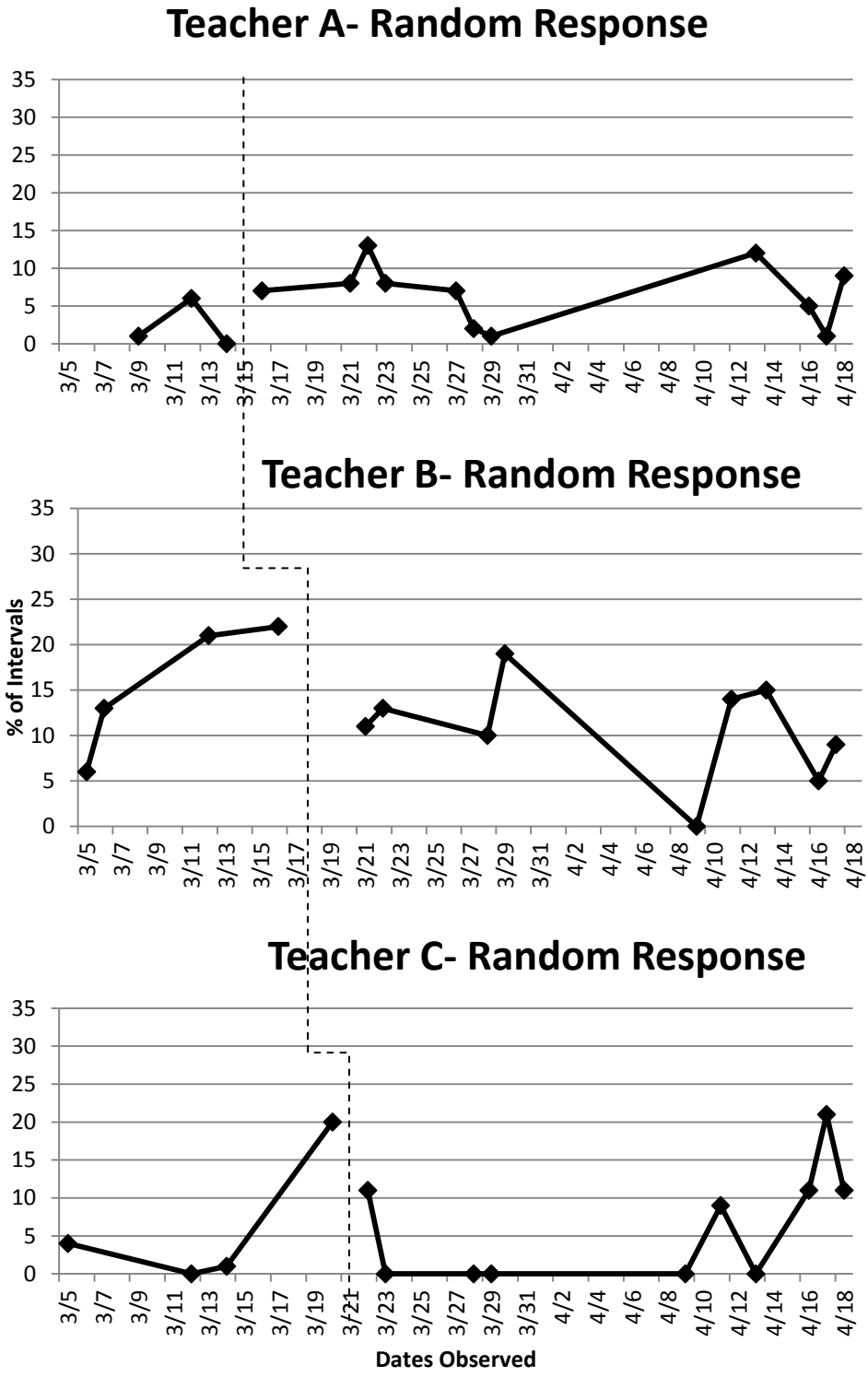

Figure 4a. Percent of intervals with random response during time observed in concurrent phase 1.

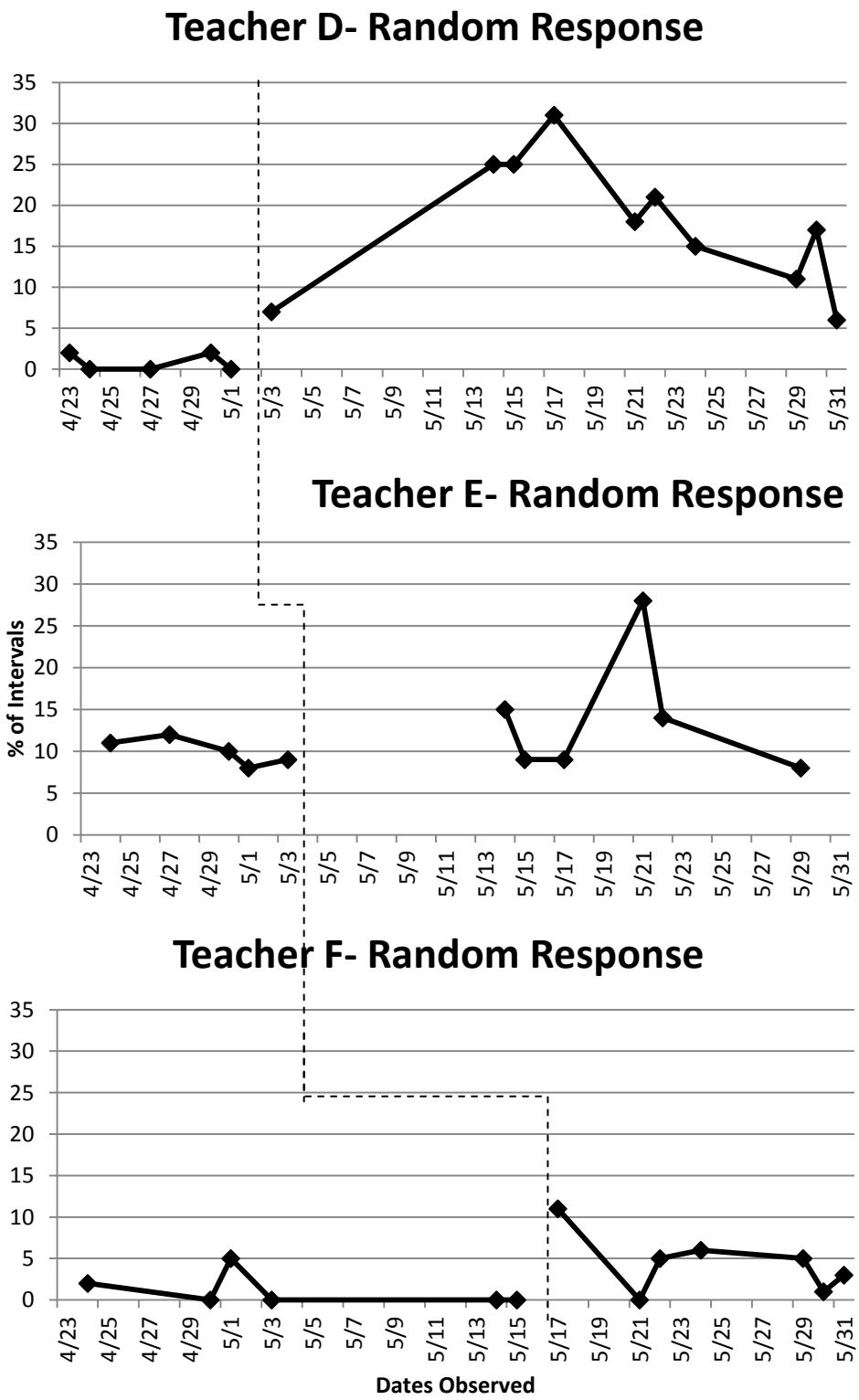

Figure 4b. Percent of intervals with random response during time observed in concurrent phase 2.

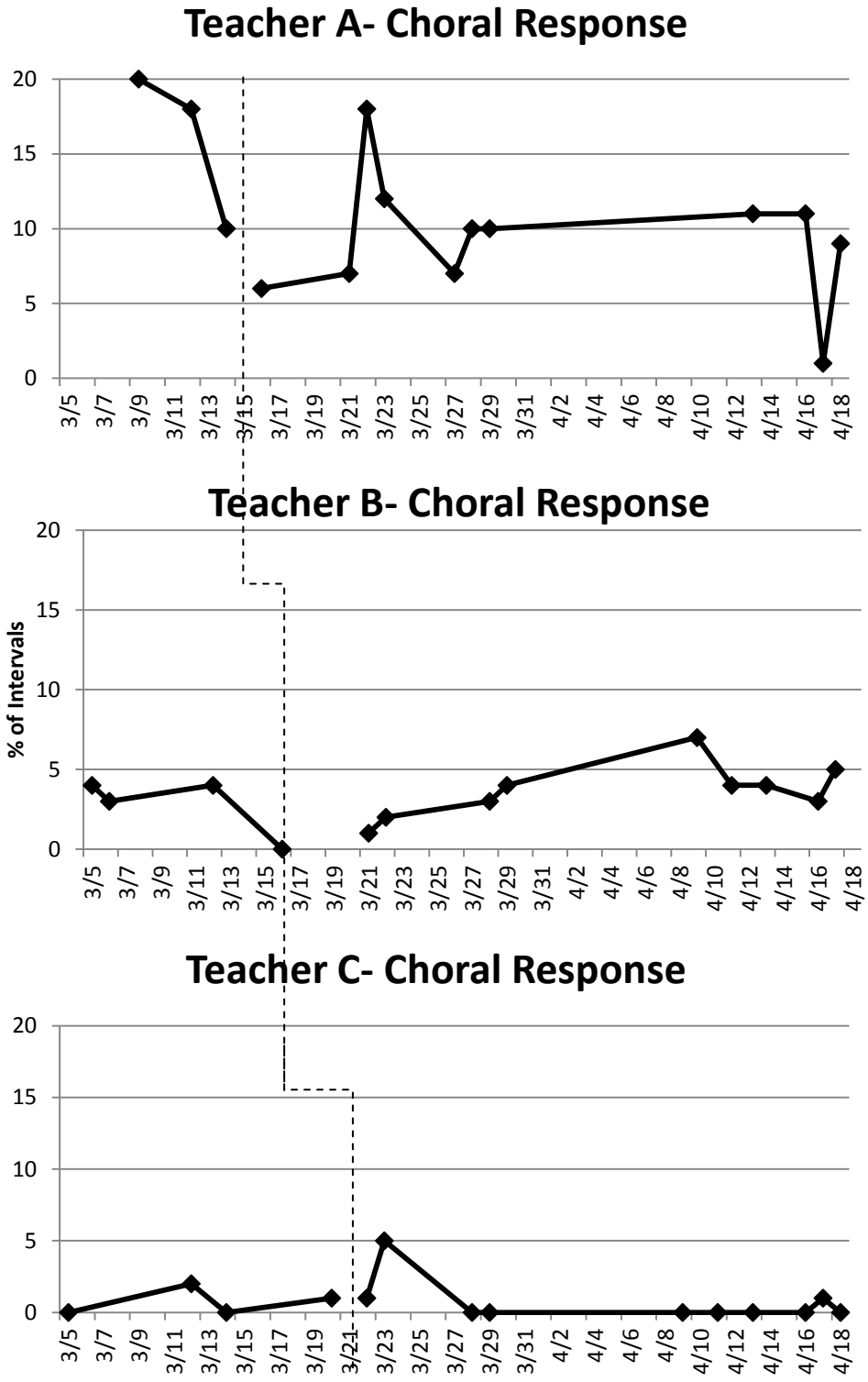

Figure 5a. Percent of intervals with choral response during time observed in concurrent phase 1.

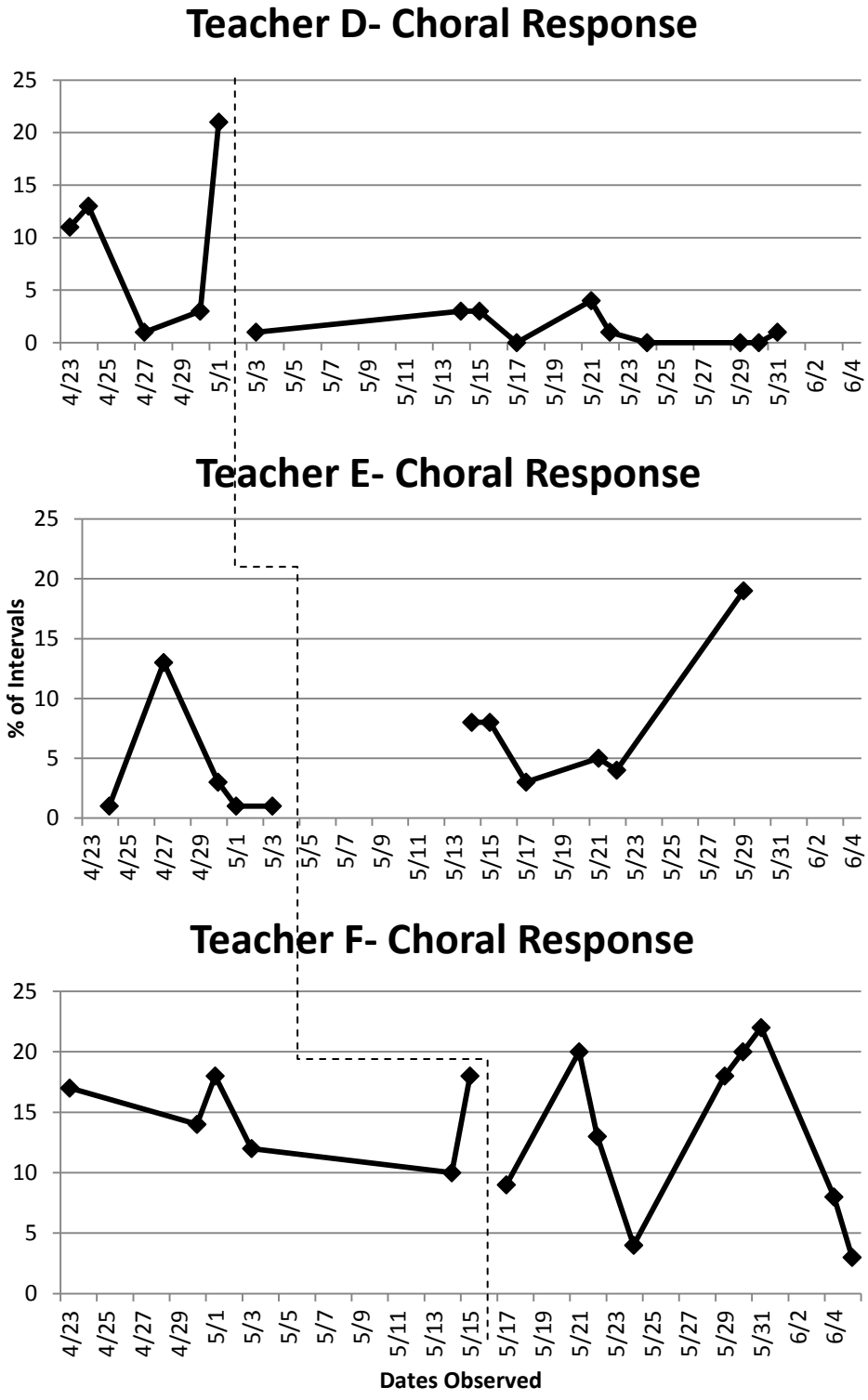

Figure 5b. Percent of intervals with choral response during time observed in concurrent phase 2.

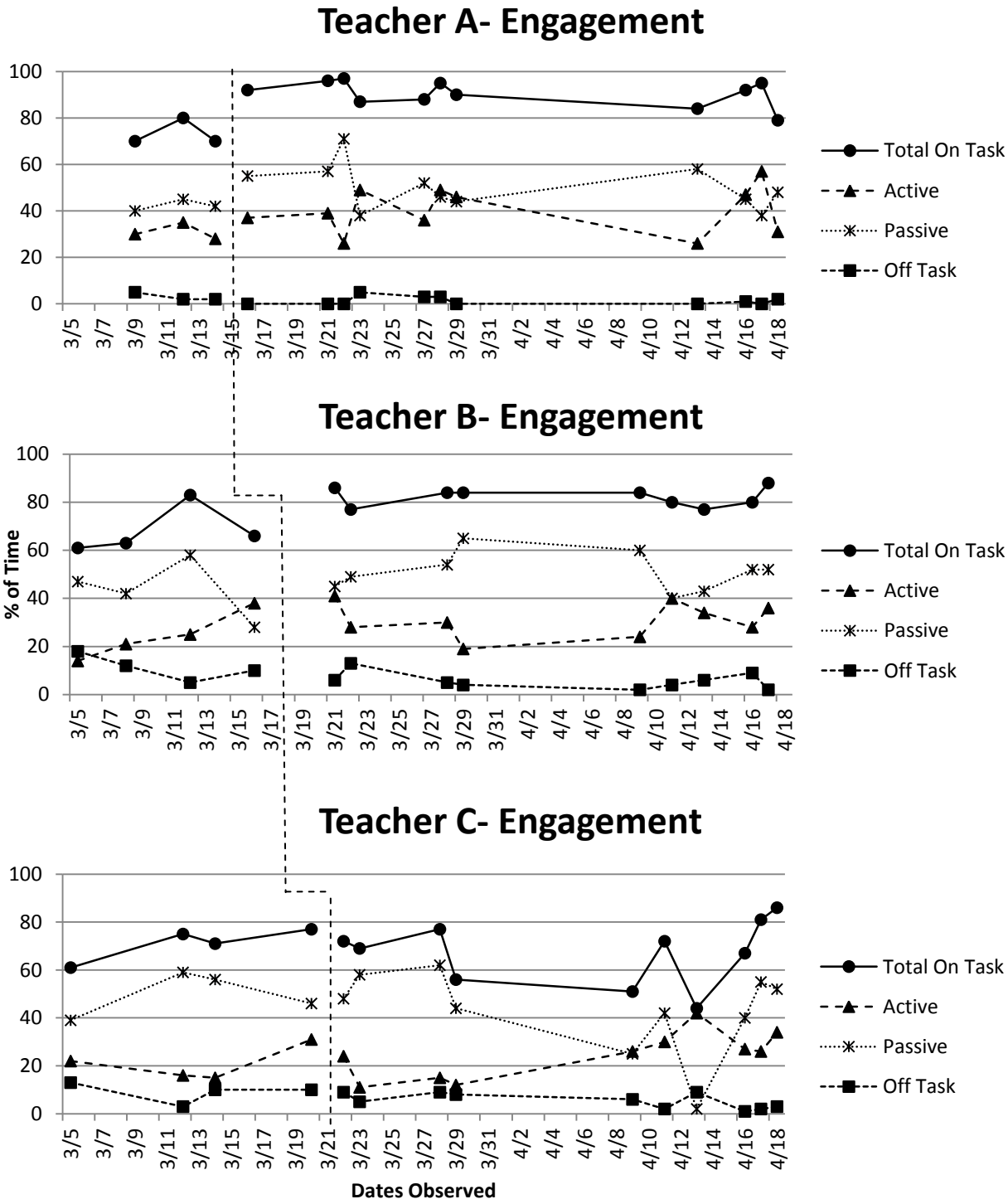

Figure 6a. Percent of intervals with student engagement during time observed in concurrent phase 1.

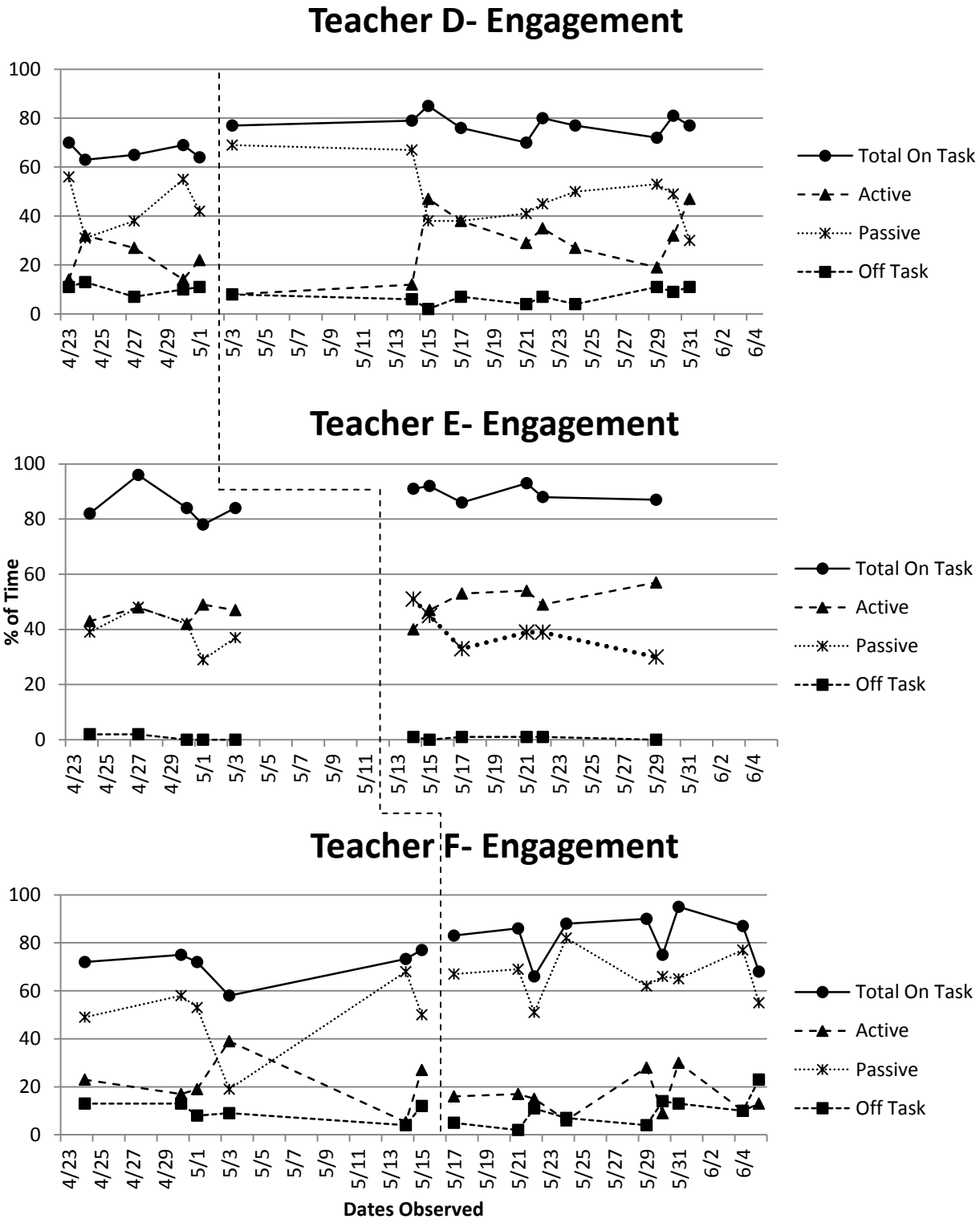

Figure 6b. Percent of intervals with student engagement during time observed in concurrent phase 2.

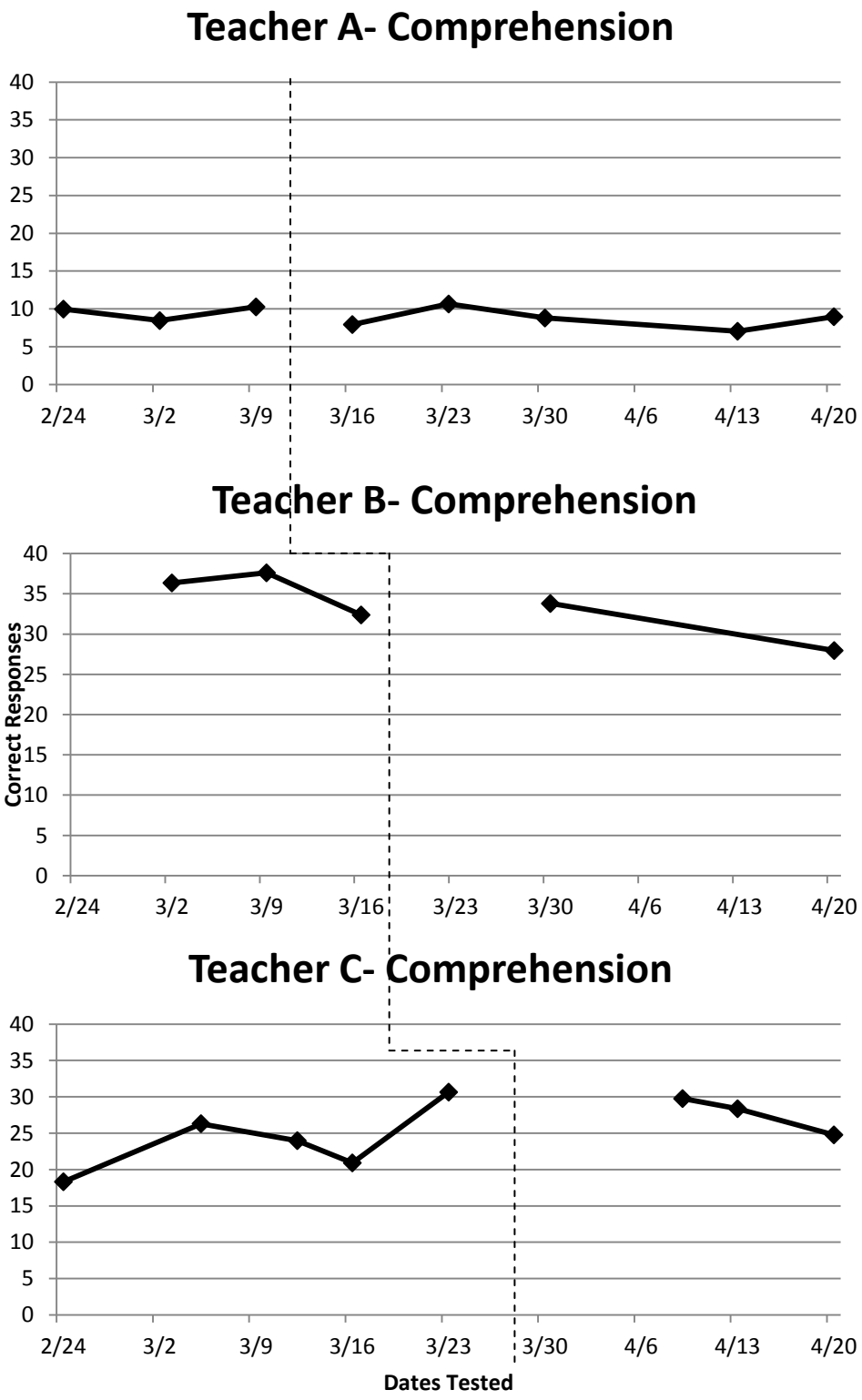

Figure 7a. Average correct responses on weekly MAZE in concurrent phase 1.

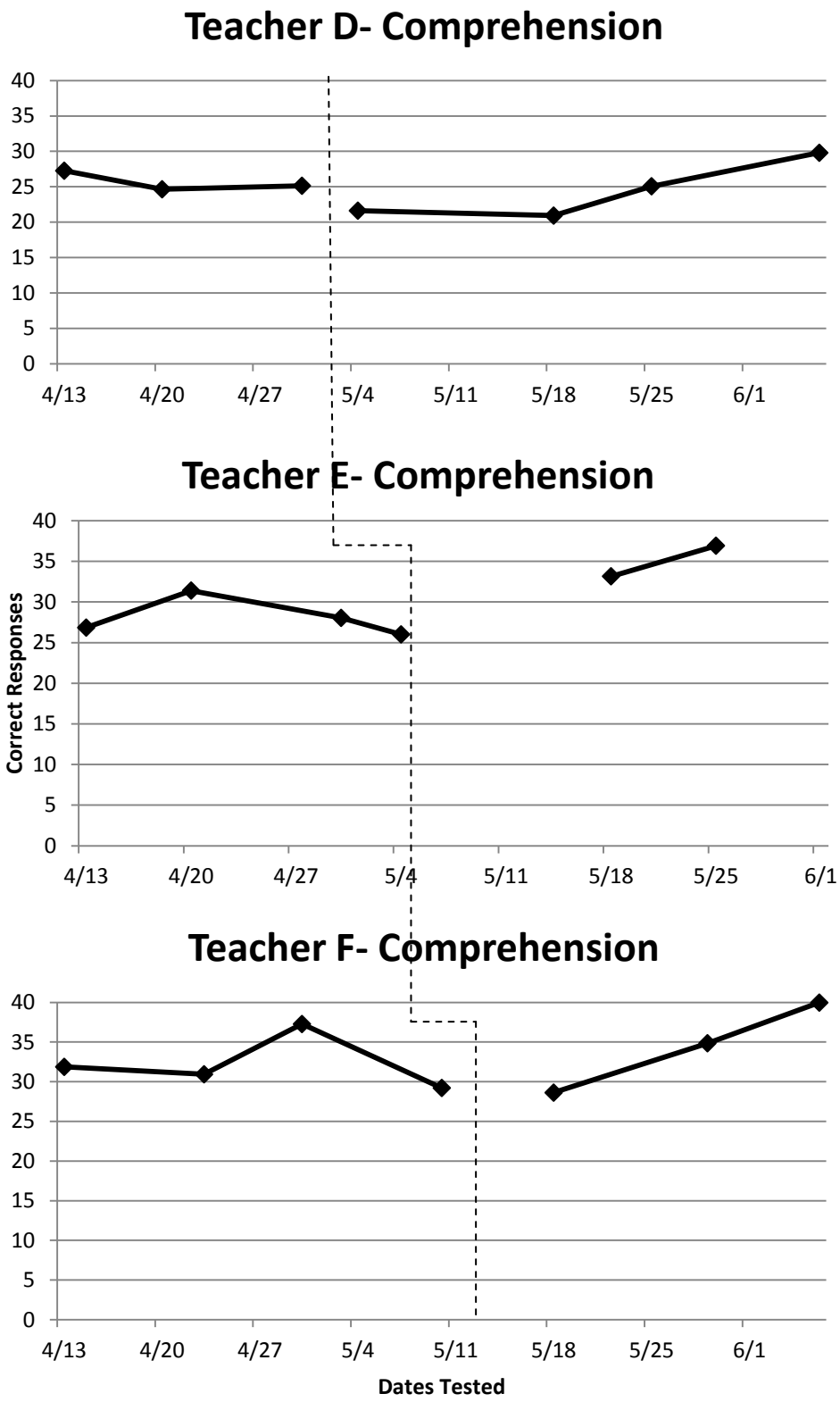

Figure 7b. Average correct responses on weekly MAZE in concurrent phase 2.