

**THE USE OF PERFORMANCE FEEDBACK TO INCREASE TEACHERS'  
IMPLEMENTATION OF INTERVENTIONS AND CLASSROOM  
MANAGEMENT STRATEGIES**

A Dissertation

by

MARCUS CHARLES FULLER

Submitted to the Office of Graduate and Professional Studies of  
Texas A&M University  
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Kimberly Vannest
Co-Chair of Committee,	Jay Ganz
Committee Members,	Christopher Thompson
	Beverly Irby
Head of Department,	Shanna Hagan-Burke

August 2020

Major Subject: Educational Psychology

Copyright 2020 Marcus C. Fuller

## **ABSTRACT**

The purpose of coaching is to support the implementation of newly acquired skills in teachers and other school staff. Using a combination of cognitive and problem-solving strategies, coaching is hard to identify as a single practice. A key component of coaching is the observation and feedback cycle. Performance feedback (PF) is an evidence-based practice as an implementation strategy for improving educators' fidelity of implementation. Yet still, there is some disagreement of what PF is due to vague terminology and PF being a part of a bigger treatment package, coaching. Supported by the literature, PF can be described as a broad definition that encompasses a personal interaction with the oral, written, or gestural communication regarding the progress towards the desired outcome. PF involves observations of the teacher to collect data on implementation and sharing that data with the teacher to improve their future performance. There are several variables that can change the effectiveness of PF on teachers' implementation of a target intervention, new curriculum, or strategy. However, research supports that any type of PF is at least somewhat effective in changing teachers' behavior. Some variables of PF include immediacy of delivering the feedback, and dosage. Several studies have looked at these variables, as well as using PF to teach a particular strategy to school staff. To date, no studies have conducted a moderator analysis of these PF variables or assessed the quality of literature for PF in isolation from other coaching components. The current dissertation analyzes PF as a way to improve teacher and school staff implementation by answering the following research questions: 1) Is PF supported by the literature as an EBP according to the WWC single-case design standards? 2) What

are the effects of PF on fidelity of implementation of classroom strategies and programs?

3) What moderator variables of PF produces the best effect for implementation fidelity?

4) Does PF increase the teachers' implementation of Opportunities to Respond? 5) What effects does high rates of Opportunities to Respond have on students' on-task behavior?

## ACKNOWLEDGEMENTS

I am honored and blessed to have accomplished this body of understanding. None of this could be possible without acknowledging several persons that made this possible, as well as giving all the honor and glory to God for granting me such opportunity.

First, I would like to thank my committee members, Dr. Vannest, Dr. Ganz, Dr. Thompson, and Dr. Irby. Your support and advice helped me accomplish this feat. I want to say thank you again to my chairs for advising me for these four years. Dr. Ganz the opportunities to work on research and participate in parent coaching helped me realize the importance to transferring knowledge and the passion to make it my own line of research. Dr. Vannest, through the conference and workshop we facilitated I realized how I can use my skills and research to support teachers. I am honored to be in the company of impactful professors.

Thanks also goes to my friends and organizations that have kept me grounded and motivated to complete my schooling. Thanks to my fraternity, Kappa Alpha Psi Fraternity Incorporated. The Lambda Alpha chapter from my alma mater was an unwavering source of support with many wise men offering great advice. The Nu Alpha chapter here on campus made Texas a home away from home for me. The experiences and memories made will be longstanding. Another organization that I would be remiss not to acknowledge the Black Graduated Student Association. These young black persons kept me centered in my culture and equipped me to conquer the attacks and slander that a person of color is still expose to in today's society. Additionally, I need to

acknowledge my friend groups that made this doctoral experience both educational and entertaining. Thank you to The A-Team, Quarantine Live, Team Bar, The Enablers, Chocolate City Crew and The Hive.

Finally, I want to thank my family for making this possible. Since a young age you have instilled in me excellence and perseverance. Thanks to my mother for always being a sound board for when I needed to vent. Thanks to my father for helping me find a silver lining in any situation and giving me something to laugh about. Thank you to my grandma for keeping my steadfast in the Lord and reminding me to be impactful. To my sister Tierra, thank you for reading and summarizing articles when I was too tired. To my brother Tyrese, thanks for inspiring me to continue achieving in my field. Finally, thank you nanny and pappy, while you are no longer here, your spirit has driven me and inspired me to be the man before you today.

## **CONTRIBUTORS AND FUNDING SOURCES**

### **Contributors**

This work was supervised by a dissertation committee consisting of my co advisors, Dr. Vannest and Dr. Ganz from the department of educational psychology. My committee member, Dr. Christopher Thompson from the department of educational psychology. As well as my outside committee member, Dr. Beverly Irby from the department of Educational Administration and Resource Development.

Inter-rater reliability was conducted by a research team encompassing four graduate level students and my Co-advisor Dr. Kimberly Vannest. Graduate students from Texas A&M University included April Haas, Mary Rose Sallese, Sandy Smith, and Tyler Evans. All other work conducted for this dissertation was completed by the student independently.

### **Funding**

No outside funding was received for the research and writing of this document.

## NOMENCLATURE

EBP	Evidence-Based Practices
OTR	Opportunities to Respond
PD	Professional Development
PF	Performance Feedback
RTPG	Research-To-Practice Gap

## TABLE OF CONTENTS

	Page
ABSTRACT .....	ii
ACKNOWLEDGEMENTS .....	iv
CONTRIBUTORS AND FUNDING SOURCES.....	vi
NOMENCLATURE.....	vii
TABLE OF CONTENTS .....	viii
LIST OF FIGURES.....	xi
INTRODUCTION.....	1
How Coaching Became Important .....	2
The Importance of Performance Feedback Within Coaching .....	4
Definition of Performance Feedback and Research Questions .....	5
THE IMPORTANCE OF DELIVERING PERFORMANCE FEEDBACK TO TEACHERS TO IMPROVE PRACTICE: AN EVALUATION OF THE QUALITY OF THE LITERATURE .....	7
Introduction .....	7
Review of problem .....	7
Working definition of performance feedback .....	10
Study purpose.....	10
Methods.....	11
Literature search procedures. ....	11
Inclusion criterion.....	13
Selection and application of quality indicators. ....	15
Interrater reliability procedures .....	16
Results .....	17
Study characteristics.....	17
Participant characteristics.....	18
Intervention implemented.....	19
Performance feedback characteristic.....	20
WWC quality indicators evaluation .....	21
Performance feedback as an evidence-based practice.....	22
Discussion .....	23



Limitations.....	26
Implications .....	27
Future research. ....	27

PERFORMANCE FEEDBACK TO INCREASE TEACHER PRACTICES: A

SINGLE-CASE META-ANALYSIS.....	31
--------------------------------	----

Study Purpose.....	34
Method.....	34
Literature search. ....	34
Inclusion criteria.....	35
Title and abstract review. ....	37
Full-text review and quality indicators.....	37
Variable codes .....	38
Effect size calculation .....	39
Statistical analysis .....	40
Inter-rater reliability .....	42
Results .....	44
Omnibus effect .....	45
Publication bias. ....	46
Moderator variable comparison.....	49
Discussion .....	51
Limitations.....	54
Future research .....	55

USING PERFORMANCE FEEDBACK TO INCREASE TEACHERS'

IMPLEMENTATION OF OPPORTUNITIES TO RESPOND.....	56
---	----

Method .....	58
Participant recruitment and inclusion criteria .....	58
Classroom characteristics .....	59
Participant characteristics.....	61
Procedures .....	61
Study design and variables .....	66
Effect size calculation .....	68
Inter-observer agreement.....	69
Results .....	70
Teacher behavior. ....	70
Classroom on-task behavior .....	74
Qualitative analysis of social validity .....	74
Discussion .....	75
Limitations.....	77

Implications for practice.....	77
Implication for research.....	78
CONCLUSION .....	79
Understanding of Results .....	80
Literature review .....	80
Meta-analysis.....	81
Single-case experiment.....	82
Limitations.....	82
Future Research.....	83
Implications .....	83
WORKS CITED.....	85
APPENDIX A SYNTAX FOR R STUDIO .....	95
APPENDIX B ARTICLE CHARACTERISTICS .....	97

## LIST OF FIGURES

	Page
Figure 2.1 PRISMA Flow Diagram .....	29
Figure 2.2 WWC quality indicator .....	30
Figure 3.1 Study Effect Size.....	44
Figure 3.2 Forest Plot .....	47
Figure 3.3 Funnel Plot.....	48
Figure 4.1 Procedural Checklist .....	65
Figure 4.2 Data Collection Sheet .....	67
Figure 4.3 IOA Percentages Across conditions and Phases .....	70
Figure 4.4 Effects of PF on Teachers' Use of OTR.....	73

## INTRODUCTION

This dissertation is focused on performance feedback (PF), but one cannot mention PF without first ensuring a brief understanding of how PF fits under the bigger umbrella of coaching and what coaching is in general. Yet, the answer to what coaching is in general warrants a dissertation all of its own. While, the research and implementation of coaching have been steadfast since first publications in the early 1980s (Joyce & Showers, 1981, 1982; and Showers, 1985), a clear and concise definition of coaching is not available. Many researchers have stated that the concept of coaching is inherently ambiguous (Kraft, Blazar, & Hogan, 2018). Nevertheless, research does present some understanding of what coaching should include (Cornett & Knight, 2009; Denton & Hasbrouck, 2009; and Kraft, Blazar, & Hogan, 2018). It is important to note that while an understanding of coaching and how PF fits into coaching is presented in this dissertation, there are multiple approaches when it comes to coaching (Cornett & Knight, 2009). Thus, an in-depth understanding of coaching is beyond the scope of this dissertation.

Coaching can mean different things for different people depending on need. There are various types of coaching; instructional coaching, cognitive coaching, peer coaching, content coaching (Cornett & Knight, 2009). Generally coaching, for educational purposes, can be described as a professional development (PD) activity that is ongoing, intimate, and teacher-oriented to gain a new skill, strategies, or understanding (Cornett & Knight, 2009; Denton & Hasbrouck, 2009; Kraft, Blazar, & Hogan, 2018; and Joyce & Showers, 1981). Coaching involves some form of an

observation and feedback cycle and is likely to have additional components embedded such as modeling (Coddling, Feinberg, Dunn, & Pace, 2005), rehearsal (Anderson, Feldman, & Minstrell, 2014), problem-solving (Burns, Peter, & Noell, 2008), goal setting (DunCan, Dufrene, Sterling, & Tingstrom, 2013), as well as varying levels of intensity (Gage, Grasley-Boy, and MacSuga-Gage, 2018). A recent meta-analysis conducted by Kraft, Blazar, and Hogan (2018) describes several key features of coaching. Coaching should be individualized with one-on-one sessions. Coaching is an intense process that involves ongoing sessions sustained for an expanded time. Another key component of coaching is that coaching is context specific. This means that the teacher is being coached on their performance of the skill in the classroom environment. Finally, coaching is focused or targeted to increase a specific skill or practice.

### **How Coaching Became Important**

The pioneers of coaching research, Bruce Joyce and Beverly Showers, began in the 1980s using coaching as a vehicle to transfer knowledge and skills that were learned during PD workshops into the classroom and the teacher's everyday practice (Joyce & Showers, 1981). The need for coaching was sparked from the realization that teachers were not transferring the knowledge they gained during PD into their everyday teachings. Observations showed that teachers were able to perform the new skill after the PD workshop and on command when asked. Yet, their daily teaching practices did not reflect the new skills or teachers failed to implement these skills correctly. Coaching was beginning to look like a viable way to maintain and implement the knowledge gained from PD into the classroom. The growing research on coaching supported its use to

increase teachers' performance by regularly conducting observation and feedback sessions. During these sessions, the coach would not only provide information on the implementation of the skill but give additional insight on why the skill was important, when to use the skills, with what students to use the skill with, and how to generalize this skill. This intensive and ongoing partnership between the teacher and the coach is linked to an increased understanding of the skill and increased implementation (Denton & Hasbrouck, 2009). During the 1990s and 2000's thousands of coaching jobs began to appear (Denton & Hasbrouck, 2009) and the research on coaching grew. Findings from early studies showed that coaching increased teachers' implementation rates and in turn, teachers with higher rates of implementation had students with better academic scores (Joyce & Showers, 1981; and Showers, 1982).

This push for coaches in educational settings was in part to federal initiatives to support, create, and maintain high-quality teachers. Research provided several Evidence-Based Practices (EBP) for numerous populations, yet a gap was still present between what was identified in research and what was being used in the classroom. This was known as the research-to-practice-gap (RTPG). The RTPG kept vital information that could increase student success out of the hands of the teachers that were in charge of ensuring that success. Legislation such as the Reading Excellence Act (1998), Reading First under No Child Left Behind (2001) and Every Student Succeed Act (2015) required schools to implement EBP and fund PD that provides information on these EBP (Hershfeldt, Pell, Sechrest, Pas, & Bradshaw, 2012). With these initiatives' schools had

money to invest in more PD for their teachers to ensure that the teachers were well equipped to tackle some of the daily issues they were faced within their classrooms.

Schools began to integrate new curriculums, classroom strategies, and school-wide intervention plans using PD workshops and teacher conferences to supply them with the information needed to implement. However, there were still several problems with the traditional (sometimes called one-shot workshops) PD model. Traditional PD models are passive and lack personalization to the teachers' individual needs; thus, they are ineffective for changing the teachers' behavior (Johnson, Pas, & Bradshaw, 2016). Gage, Grasley-Boy, and MacSuga-Gage (2018) suggest that PD models without sustained support and follow-up are ineffective. Furthermore, one-shot workshops, seldom cause measurable, sustained change when looking at student outcomes (Denton & Hasbrouck, 2009). One way to combat these problems is to incorporate coaching into current PD models.

### **The Importance of Performance Feedback Within Coaching**

Despite coaching becoming more common when deciding on PD, it is still being understood by the research community and practitioners alike. Many different coaching models exist as well as many different needs for coaching. Amongst all the different variations of coaching one commonality they share is the incorporation of an observation and feedback cycle (Kraft, Blazar, Hogan, 2018). This part of the coaching process involves the coach observing the teacher implement the targeted skill in the desired environment (i.e. classroom). After the observation, the coach delivers feedback on the teachers' performance. This observation and feedback cycle is repeated numerous times

throughout the coaching experience and is designed to focus the purpose of the coaching and heighten the teachers' understanding of the targeted skill (Blazar & Kraft, 2015; Denton & Hasbrouck, 2009; and Joyce & Showers, 1981).

Performance feedback (PF) is an EBP for improving educator's fidelity of implementation (Fallon, Collier-Meek, Maggin, Sanetti, & Johnson, 2015). Yet as in coaching itself, there is some disagreement of what PF is due to vague terminology and PF being a part of a bigger methodology of coaching. PF involves observing the teacher, collect data on implementation, and sharing the data with the teacher to improve their future performance (Brock & Cater, 2017; Brock, et al., 2017; Cavanaugh, 2013; Kraft, Blazar, & Hogan, 2018; and Solomon, Klein, & Politylo, 2012). Research has begun to identify several contextual variables that may impact the effectiveness of PF on teachers' implementation of an intervention. However, virtually any type of PF is at least somewhat effective in changing teachers' behavior (Solomon, Klein, & Politylo, 2012). Some variables of PF included online or in-person feedback, the immediacy of delivering the feedback, how often observation and feedback should occur, and the type of feedback (i.e. narrative feedback, goal setting, or rating). A few studies have looked at these components of PF in isolation, as well as using PF to better a particular strategy to teachers and school staff (Brock & Carter, 2017; Brock, et al., 2017; Cavanaugh, 2013, Kraft, Blazar, & Hogan, 2017; and Solomon, Klein, & Politylo, 2012).

### **Definition of Performance Feedback and Research Questions**

Currently, no study has assessed the quality of the PF research literature nor conducted a component analysis of PF. The current dissertation analyzes PF as a way to



improve educator implementation. For understanding and consistency, throughout this dissertation, PF is described broadly and encompasses a personal interaction with the oral, written, or gestural communication regarding the progress towards the desired outcome. This understanding of PF was investigated to answering the following research questions:

- 1) Is PF supported by the literature as an EBP according to the WWC single-case design standards?
- 2) How is PF used in educational settings to support teachers' implementation of classroom interventions?
- 3) How do different moderators change the effect of PF?
- 4) Does PF increase the teachers' implementation of Opportunities to Respond?
- 5) What effect does PF have students' on-task behavior (PF → Teacher increase in OTR → student behavior)?

# **THE IMPORTANCE OF DELIVERING PERFORMANCE FEEDBACK TO TEACHERS TO IMPROVE PRACTICE: AN EVALUATION OF THE QUALITY OF THE LITERATURE**

## **Introduction**

**Review of problem.** The implementation of EBP in schools remains problematic (Cook & Cook, 2013; Massar, 2018; Weston & Bain, 2015). This concept describes the RTPG (Massar, 2018) that indicates these practices are not widely adopted (Brock & Carter, 2017) and when adopted, not sustained (Kraft, Blazar, Hogan, 2018), and when sustained not implemented with fidelity (Cook & Odom, 2013). Research has shown that this gap may cause an increase in student reprimands and teacher burn-out (Wang, 2017). Commonly, PD workshops are used to combat the RTPG. However, teachers do not feel confident implementing these EBP after participating in PD (Kim, Koegel & Koegel, 2017).

One proposed solution is training with specific components such as in consultation or coaching (Denton & Hasbrouck, 2009; Garcia, 2015; Joyce & Showers, 1982; and Kraft, Blazar, Hogan, 2018). Components such as modeling and PF have been consistently associated with larger effects in both single case and group designs. (Brock & Carter, 2017; Brock et al. 2017). Literature reviews found that PF (of a variety of types) improved teacher behavior in the use EBP such as praise and opportunities to respond in both group and single case research designs (Cavanaugh, 2013). PF also appears to moderately improve the treatment integrity of interventions in general and special education teachers across grade levels and skill types (Solomon, Klein, &

Politylo, 2012). Follow up performance evaluation using data appears to be a robust method in education and psychology (Noell, Gansle, Mevers, Knox, Mintz & Dahir, 2014).

According to Fallon and colleagues (2015) PF is an EBP for improving educator's fidelity of implementation. However, the term is vague, used inconsistently, and often studied as a component in a treatment package, thus creating disagreements with the field (Brock & Carter, 2017; Brock, et al., 2017; Cavanaugh, 2013; Kraft, Blazar, & Hogan, 2018; and Solomon, Klein, & Politylo, 2012). PF encompasses a personal interaction between two people with an oral, written, or gestural communication regarding progress towards the desired outcome. PF is important because teachers need support when using EBP and classroom strategies to effectively implement them. Using observation and PF loops, teachers and school staff are supported by having intense contact hours and an understanding of how well they are implementing the intervention. Providing ongoing support and feedback is the most direct, effective way to produce high-quality implementation (Downer, et al., 2013). Additionally, Baker (1983) reported that when PF was used, teachers maintained high rates of performance six months after.

Four studies previously looked at PF in some context. A review of group design studies conducted by Brock and Carter (2017) looked at teacher training to improve the implementation of classroom interventions. The results concluded that when modeling and PF were used to train teachers, increased implementation was observed. Brock and Colleagues (2017) conducted another meta-analysis of 118 single-case studies for

teachers or paraprofessionals of students with learning disabilities or Autism Spectrum Disorder (ASD). Using visual analysis, they determined 521 effects out of 626 AB phases, with a d-hedges (Hedges, Pustejovsky, & Shadish, 2013) effect size of  $d=2.48$ . A dissertation on PF looked at PF to increase teachers' use of praise statements and opportunities to respond (Cavanaugh, 2013). Findings from the review conclude that PF is an effective strategy to increase the number of praise statements in teachers. Another review sought to identify the effects of PF on teachers' treatment integrity across studies utilizing single-case-research design (Solomon, Klein, & Politylo, 2012). The results suggest that PF is effective for increasing integrity and maintaining the high-quality implementation of the intervention. While these reviews are important to the field, the data is restricted to the targeted outcomes (i.e. LD, ASD, OTR, and Praise).

Fallon and colleagues (2015), conducted a What Works Clearinghouse (WWC) review on PF across 47 studies with 169 cases in articles published prior to 2011. The review identified 102 cases that met WWC standards with reservation, with 54 cases showing strong evidence of effect, and 48 cases showing moderate evidence of effectiveness. The authors conclude from the study that PF can be identified as an EBP aligned with the WWC standards (Fallon, Collier-Meek, Maggin, Sanetti, and Johnson, 2015), however, the article presented several limitations. First, articles published after 2011 are not included in the review. Secondly, the prior review indicated a lack of a concrete definition for what PF includes. One major concern was that the article did not limit the review to studies that isolated PF but included PF with additional strategies

such as instruction or modeling (Gage, Grasley-Boy, and MacSuga-Gage, 2018). Thus, the Fallon and colleagues (2015) review causes doubt for if PF is a truly EBP.

**Working definition of performance feedback.** While PF is defined broadly with heterogeneous components and attributes, for the purpose of the study the definition is operationalized and specified to educational setting. The operational definition is a personal interaction between two people with the oral, written, or gestural communication targeting/regarding progress towards a desired outcome. This definition aligns with the literature search and inclusion criteria described below.

**Study purpose.** Recent studies and literature reviews found PF increases implementation fidelity of classroom interventions and strategies in teachers with concomitant promising results for increased outcomes in students' academics and behaviors (Kraft, Blazar, & Hogan, 2018). The heavy interest in using coaching models to replace traditional professional development models means many schools have implemented PF without knowing if it is supported as an EBP. EBP are the gold standard when districts are looking to implement any intervention whether it be teacher-focused or student-focused. Therefore, identifying PF as an EBP for improving teachers' use of interventions across settings and rates of feedback is crucial information when considering using this strategy to support teachers and improve their practices. One way to identify an intervention or practice as evidence-based is by applying the What Works Clearinghouse (WWC) standards to the body of research available. As an initiative of the U.S. Department of Education Institute of Education Sciences (IES), WWC serves to identify and report on the "findings of scientific evidence for what works in education,"

(Institute of Education Sciences, 2014). In 2014, WWC created a protocol for assessing the rigor or quality of research on a particular strategy/intervention to determine if that strategy can be deemed as an EBP and recommended to be used in education.

The current review applied the WWC standards for single-case research to determine if PF is support in the literature as an EBP for improving teachers use of classroom interventions. This review seeks to answer two research questions: RQ1: What is the quality of research for PF according to WWC standards for single-case research? RQ2: How is PF used in educational settings to support teachers' implementation of classroom interventions?

## **Methods**

**Literature search procedures.** Working with a research librarian, the author conducted a systematic review of the literature through a two-phase process. First, we identified potential articles across three databases using search terms that were unique to that database's thesaurus terms in the fall of 2019. For the Education Resources Information Center (ERIC) database, the research librarian used the search phrase (TI teacher\* OR AB teacher\*) AND (TI (performance n2 feedback) OR AB (performance n2 feedback)) and yielded 535 articles. For the Academic Search Ultimate database, the research librarian used the same phrase as in the ERIC search and yielded an additional 171 articles that were not present in the ERIC search. Finally, the research librarian conducted a search in PsycINFO using the search phrase (DE "Elementary School Teachers") OR (DE "High School Teachers" OR DE "Junior High School Teachers" OR DE "Middle School Teachers") OR (TI teacher\* or AB teacher\*) AND (TI (performance

n2 feedback) OR AB (performance n2 feedback), which yielded 367 new articles. These searches were restricted to only include articles published in peer-reviewed journals. No gray literature or dissertations were included, this concept is aligned with WWC standards (Institute of Education Sciences, 2014).

The research librarian initially imported the ERIC search into RefWorks (Marsalis & Kelly, 2008), a commercial web-based reference management service that can be used by researchers to house literature searches. RefWorks was used to easily removed duplicates (i.e. dedupped; Foster, 2018), and create a backup to save the searches and progress. The searches obtained from Academic Search Ultimate and PsycINFO were imported into RefWorks and dedupped. The remaining articles (n=1,702), were imported into Rayyan to conduct interrater reliability and determine the articles for inclusion. Rayyan is a web/app-based reference management service that allows multiple raters to quickly sort through articles against an inclusion criterion while being unaware of the decisions of other raters (i.e. blind rating).

Forty-eight articles included after title and abstract review were moved to full text review. During full review, articles were re-reviewed by reading the complete article to ensure that they met the inclusion criteria. Ten articles met inclusion criteria after full text review. Three articles were excluded for not using a single-case study design. Another five articles were excluded because the participants were not school staff (i.e. children or parents as participants). Another 30 articles were excluded because the design of the study did not allow for the effects of PF to be observed in isolation without packaging PF with instruction, modeling, or some other variable. Studies needed to have

clear adjacent AB phases with A being baseline and B being PF sessions, or have a return to baseline phase before implementing PF. These 30 articles did not conduct studies that followed this format. Some of these studies conducted baseline observation, delivered a PD (instruction), then introduced PF sessions. Other articles were combining multiple components of a coaching model (i.e. modeling and PF, or self-monitoring and PF) all at once. While these concepts are efficient for conducting PD and training teachers, it only allows the researcher to see the effects of the combined components and not PF in isolation.

***Phase two.*** The 10 articles that met inclusion were used to identify additional articles during the second phase. The second phase consisted of an ancestral search and a forward search of each included article using SCOPUS (Burnham, 2006). Scopus is an online literature search engine that allowed the researcher to identify publications from the reference section of an article (ancestral search), as well as publications that cited the specific article (forward search). The second phase search rendered an additional 1,062 articles, after removing duplicates. Title and abstracts were reviewed against the inclusion criteria. Articles that were included during the second phase of title and abstract review were again examined during full text. An additional five articles were included during this phase. A total of 15 articles were included in this quality review (See Figure 2.1).

**Inclusion criterion.** The inclusion criterion included seven statements used to decide if an article met threshold. First, the article had to be written or translated into English, any article not available in English was excluded (n=1). Secondly, the article



described a single-case experimental design. Publications that were group study designs, literature reviews, meta-analysis, surveys, books/book chapters, white papers, or practitioner articles were excluded (n=453). Next, the author looked at the independent variable (IV). PF had to be the IV or part of the IV package, articles that did not include PF as the IV was excluded (n=263). Fourth, the study had to take place in a public, private, charter school, or an alternative school setting. Articles where the study took place in clinical settings, juvenile setting, hospice care, residential or home care, medical base settings, or athletic setting were excluded (n=150). Along with the type of school setting, articles were only included when the study took place within the Kindergarten to 12th grade. Articles that focused on early education, job-related education, or higher education settings were excluded (n=57). Sixth criteria focused on the person receiving the feedback. Studies where general education teachers, special education teachers, or para-educators/school aids were the focus of PF was included, articles with any other type of participants were excluded (n=68).

The final criterion focused on the progression of instruction, observation, and feedback as they pertain to the baseline condition. This review focused on studies that delivered instruction before conducting baseline and observation trials, then introduced PF during the intervention phase. Studies that implemented instruction and PF together after collected baseline data was excluded (n=32). By excluding these studies, the review was able to look at the quality of PF in isolation as opposed to a package that encompasses PF along with other components such as coaching.

**Selection and application of quality indicators.** A quality indicators review was conducted using the What Works Clearinghouse (WWC) standards for single-case research (Institute of Education Sciences, 2014). WWC created a protocol for assessing the rigor or quality of research on a particular strategy/intervention to determine if that strategy can be deemed as an EBP and recommended to be used in education. This protocol was influenced by the quality indicators identified by Horner and colleagues (2005). Originally this protocol was designed to assess studies that employed a randomized control trial study design. However, in 2014 WWC created an appendix that outlined how to assess the rigor and quality of studies utilizing single-case research. The WWC standards for single-case research address the following criteria: (a) Is there experimental control, (b) was interobserver agreement (IOA) collected, (c) was IOA collected for 20% of each phase, (d) did IOA meet criteria (i.e., 80% agreement or 0.60 Cohen's kappa), (e) were there at least three demonstration of effects, and (f) were there sufficient data points. One has to be able to identify the type of single-case design that the author(s) used to determine if the study design has sufficient data points. Reversal or withdrawal designs, the study must have at least four phases with three to five points in each phase. If the study uses a multiple baseline or multiple probe design, then the study must have at least six phases with at least three data points in each phase. A study using an alternating treatment design, there needs to be at least four points for each condition with no more than two points per phase. Additional considerations must be met when using a multiple probe design: i) there must be at least one data point for each case at the beginning of the experiment, ii) at least one data point needs to be present directly

before and directly after a participant enters intervention. Studies are scored on a three-point scale as meeting standards, meeting standards with reservations, or not meeting standards.

**Interrater reliability procedures.** Training raters involved the use of 10 practice articles to build understanding across the research team and revise inclusion criteria when necessary. The research team consisted of one faculty member, three doctoral students, and one master students who were proficient in using performance feedback for teacher training. Interrater reliability (IRR) was conducted on 100% of the 636 articles derived from the three databases. Agreements were calculated by dividing the number of agreements by the sum of disagreements plus agreements, then multiplied by 100. Point-by-point agreement was 91.5% for screening articles during titles and abstract review. Disagreements were discussed with research team and final decisions were made. Interrater reliability was conducted on 92.1% of the 38 articles during full text review. Point-by-point agreement was 97%. Disagreements were discussed and final decisions were made to determine articles that meet full text inclusion.

The same research team was used to conduct IRR during the second phase of the literature. IRR was conducted on 37.6% (n=400) of the 1062 articles reviewed during the second phase. Point-by-point agreement was 93.4% for reviewing articles during titles and abstract. The research team discussion all disagreements and a final decision was made. For the full text review during the second phase, IRR was conducted on 100% of the articles. Point-by-point agreement for full text of the second phase literature search was 92%. Disagreements were resolved by the research team.

Interrater reliability for WWC standards was conducted by the same research team as the literature review. Reliability was conducted on five of the 15 articles that were included in this review (30%). IRR agreement was 87% for WWC quality indicators. The disagreements were due to two WWC indicators; "Each outcome is measured overtime by more than one assessor, with inter-assessor agreement collected in each phase and in 20% of data points," and "Attempts to demonstrate effect over time and data points per phase." Disagreements were discussed with the research team and unanimous decisions were made.

## **Results**

A literature review was conducted on the use of PF to increase teachers' success of implementing classroom interventions and strategies. The review identified a total 2,764 articles across the two phases and 15 articles met inclusion criteria and was included in the review. The review described how PF was used in the field of education across number of participants, participant type, the intervention implemented by PF, the immediacy of PF, and the dosage of PF (See Appendix B). Finally, an evaluation of the literature was presented using WWC standards for single-case research.

**Study characteristics.** This review included 15 articles with 89 single case AB phases. Most of the articles utilized a multiple baseline design across participants or settings (n=13). Two of the studies used a variation of a A/B/A/B design (Sanetti, Luiselli, & Handler, 2007; and Zoder, Dufrene, Sterlind, Tingstrom, Blaze, Duncan, & Harpole, 2013). This review included articles from 10 different author groups, with Noell authoring four studies, Witt co-authoring two of those articles. Fallon authored

two studies, Dufrene authored two studies, and Sanetti authored another two articles. Articles were published from 11 different journals between the years 1982 and 2018. Three articles were published by School Psychology Review, two articles were published by Psychology in the schools, two articles were published in Journal of Applied Behavior Analysis.

**Participant characteristics.** A total of 55 participants were identified. Most of the studies enrolled general education teachers as participants (n=10). Three studies recruited special education teachers as their participants (Coddling, Feinberg, Dunn, & Pace, 2005; McKenney & Bristol, 2015; Zoder, Dufrene, Sterlind, Tingstrom, Blaze, Duncan, & Harpole, 2013). Across the review, three participants were identified as paraeducators, teacher aids, or classroom support staff (Maggin, Fallon, Sanetti, & Ruberto, 2012; McKenney & Bristol, 2015; and Zoder, Dufrene, Sterlind, Tingstrom, Blaze, Duncan, & Harpole, 2013). Burns, Peters, and Noell (2008), used three problem solving teams as participants in the study. One study conducted by Maher (1982) used two schools as their participants.

Six articles did not identify the number of years teaching for their participants. Three of the articles had participants with under 10 years of experience. Coddling, Feinberg, Dunn, & Pace (2005), had participants whose highest year of teaching was 2.5 years, O'Handley, Dufrene, & Whipple (2018), had a participant with eight years of teaching. One of the articles expressed that the teachers had more than one year of teaching experience (Noell, Witt, Gilbertson, Ranier, & Feeland, 1997). Two articles enrolled teachers with 10-15 years of teaching experience (Leach & Conto, 1999 and

Maggin, Fallon, Sanetti, & Ruberto, 2012). A few of the studies recruited participants with 25 or more years of teaching experience (n=4).

Nine articles did not identify the level of education of the participants recruited in their studies. Maggin, Fallon, Sanetti, and Ruberto (2012), used participants with a high school education. Two studies employed teachers with bachelor's degrees (Coddling, Feinberg, Dunn, & Pace, 2005; and DiGennaro, Martens, & McIntyre, 2005). O'Handley, Dufrene, and Whipple, (2018) and Sanetti, Luiselli, and Handler (2007) enrolled teachers with master's degrees.

**Intervention implemented.** Within this review, researchers used PF to increase the participants' implementation across various interventions, strategies, and plans. Four of these studies used PF to increase the fidelity of plans or processes, rather than a specific intervention or strategy. Within two of the studies PF was used to increase the fidelity of implementation of behavior support plans (e.g. Coddling, Feinberg, Dunn, & Pace, 2005; and Sanetti, Luiselli, & Handler, 2007). Similarly, Noell, Duhon, Gatti, and Connell (2002) used PF to increase the fidelity of the participants implementing behavior management plans. One study implemented PF to increase the efficiency of problem-solving team processes during team meetings.

The remaining 11 studies used PF to either increase the frequency of implementation or the fidelity of implementation of a specific intervention. Performance Feedback (PF) was used to increase the use of teachers' praise statements (i.e. behavior specific praise, and positive praise) in three of the studies within this review. Two of the studies sought to increase the implementation of group contingency interventions, such

as class-wide token economies, through the use of PF. Sanetti, Fallon, and Collier-Meek (2013) sought to improve the fidelity of a self-monitoring by using PF with the participants. Within this review PF was used to increase the fidelity of discrete trial training (McKenney & Bristol, 2015), teachers implementing strategies to improve correct words per minute in their students (Zoder, Dufrene, Sterlind, Tingstrom, Blaze, Duncan, & Harpole, 2013), and implementing strategies to increase time-on-task in their students (Leach & Conto, 1999). Finally, two studies used PF to increase the teachers' use of a reinforcement-based intervention (Noell, Witt, Gilbertson, Ranier, & Feeland, 1997; and Witt, Noell, LaFleur, & Mortenson, 1997).

**Performance feedback characteristic.** Four of the studies within the review did not mention the immediacy of the performance feedback session after the observation (Maher, 1982; McKenney & Bristol, 2015; Reinke, Lewis-Palmer, & Merrell, 2008; and Zoder, Dufrene, Sterlind, Tingstrom, Blaze, Duncan, & Harpole, 2013). Another four studies delivered PF immediately after the observation sessions, while an addition two studies delivered PF the same day as the observation session. Three studies conducted by Burns, Peters, and Noell (2008), Noell, Witt, Gilbertson, Ranier, and Feeland (1997), and O'Handley, Dufrene, & Whipple (2018) delivered the PF sessions before conducting the next observations. Finally, there was two studies that delivered PF to their participants the same week as they conducted the observation sessions.

This review also described how often the studies delivered PF to the participants. Only two articles did not mention the dosage of PF delivered (Maher, 1982; and Zoder, Dufrene, Sterlind, Tingstrom, Blaze, Duncan, & Harpole, 2013). Most of the studies

delivered PF sessions daily (n=6). A few of the studies delivered PF weekly (Coddington, Feinberg, Dunn, & Pace, 2005; Maher, 1982; O'Handley, Dufrene, & Whipple, 2018; and Sanetti, Fallon, & Collier-Meek, 2013) or multiple times a week (Leach & Conto, 1999). Two of the studies included in the review delivered PF less than once a week (Coddington, Feinberg, Dunn, & Pace, 2005; and Sanetti, Luiselli, & Handler, 2007).

**WWC quality indicators evaluation.** According to the WWC standards for single-case research design, none of the studies identified in this review fully met standards (See figure 2.2.). However, several of the studies did meet WWC standards with reservation (n= 10). The 10 studies met WWC standards with reservation due to two quality indicators *2B: Was IOA collected for at least 20% of each phase* and *4: Are the number of data points in each phase appropriate for the study design*. Six of the studies meet WWC standards with reservation due to indicator 2B (Leach & Conto, 1999; Maggin, Fallon, Sanetti, & Ruberto, 2012; McKenney & Bristol, 2015; Noell, Duhon, Gatti, & Connell, 2002; Sanetti, Luiselli, & Handler, 2007; and Zoder, Dufrene, Sterlind, Tingstrom, Blaze, Duncan, & Harpole, 2013). These studies conducted IOA for at least 20% of all the data points collected across each condition but did not conduct 20% for each phase (i.e 20% IOA for each baseline phase as well as 20% IOA for each intervention phase).

Two studies met WWC standards with reservation due to indicator 4 (O'Handley, Dufrene, & Whipple, 2018; Reinke, Lewis-Palmer, & Merrell, 2008). Both of these articles used a multiple baseline design. These studies had only three or four data points in at least one phase within the design. The studies conducted by Burns, Peters, and



Noell (2008) and DiGennaro, Martens, and McIntyre (2005) were rated as meeting WWC standards with reservation due to both indicator 2B and indicator 4.

Five of the articles in this review did not meet WWC standards for single case research (Coddington, Feinberg, Dunn, & Pace, 2005; Maher, 1982; Noell, Witt, Gilbertson, Ranier, & Feeland, 1997; Sanetti, Fallon, & Collier-Meek, 2013; and Witt, Noell, LaFleur, & Mortenson 1997). These studies did not meet WWC standards due to one of two quality indicators, 2A: *Was IOA collect* or 3: *Does the study demonstrate three effects of the intervention*. The studies conducted by Noell, Witt, Gilbertson, Ranier, and Feeland (1997), Sanetti, Fallon, and Collier-Meek (2013), and Witt, Noell, LaFleur, and Mortenson (1997), did not report any type of IOA producer or data thus not meeting the criteria for WWC standards. The other two studies did not demonstrate three effects of the intervention (Coddington, Feinberg, Dunn, & Pace, 2005; and Maher, 1982). Both of these studies only demonstrated two effects of the intervention.

**Performance feedback as an evidence-based practice.** WWC minimum standards specify five studies, with 20 participants, across three author groups (Clearinghouse, 2014). Overall, 10 studies met WWC quality standards with reservations (Burns, Petters, & Noell, 2008; DiGennaro, Martens, & McIntyre, 2005; Leach & Conto, 1999; Maggin, Fallon, Sanetti, & Ruberto, 2012; McKenney & Bristol, 2015; Noell, Duhon, Gatti, & Connell, 2002; O'Handley, Dufrene, & Whipple, 2018; Reinke, Lewis-Palme, & Merrell, 2008; Sanetti, Luiselli, & Handler, 2007; and Zoder, Dufrene, Sterlind, Tingstrom, Blaze, Duncan, & Harpole, 2013). These 10 studies surpass WWC standards minimum. There are 10 different author groups across this

review. This review identified a total of 55 classroom staff, 3 school teams, and 2 schools. Given the criteria to determine an EBP, PF exceeds the conditions, meaning that PF would be considered an EBP according to the WWC standards for single case research.

## **Discussion**

This review sought to understand the quality of research on the use of PF to increase school staffs' fidelity of implementation. Additionally, the author sought to understand how PF was used across different types of school personnel, as well as the type of intervention these studies wanted to improve while using PF. Finally, the review showcased how these studies used PF in terms of immediacy of PF sessions and dosage of PF sessions.

According to the WWC standards for single-case research PF meets the criteria to be an EBP. The results from this review show that 10 studies met WWC standards with reservation, and that there were more than 20 participants across these studies. Results from this review align with results from the review conducted by Fallon and colleagues (2015). This review deepens the understanding of PF as an EBP by presenting the quality of literature that uses PF in isolation, rather than in a multi-component model that was previously investigated (Fallon, et al., 2015). These results present PF as an effective implementation strategy that can be used to achieve desired results when implementing other EBP in education. One of the goals of coaching is to get EBP into the hands of educators that would benefit their students with disability and equip these educators with the skills and wherewithal to implement them with high

fidelity. Using an evidence-based implementation strategy such as PF, gives coaches a model to use when training educators.

A commonality across the review was that PF was used more often to increase the fidelity of general education teachers than with special education teachers or paraprofessionals. These results align with the results Fallon and colleagues (2015) found with over half of the studies included in the review using general education teachers. However, a group meta-analysis conducted by Brock and Carter (2017) and a review on single-case designs (Brock, et al., 2017) found that more special educator and paraprofessionals participated in PF studies than general educators. While these results might seem contradicting, an interpretation can be made that PF is effective for both general and special educators (Brock & Carter, 2017; Brock, et al., 2017; and Fallon, et al., 2015). This information can be beneficial in that coaches can recruit heterogeneous groups when using PF without the threat of it being less effective. The findings promote coaches training both general educators and special educators together while using PF to increase the educators' fidelity of implementation. This notion aligns with the results from a study conducted Zoder and colleagues (2013) that implemented PF with general educators, special educators, and paraprofessionals and found promising results.

This review found that studies used PF to increase implementation of interventions that were behavioral based interventions (i.e. increasing opportunities to respond or implementing reinforcement) more often than interventions that were academic based (i.e. DIBELS) or curriculum based (i.e. STARR DTT). Findings from this review support previous research that PF is utilized more often when training a

specific practice, intervention, or skill rather than a curriculum, theory, or new pedagogy (Brock & Carter, 2017; and Brock, et al., 2017). Across the literature there are studies that use PF to implement more complexed practices such appropriate curricular focus or prescribed instructional sequence . However, the majority of the studies align with the findings from this review. PF is clear, concrete, and precise, which lends itself well as an implementation strategy to increase a specific intervention. When using PF for more complex practices, additionally coaching components such as modeling and rehearsal should be considered.

When analyzing how PF was used, more commonly the studies delivered PF daily and either immediately after the observation session or the same day as the observation sessions. Results from Fallon and colleagues (2015) had more variability in terms on immediacy and dosage of PF than presented in the current review. Yet, results from both reviews are aligned. Fallon and Colleagues stated that about 28% of their studies delivered PF daily, and 17% delivered sessions multiple times a week, while 20% of the studies delivered PF on a weekly basis. The findings from the current review showed that 40% of the studies delivered PF sessions daily. Over half of the studies within the current review delivered PF immediately after or the same day as the observation session (n=6), while the most prevalent immediacy in the former review was tied with same day and next day delivery at about 20% for each. The findings from both of these reviews do not present a clear understanding of the consistency of PF sessions in terms of immediacy and dosage, yet it does provide evidence that PF is effective despite how often you provide PF or when. This information gives coaches the flexibility when

conduct PF sessions to tailor these sessions to the need and the availability of their teachers.

**Limitations.** There were three major limitations in this review of the literature. First, while the review was conducted systematically with the support of a research librarian, there is still a possibility that an article that would meet the inclusion criteria was not included within this review. This review only included published peer-reviewed articles, therefore limited by not including grey literature and potentially missing a book chapter or dissertation that would have met the inclusion criteria. This was deliberately done for two reasons. The author wanted to align the search with WWC standards that tend to only use published peer-reviewed studies. Additionally, the author sought out literature that could be accessed by school staff. Many districts do not have access to online journals or scholarly databases; therefore, these districts are already at a disadvantage when looking for research. By including gray literature this disadvantage grows and makes it harder for teachers and coaches to find the articles they need to implement EBP or PF.

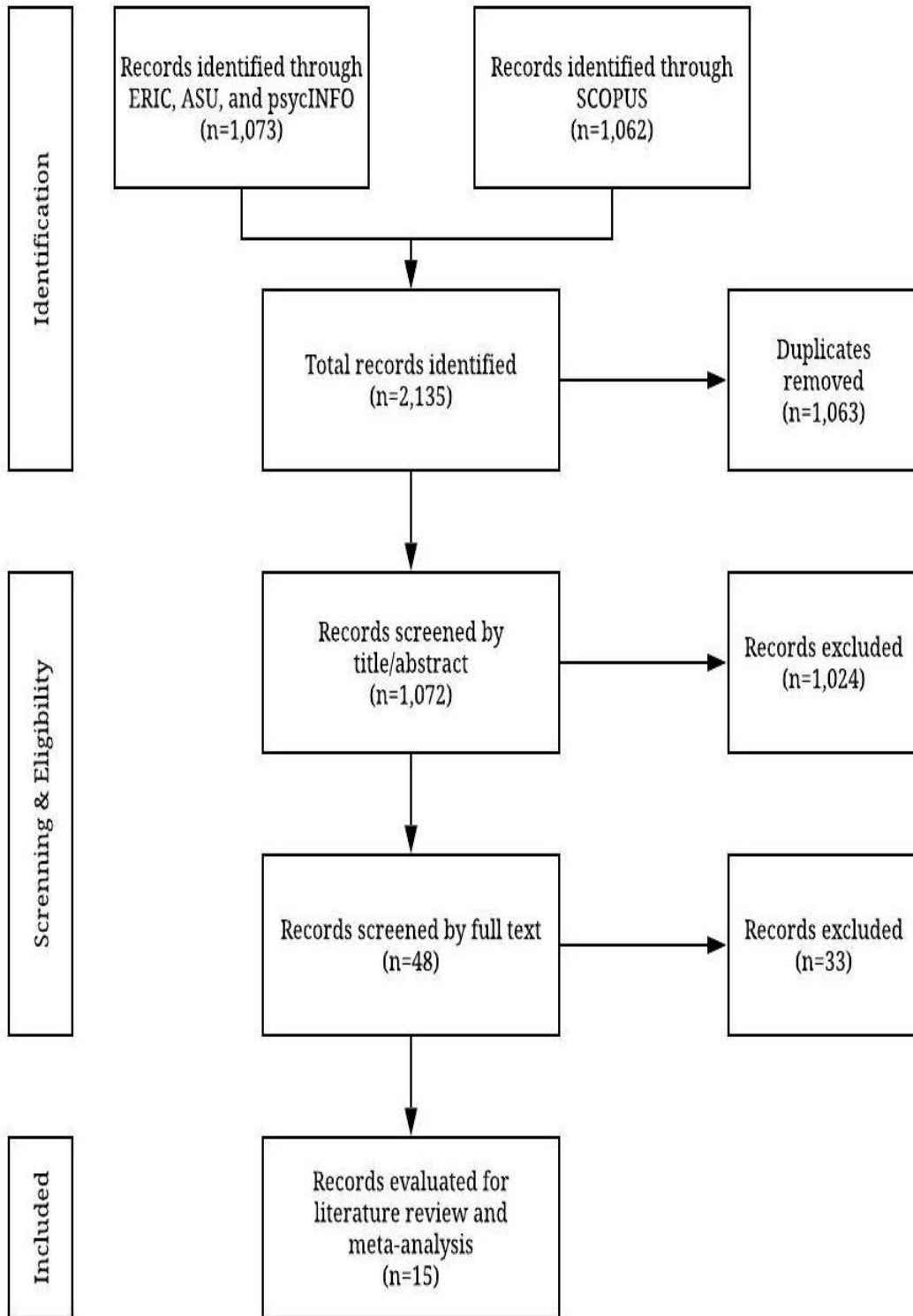
Secondly, the author only included studies that isolated the delivering of PF sessions from other coaching components such as instruction, modeling, or rehearsal. Additional studies that use PF in concert with other components are in the body of literature but not sought out for this review. Finally, the review focused on how PF was used with school personnel in educational settings. Understanding of how PF can increase fidelity of implementation with other groups such as parents, counselors, or peers cannot be concluded from the current review.

**Implications.** This review presented some important caveats for using PF in educational settings. Instructional coaches, supervisors, and professional development coordinators can use PF to increase the fidelity of implementation in their school personnel. The results from this rigorous evaluation using the WWC standards, allows coaches to be confident when selecting PF as the implementation strategy to obtain high fidelity in their teachers. With a high fidelity of implementation, school personnel can implement these EBP with assurance and have a better chance of achieving desired results among the students. Also, if school personnel are supported and coached on how to properly implement intervention, they are more likely to not be affected by burn-out and can have a bigger impact on classroom instruction and increase student success.

**Future research.** Future researchers should investigate the use of PF across different settings such as vocational settings, hospice care, or home base. Another avenue of research should look at different types of personnel that receive feedback, whether that be parents, community persons (i.e. clergy, athletic coaches, or after school staff), or extended relatives. Research should continue to investigate and conduct systematic reviews on aspects of PF that were not discussed within this current review. Looking at the impact the relationship between the person delivering PF and the person implementing the intervention is worth more investigation. Another aspect of investigation should involve looking at the lasting impact PF has on implementation long after the intervention has concluded (i.e. maintenance). Finally, research studies using single-case experimental design should continue being conducted across different

context to build this body of research as well as reinforce the use of PF as an EBP to support teachers' implementation of classroom strategies.

**Figure 2.1 PRISMA Flow Diagram**





**Figure 2.2 WWC Quality Indicator**

Article	Indicator 1	Indicator 2A	Indicator 2B	Indicator 2C	Indicator 3	Indicator 4
Burns	M	M	MWR	M	M	MWR
Codding	M	M	MWR	M	DNM	M
DiGennaro	M	M	MWR	M	M	MWR
Leach	M	M	MWR	M	M	M
Maggin	M	M	MWR	M	M	M
Maher	M	M	MWR	M	DNM	M
McKenney	M	M	MWR	M	M	M
Noell (A)	M	M	MWR	M	M	M
Noell (B)	M	DNM	N/A	N/A	M	M
O'Handley	M	M	M	M	M	MWR
Reinke	M	M	MWR	M	M	MWR
Sanetti (A)	M	DNM	N/A	N/A	M	M
Sanetti (B)	M	M	MWR	M	M	M
Witt	M	DNM	N/A	N/A	M	M
Zoder	M	M	MWR	M	M	M

M= meets, MWR= meets with reservation, DNM= does not meet, N/A= information not available.

## **PERFORMANCE FEEDBACK TO INCREASE TEACHER PRACTICES: A SINGLE-CASE META-ANALYSIS**

Results from the literature review support PF as an evidence-based implementation strategy for increasing teacher's use and fidelity of classroom strategies. However, it is still important to determine how effective PF is for increasing teacher's implementation and the relative effects if one type of treatment condition produces larger effect sizes than another. A search of the coaching literature found more than 13 reviews or meta-analyses from 1997 until the most recent, 2018 (e.g. Kraft, Blazar, & Hogan). The author recognizes that while PF may be a sub-component within coaching interventions, PF was not the direct intervention under investigation. Additionally, a mega-analysis (meta-analysis of meta-analyses) of the entire coaching literature to identify PF components, is beyond the scope of this dissertation. Instead, only the experimental manipulation of PF as a variable of primary interest is the topic of interest for this study.

When looking at PF outside of coaching models, there are currently four reviews dating between 2012 and 2017 (Brock & Carter, 2017; Brock, et al., 2017; Cavanaugh, 2013; Solomon, Klein, & Politylo, 2012). These include studies through 2016 indicating a minimum of four years of unevaluated work in the professional literature. The existing meta-analyses have identified several potential contextual variables in addition to naming PF as an EBP.

Furthest from the scope of the current meta-analysis, Brock, and Carter (2017) conducted a meta-analysis on group design studies of training to improve the

implementation of classroom interventions. The study included 12 articles with an overall mean effect size of  $g=1.08$ . Hedges'  $g$  is commonly used when sample sizes are under 20 participants. The equation for Hedges'  $g$  involves the difference in the means scores from the control and experimental group divided by the pooled standard deviation ( $SD_{pooled}$ ). The  $SD_{pooled}$  differs from a basic  $SD$  in that, the  $SD_{pooled}$  is the weighted average of each group's  $SD$ . The weighting gives larger sample sizes more effect on the overall estimate. The results concluded that when modeling and PF were used to train teachers, increased implementation was observed.

Two reviews looked at PF in single-case research but restricted the scope to a specific outcome variable. One meta-analysis evaluated 118 single-case studies for teachers or paraprofessionals of students with learning disabilities or Autism Spectrum Disorder (ASD; Brock, et al., 2017). Using visual analysis, they determined 521 effects out of 626 AB phases, with a  $d$ -Hedges (Hedges, Pustejovsky, & Shadish, 2013) effect size of  $d=2.48$ .  $D$ -Hedges uses a hierarchical model to produce a between-subject effect size. This effect size computation works best for withdrawal designs or multiple baseline designs that provide adequate variability across cases to yield an estimate of variance greater than zero.  $D$ -hedges was designed to correspond to the parameters of Cohen's  $d$ , a commonly used effect size for group studies. A dissertation on PF looked at PF to increase teachers' use of praise statements and opportunities to respond (Cavanaugh, 2013). Findings from the review conclude that PF is an effective strategy to increase the number of praise statements in teachers, the effectiveness PF has on opportunities to respond was ambiguous from the results of this review.

The most similar meta-analysis to the scope of the current meta-analysis sought to identify the effects of PF on teachers' treatment integrity across studies utilizing single-case-research design (Solomon, Klein, & Politylo, 2012). The results suggest that PF is effective for increasing integrity and maintaining the high-quality implementation of the intervention. The authors did find changes in effects when moderated for the intervention (Solomon, Klein, & Politylo, 2012).

There are several limitations in the existing literature hampering a complete understanding of PF as an intervention to increase implementation fidelity. First, two meta-analyses looked at PF in relationship to single dependent variables of interest (i.e. Opportunities to respond and Praise; Cavanaugh, 2013), or with specific and exclusive student populations. (i.e. Learning Disabilities and Autism; Brock, et al., 2017). Secondly, Brock and Carter (2017), used a narrow methodological focus (e.g. only reviewed PF in studies that used group designs). Solomon, Klein, & Politylo (2012), expressed limitations in the selection and application of parametric effect sizes for single case (e.g., did not use non-parametric ES such as NAP and Tau).

The meta-analyses currently in print do not adequately show evidence in comparing the impact of different moderators such as number of participants, immediacy of PF, or dosage of PF. Thus, the literature lacks the understanding of important factors a coach should consider when using PF. Such as does PF work better for small groups of three participants or less, or can PF still be effective when upscaled. Another considering factor is whether PF is more successful with general education instructors or special education instructors. Additional evidence is needed to determine the optimal immediacy

and dosage of PF session teacher's need to increase the implementation of the targeted intervention. A few studies have eluded to various contextual variables of PF that change the effectiveness of teachers' implementation of an intervention. However, no comprehensive review is available to validate how effective are each of these variables. Furthermore, Solomon, Klein, and Politylo (2012) stated that any variation of PF appears at least somewhat effective in changing teachers' behavior. These contextual variables are vastly different and have the potential to alter the effectiveness of PF, thus needs to be looked at comparatively.

### **Study Purpose**

The purpose of this meta-analysis was to look through the literature on PF in educational settings to answer the following research questions: RQ1: What are the effects of performance feedback on teacher's implementation of fidelity directed classroom strategies and programs? RQ2: What contextual variables of performance feedback produce the best effect on teacher's implementation fidelity?

### **Method**

**Literature search.** The literature search was conducted in two phases. The author conducted the literature search with support from a research librarian specializing in systematic reviews. A literature search was conducted during the Fall 2019 semester. The first phase of the literature search identified potential articles using three databases, Education Resources Information Center (ERIC), Academic Search Ultimate (ASU), and PsycINFO. A separate search was conducted in each database, using terms unique to the database. The search within the ERIC databased used the phrase (TI teacher\* OR AB

teacher\*) AND (TI (performance n2 feedback) OR AB (performance n2 feedback)). This search yielded 535 articles. When the literature search was conducted for ASU, the author used the same phrase (TI teacher\* OR AB teacher\*) AND (TI (performance n2 feedback) OR AB (performance n2 feedback)) and yielded 171 articles that were not found during the ERIC search. The final search in phase one was conducted using the PsycINFO database. For this database, the author used the search phrase (DE "Elementary School Teachers") OR (DE "High School Teachers" OR DE "Junior High School Teachers" OR DE "Middle School Teachers") OR (TI teacher\* or AB teacher\*) AND (TI (performance n2 feedback) OR AB (performance n2 feedback)). This search yielded an additional 367 articles that were not gathered from the previous two searches. All of the searches only included published, peer-reviewed articles. No gray literature, book chapters, or dissertations were included in this search.

***Phase Two.*** The second phase was conducted using an online literature search engine called SCOPUS (Burnham, 2006). This site allowed the author to conduct an ancestral search and forward search of articles that met inclusion. An ancestral literature search looked through each of the articles' work cited pages to sort through the cited references. A forward search compiles a list of references that have cited a specific article. The SCOPUS search yielded 1,062 articles.

**Inclusion criteria.** An article needed to meet seven criteria to be included. First, the article had to be written in English. Any article that was not written in English or was unable to obtain a translated copy was rejected from the study (n= 1). Next, the article had to use a single-case design, articles using group designs, quasi-experiments, or case-

studies were not included. Additionally, articles that were literature reviews, meta-analysis, books/book chapters, survey data, white papers, or practitioner papers were rejected from the review (n= 453) Third, the independent variable (IV) described in the article had to include PF. Articles that did not use PF as the IV were rejected (n= 263). The next criteria involved the type of setting in which the study was conducted. Articles describing the setting as a public, private, charter, or alternative school were included. Articles that described the setting in a clinic, hospital, juvenile, residential, or home base were rejected. Additionally, studies conducted in a medical field or with athletes were also rejected. One hundred and fifty articles were rejected that did not meet the setting criteria. The fifth criterion focused on the grade level in which the study took place. Studies that were conducted between grades Kindergarten and 12th were included. Articles that conducted studies in early education, higher education, or job-related settings were rejected (n= 57). The next criteria pertained to the type of participant that received the performance feedback. The study had to have at least one participant who was a general education teacher, special education teacher, or a supporting staff (i.e. paraprofessional, teacher aid) to be included. Studies using counselors, behavior therapist, or outside personnel were rejected (n= 68).

The last criterion involved the progression of instruction, observation, and feedback to the baseline data. The author focused on studies that delivered instruction before conducting baseline and observation trials, then introduce PF during the intervention phase, or studies that conducted an additional baseline condition after training. Studies that implemented instruction and PF together after collected baseline

data were excluded (n= 32). By omitting these studies, this review can look at the quality of PF in isolation as appose to a package that encompasses PF along with other components (See Figure 2.1.).

**Title and abstract review.** The research librarian imported the results from all three searches in an online web-based reference management service called RefWorks (Marsalis & Kelly, 2008). RefWorks was used by the librarian to combine to the searches from ERIC, ASU, PsycINFO, and SCOPUS databases and remove duplicated articles. After de-duping (Foster, 2018), 1,702 articles remained. These articles were imported into another online reference management system called Rayyan. This system allowed for multiple raters to sort through all or a percentage of the articles to determine if an article met inclusion for the review. Rayyan allowed the rater to see various parts of the article such as the authors, title, the journal, and the abstract. The raters used this information to determine if an article should be included during the title and abstract review. After the title and abstract review was complete, 48 articles were included.

**Full-text review and quality indicators.** All 48 articles were then put through a full-text review where the author and raters took an in-depth look through each article to ensure that the article met the inclusion criteria. The raters used a google forms sheet to conduct the full-text review. The form had a place to mark that the article met inclusion against each of the criteria, as well as an evaluation of the study design using the WWC single-case standards (Institute of Education Sciences, 2014). An article had to meet all of the inclusion criteria and meet WWC single-case standards or meet WWC single-case



standards with reservation to be included. A total of 15 articles were included in this meta-analysis.

**Variable codes.** Articles that met inclusion criteria and WWC single-case standards with or without reservations (n=15) were coded according to three coding variables. These variables included a) type of participants, b) immediacy of feedback, and c) dosage of feedback. These variable codes were split into two categories, participants' characteristics, and PF characteristics.

The first group of coding variables was participant characteristics. This group included the type of participants. The author recorded the participants' job title as either general education teachers; special education teachers; or support staff (paraprofessionals, classroom aids, teacher aids). The author could select multiple types, or the "other" tab for types such as teams (i.e. Burns, Peters, & Noell, 2008; and Sanetti, Luiselli, & Handler, 2007).

The second group of coding variables was PF characteristics. There were two codes in this category, the immediacy of feedback, and the dosage of feedback. The immediacy of feedback was to identify the length of time between the observation and the 'PF session. The author selected either immediately after the observation, the same day as observation, the week of observation, or that immediacy was not mentioned in the article. For the dosage of feedback, the author identified how often the participant received feedback during the study. The author selected either daily, multiple times a week, weekly, less than once a week, or not mentioned.

**Effect size calculation.** Data were extracted from the appropriate graphs within all of the included studies (n=15) using an online plot digitizer called WebPlotDigitizer (Rohatgi, 2017). Using this website, the author imported a screenshot of a graph showing only one case (i.e. data from one participant) from one of the included studies. Next, the author identified two points on the x-axis and two points on the y-axis to calibrate the digitizer to the graph. Then the author clicked on each data point on the graph in chronological order. The website would record the x, y coordinate, and produce a spreadsheet of the outputs. Finally, the author reviewed spreadsheets and rounded the x-axis data points to the next nearest integer, to reflect the study design. For example, if a data point had an  $x=3.658$ , the author adjusted to simply show as session 4. However, the y-axis values were not adjusted, even if the value was impossible to obtain for the given variable in the study (i.e. decimal number for a frequency count or the number of steps completed correctly). The author extracted data from 89 AB contrasts across the studies. After extracting data using WebPlotDeigitizer, the author inputted the AB contrast into a free online Single-case effect size calculator (Pustejovsky & Swan, 2018). The online calculator estimates non-overlap effect sizes, as well as parametric effect sizes. This calculator was used to compute Tau scores.

Tau scores are an extension of Nonoverlap of all pairs (NAP). NAP uses AB comparisons to determine the percentage of data that improve from baseline to intervention phase. This calculation uses all the pairwise AB comparisons and determines the number of positive pairs (i.e. improving pairs) as well as half of the pairs that tied and divides by all pairs (Parker, Vannest, & Davis, 2011). Tau extends NAP by

subtracting the number of overlapped pairs (i.e. negative pairs) from the positive pairs in the numerator. This added variable creates a different equation reflected as the positive pairs minus negative pairs divided by the total number of pairs (Parker, Vannest, & Davis, 2011). Tau have been used frequently in single-case research to compute effect sizes for studies as well as utilized in meta-analyses (Bowman-Perrott, Burke, Zaini, Zhang, & Vanest, 2016; and Whalon, Conroy, Martinez, & Werch, 2015).

**Statistical analysis.** Using the R studio software statistical significance, effect size means, and confidence intervals were calculated to analyze differences between the studies (n=15) for an omnibus as well as individual cases (n=79) during moderator analysis. For the omnibus, the effect size scores were calculated by averaging the effect sizes of all participants within the study. Calculations were run using Tau scores. To test homogeneity, the author ran a random effect (RE) model and created a Forest plot. A RE model estimates the overall mean effect size ( $\hat{\mu}$ ). A RE model was used because it was assumed that the effects sizes included in this meta-analysis do not have the same distribution and that all the effect sizes are not equal. The Forest plot compares effect sizes by representing them vertically with their confidence intervals. Confidence intervals (CI) are represented by a box along each line. Smaller boxes mean that there is a larger CI equating to a less precise effect size measurement, larger boxes imply a smaller CI equating to a more precise effect size measurement (Borenstein, et al., 2011).

Tests for publication bias were conducted using a Funnel plot and Egger's Regression test. Publication bias statistically describes the difference between the results available in publications and the true results of the independent variable. Articles that

describe their intervention to have desired effects on the dependent variable tend to be published more frequently than those with lesser effects. Thus, may lead to an inflation of the true effect of the independent variable. Tests for publication bias cannot confirm bias or that additional effect sizes exist that are unpublished. Whereas, publication bias tests provide evidence for or against publication bias. With a Funnel plot, a graphic is created plotting the effect sizes with a funnel/pyramid-like backdrop. If the effect is caused by a single variable and all effect sizes are reported, the plot would look like a funnel with one side symmetrically reflecting the other side. If the graphic has many effect sizes outside of the funnel and the sides are asymmetrical, then there is evidence to support publication bias (Borenstein, et al., 2011). A second test, Egger's Regression test, was conducted to provide an additional source in determining publication bias. This test takes the funnel plot, turns it on its side, and determines the slope. If the slope is zero than there is evidence against publication bias. This is a null test, in that failing to reject the null results in evidence supporting publication bias (Borenstein, et al., 2011).

Finally, RE models were run for the moderators collected within this meta-analysis to determine if PF has a greater effect with one category than another. Effect sizes were weighted at the case level ( $n=.79$ ). This was designed to have each replication equal regardless of the particular study or number of data points per participants. Estimates were used to compare a) general education instructors vs special education instructors, b) immediate or same-day delivery of PF sessions vs delayed delivery, and c) daily or multiple times a week PF sessions vs weekly or less dosage). *Appendix A* provides a full syntax of all codes used in R studio.

**Inter-rater reliability.** Inter-rater reliability (IRR) was conducted at three different points during this meta-analysis; title and abstract review, full text and WWC quality indicator review, and moderator coding and data extraction. Before reviewing, 12 practice articles were used to build understanding across the research team, revise inclusion criteria when necessary, and to revise coding form for better reliability. The research team consisted of one faculty member, three doctoral students, and one master students who were proficient in using performance feedback for teacher training. IRR was conducted on 60.8% of the 1,702 articles that were imported into Rayyan to conduct the title and abstract review. A point-by-point agreement was used for all IRR, meaning that if both raters gave the same response for one point of the inclusion criteria or the coding sheet, then that one point would be considered an agreement. This method created between 14 to 48 different points to which assess agreement for each article. An agreement was then calculated by dividing the number of agreements by the sum of disagreements plus agreements, then multiplied by 100. The agreement was 92.45% during the title and abstract review. Disagreements were discussed with the research team and a final decision was made.

The same research team conducted IRR on 83% of the 48 articles during full-text review and WWC quality indicator analysis. The point-by-point agreement was 92%. Disagreements were discussed and final decisions were made to determine articles that meet full-text inclusion. Most of the disagreements were due to the two WWC indicators i) "Each outcome is measured over time by more than one assessor, with inter-assessor agreement collected in each phase and 20% of data points," and ii) "Attempts to

demonstrate effect over time and data points per phase." Another point often disagreed upon was the seven criteria, the study delivered instruction before conducting baseline trials. To increase understanding of this concept the author conducted additional training with the research team. This training included examples and nonexamples of the inclusion criteria as well as practice articles.

The third point in which IRR was conducted was moderator coding and data extraction. Reliability was conducted on 30% of the 15 articles included in the meta-analysis. IRR agreement was 94% for moderator coding. Most of the disagreements were due to the name of the intervention and the immediacy of feedback. Not all of the studies gave a specific name of the intervention being implemented, this caused confusion among the coders. Similarly, some of the studies didn't state when the PF sessions happened in conjunction with the observation session. Disagreements were discussed and a final decision was made.

A point-to-point agreement was used for IRR on 20% of the data extraction resulting in 120 points across three articles. Agreement for data extraction was 99%, with one disagreement amongst coders. An agreement was determined if both coders recorded the x, y coordinate within one integer of each other. For example, if one coder recorded a data point as 4.5, 89.7 and the other coder recorded the same data point as 3.9, 90.2 then this was considered an agreement. However, if either of those differences was greater than 1, the data point would be considered a disagreement. A brief meeting was held with the author and the secondary coder to resolve the disagreement and make final decisions.

## Results

Fifteen articles published between 1982 and 2018, involving 10 author groups were included in this meta-analysis. Results from this meta-analysis look at the overall effect size (omnibus effect) of the 15 included studies, as well as comparing cases (n=79) against three different moderators: type of participants, the immediacy of PF, and dosage of PF. A total of 89 AB contrast across 55 participants were used to conduct the effect sizes. *Figure 3.1* presents a table of the effect size scores, at the study level (n=15), for Tau and standard error.

**Figure 3.1 Study Effect Size**

Study	Tau ES	Tau SE
Burns	0.932	0.068
Codding	0.552	0.204
DiGennaro	0.947	0.053
Leach	0.307	0.147
Maggin	0.881	0.120
Maher	1.000	0.000
McKenney	0.332	0.146
Noell (A)	0.182	0.288
Noell (B)	0.396	0.232
O'Handley	1.000	0.000
Reinke	0.889	0.065
Sanetti (A)	0.250	0.177

**Figure 3.1 Study Effect Size Continued**

Study	Tau ES	Tau SE
Sanetti (B)	0.000	0.337
Witt	0.190	0.260
Zoder	0.595	0.322

**Omnibus effect.** A Forest plot of the effect sizes was run using Tau scores at the study level. *Figure 3.2* shows the Forest plot of the effect sizes of the included studies. None of the effect sizes were negative, with a range of 0.00 to 1.00 and a mean of 0.61 CI [0.42, 0.80]. Seven of the studies had a CI that included zero, meaning that this effect size was not statistically significant (Maggin, Sanetti, & Ruberto, 2012; Noell, Duhon, Gatti, & Connell, 2002; Noell, et al., 1997; Sanetti, Fallon, & Collier-Meek, 2013, Sanetti, Luiselli, & Handler, 2007; Witt, Noell, Lafleur, & Mortenson, 1997; and Zoder-Martell, et al., 2013). A visual analysis of the Forest plot revealed boxes of various sizes. The large boxes express a smaller CI, equating a more precise effect size measurement, while smaller boxes express a large CI, equating to a less precise effect size measurement (Borenstein, et al., 2011). Using the RE model, the omnibus effect of PF on teachers implementing classroom interventions and strategies was  $\hat{\mu}= 0.6116$ , CI [0.4249, 0.7984],  $p<.0001$ . The CI did not include 0 and the p-value is less than 0.0001, suggesting an effect size of statistical significance. Results from the test for heterogeneity were  $Q(df=14)= 86.0344$ ,  $p<.0001$  indicating that this sample is homogeneous.



**Publication bias.** Two tests were conducted to help determine if there is publication bias within the literature, a Funnel plot, and an Egger's Regression test. *Figure 3.3* illustrates the Funnel plot analysis of the effect sizes using the standard error. The figure shows several of the effect sizes partially or fully outside of the funnel, with the two sides being extremely asymmetrical. The visual analysis of the Funnel plot gives evidence to publication bias within this body of literature.

The second test conducted was an Egger's Regression test. This test reveals a z score along with a p-value. The null hypothesis implies that there is publication bias. Therefore, failing to reject the null would result in evidence that there is publication bias. The test computed a score of  $z=8.9374$  with a p-value of .0006. Because of this value, we fail to reject the null, thus providing evidence in support of publication bias.

**Figure 3.2 Forest Plot**

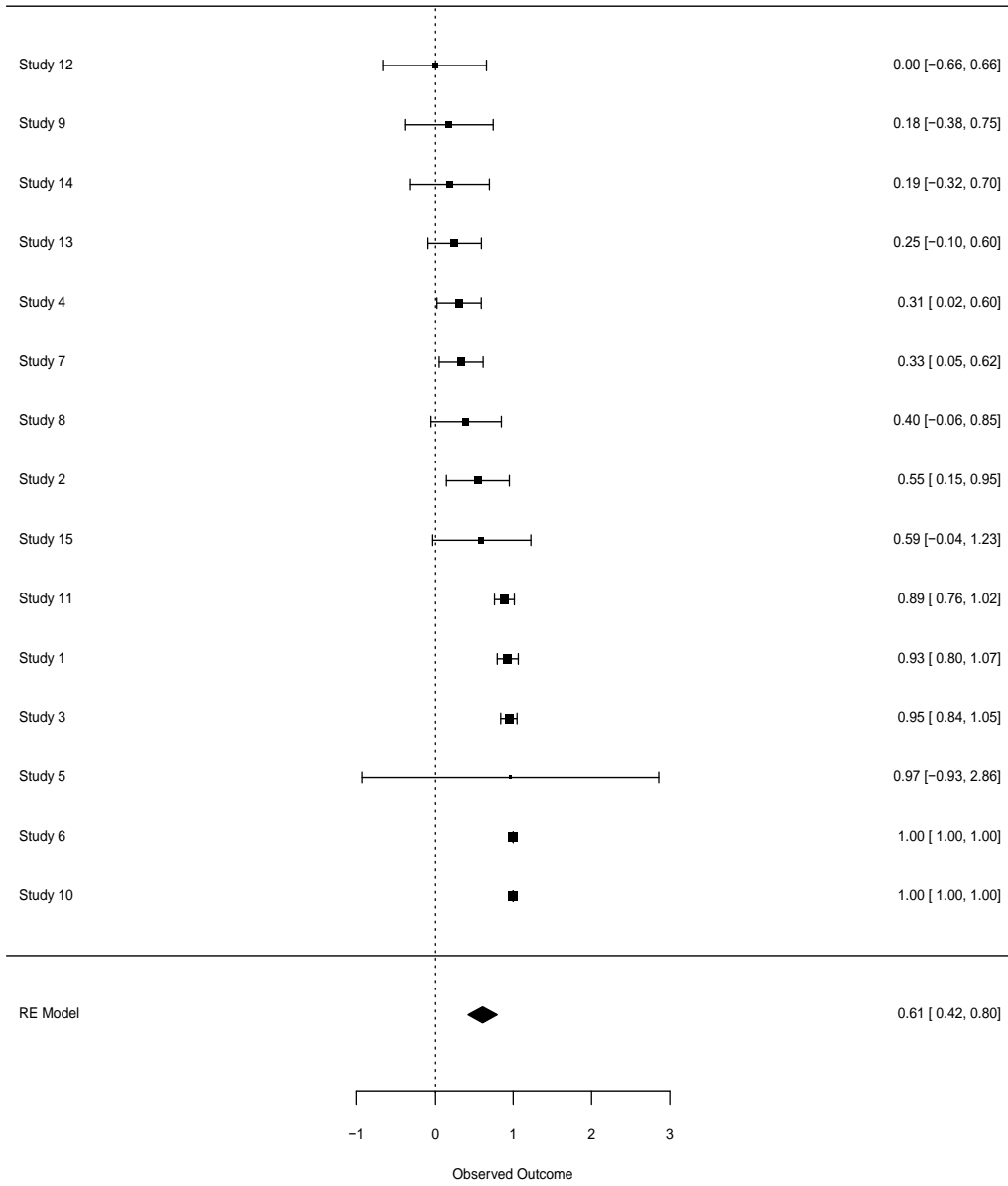
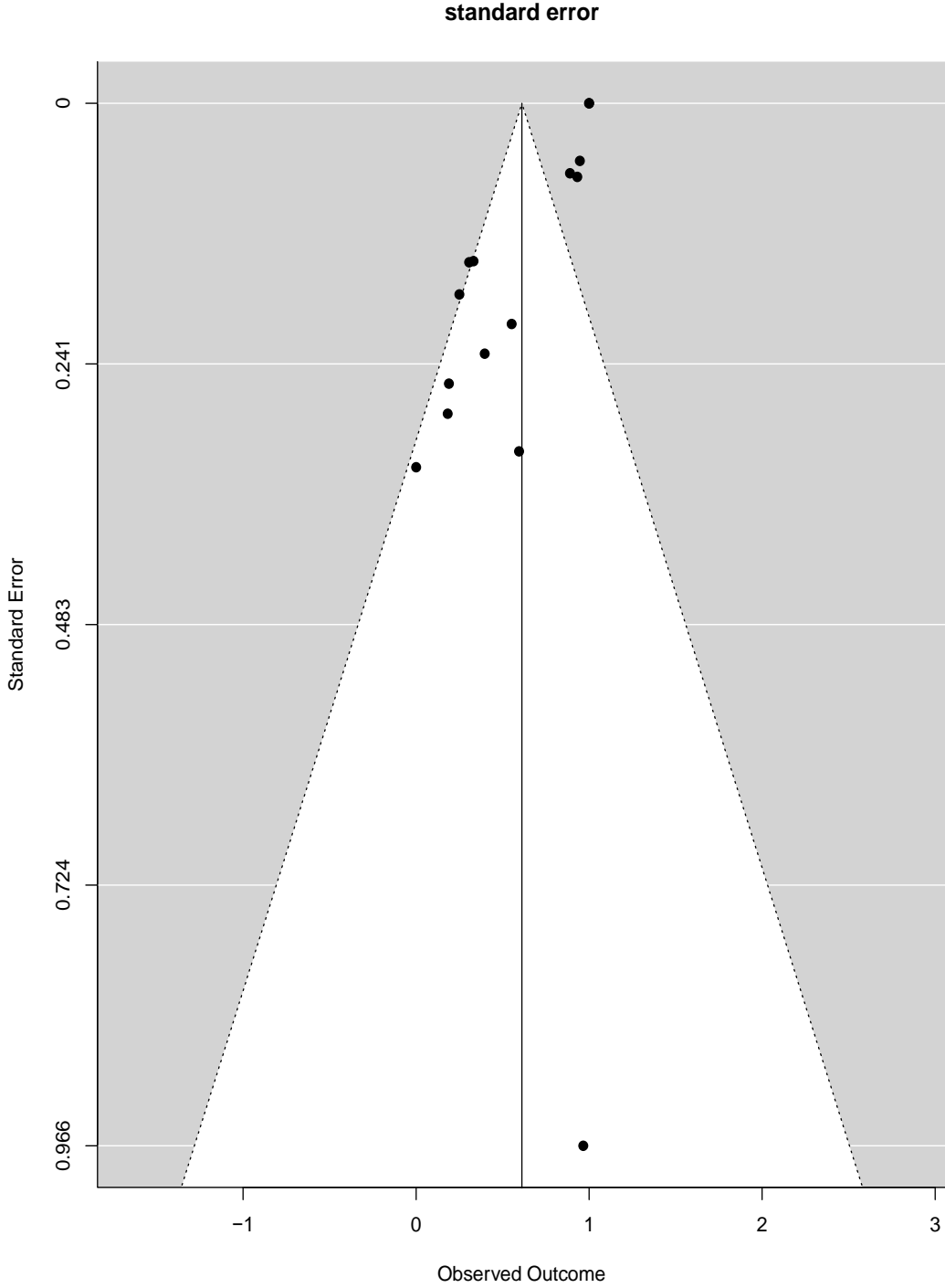


Figure 3.3 Funnel Plot



**Moderator variable comparison.** RE models were run on four moderating variables; type of participant (gen ed. vs sped), the immediacy of PF (same day vs delayed), and dosage of PF (multiple times a week vs once a week or less). Concerning the type of participants, studies were coded as a 1 if they utilized general educator instructors, or as 2 for studies that used special education teachers or aids. Two studies were coded as 3 and not used in the moderator analysis because they employed "teams" as their participants (Burns, Peter, & Noell, 2008; and Sanetti, Luiselli, & Handler, 2007). Coding for the immediacy of PF created three codes. Code 1 for studies that delivered PF immediately after the observation session or the same day as the observation session. Studies that delivered PF sessions before the next observation session or the same week of the observation session were coded as 2. Code 3 was reserved for studies that did not mention the immediacy of PF and was not included in the moderator analysis. The last moderator categorized how often the study used PF sessions. Code 1 was reserved for studies delivering PF sessions on a daily or multiple time a week basis. Studies delivering PF sessions on a weekly or less than once a week basis were coded as a 2. Two studies were coded as a 3 because the dosage of PF sessions was not mentioned in the article (McKenney & Bristol, 2015; and Zoder, et al., 2013). Results for moderator RE models slightly differ when using NAP scores or Tau scores. Results are described for each moderator comparison using NAP and Tau scores.

***Type of participant.*** Nine studies used general education instructors as the participants, four studies used special education teachers or teacher aides, and two studies used school groups like entire schools or school teams. RE model was run to

analyze the comparison between studies using general education instructors and studies using special education instructors. Results from the heterogeneous test of general education cases were  $Q(df=46) = 122.9375, p < .0001$ . The Tau score is statistically significant, meaning that there is uncertainty in the heterogeneity of this group. Results for studies using special education instructors were  $Q(df=21) = 21.4793, p = .4300$ . Some variability still exists in this group that is not controlled by this moderator. Mean estimates were the same for both the general education group and the special education group. Mean estimate was  $\hat{\mu} = 1.00, p < .0001, CI [0.9999, 1.0001]$ . This estimate is statistically significant. Tau estimates show no difference between moderator groups.

***Immediacy of PF.*** Six studies delivered PF sessions either immediately after the observation session or the same day as the observation session. Five studies delivered PF sessions either before the next observation session or at least the same week as the observation session. Four studies did not mention the immediacy of delivering PF sessions and were not included in the RE model. Results from the Q test for immediate delivery were  $Q(df=26) = 41.4033, p = .0283$ . We reject the null, expressing some variability in this group. For studies using a delayed delivery of PF sessions results were  $Q(df=32) = 89.8214, p < .0001$ . This p-value indicates that this subgroup is homogenous. RE model estimates were the same for both groups, yet the groups did have a slight difference in their confidence intervals. For immediate delivery mean estimate was  $\hat{\mu} = 1.00, p < .0001, CI [0.9983, 1.0016]$ . Mean estimate was  $\hat{\mu} = 1.00, p < .0001, CI [0.9999, 1.0001]$  for studies that used a delayed delivery. These mean estimates are statistically significant.

***Dosage of PF.*** Seven studies conducted PF sessions daily or multiple times a week, while six studies conducted PF sessions weekly or less than once a week. Two studies did not mention the dosage of PF and were not included in the moderator analysis. RE model was run to compare the studies using a higher dosage of PF with the studies using a lesser dosage of PF. Results from the Q test for the higher dosage group was  $Q(df=26) = 50.0751, p=.0031$ . We reject the null, stating that some variability exists amongst the studies in this group. Results for studies using a lesser dosage of PF sessions were  $Q(df=38) = 81.7173, p<.0001$ . Results for this group show that little variability exists within the group. RE model mean estimates were the same for both subgroups with the same confidence interval and p-value. Mean estimates were  $\hat{\mu} = 1.00, p<.0001, CI [0.9999, 1.0001]$  The estimates are statistically significant.

## **Discussion**

Related to the purpose of this study, the author found that PF was effective overall as well as within the moderators. Seven of the studies were not statistically significant (Maggin, Sanetti, & Ruberto, 2012; Noell, Duhon, Gatti, & Connell, 2002; Noell, et al., 1997; Sanetti, Fallon, & Collier-Meek, 2013, Sanetti, Luiselli, & Handler, 2007; Witt, Noell, Lafleur, & Mortenson, 1997; and Zoder-Martell, et al., 2013). The other eight studies had a statistical significance that these results were not by chance, despite a few studies demonstrating weaker effects. The omnibus estimate shows a strong effect for using PF with teachers to increase implementation, the moderator analysis provided little details to answer for whom and under what condition PF session

might be more effective. This meta-analysis investigated three moderator variables: a) type of participants, b) immediacy of PF sessions, and c) dosage of PF sessions.

First, the author investigated the effects of employing general education instructors or special education instructors. Each moderator group had statistically significant effects. Results for this moderator variable show no difference between groups. Nonetheless, effects sizes are extremely strong (Vannest & Ninci, 2015), indicating that PF sessions can be effective despite the type of participant that is being employed. These results align with results published by Cavanaugh (2013) that stated PF was an effective implementation strategy for teachers across grades from kindergarten through high school as well as both general and special educators. However, results from a meta-analysis conducted by Solomon, Klein, and Politylo (2012) found that elementary school general education teachers implemented their interventions with higher rates of fidelity than high school teachers and special education teachers. Although the authors continued to state that grade alone was not a significant factor. Despite the conflict amongst the reviews, these results are still promising. The results demonstrate that school staff is responsive to PF, despite the grade level or population that they teach.

The author evaluated the effects of delivering PF sessions immediately versus delaying the delivery of PF sessions. Results from both moderator groups were statistically significant. Analysis of both moderator groups produced extremely strong effects. While these results suggest that immediacy of PF does not have a large impact on the effectiveness of PF as an implementation strategy, a past meta-analysis (Solomon, Klein, & Politylo, 2012) suggested that PF should be delivered as close to the

observation session as possible (i.e. immediately after). Results from their meta-analysis found little difference between studies that used immediate feedback and studies that delivered same-day feedback. Studies that implemented PF every week were less effective, though the authors noted the difference was not significant. Theoretically, delivering PF soon after the observation (i.e. within 24 hours) allows the teacher to more easily reflect on the feedback and make changes to the teaching quicker. However, weekly feedback may be more sustainable to coaches that have a larger group of participants and can keep the cost of coaching low. Being able to manipulate the immediacy and dosage of delivering PF without heavily impacting the effectiveness of this implementation strategy, gives coaches and administrators the leeway to tailor PF to the needs and availability of the teachers.

Finally, the author investigated how the dosage of PF sessions impacts the effect of using PF. Studies conducting PF sessions daily or multiple times a week were compared to studies conducting PF sessions every week or less. Both estimates illustrated an extremely strong effect (Vannest & Ninci, 2015). The NAP scores produced the strongest effects. Both estimates were statistically significant. These results are slightly conflicted. The estimated mean scores show no difference between the group. Despite showing no difference, these results are favorable, indicating the possibility that coaches can deliver PF sessions every week and still achieve desired results. This can free up time and involve fewer contact hours with each participant.

This meta-analysis provides understanding and extends previous research on using PF to increase implementation fidelity with teachers. There are already meta-



analyses and literature reviews investigating the impact PF session when in concert with an additional variable (Noell, et al., 2014) or when using PF sessions to train a particular skill (Cavanaugh, 2013). However, this is the first meta-analysis to look at PF sessions as a single variable removing the combined effect often seen in coaching packages (Kraft, Blazar, & Hogan, 2018). Investigating this PF variable in isolation is important because understanding how effective PF sessions are alone will help coaches when constructing implementation interventions with their teachers. Furthermore, understanding which PF variables have the best potential in producing the strongest effect allows coaches to manipulate these variables while maintaining confidence in the potential results.

**Limitations.** There are several limitations that the author has identified while conducting this meta-analysis. First, some of the moderator groups had small sample sizes were and compared them to unequal sample sizes groups. This may have screwed the results and created inadequate reporting of some of the findings. Additionally, some of the information was not available for moderator analysis. Another limitation is that this meta-analysis focused on only single-case studies due to a lack of effect size calculations that combine single-case and group design research with great confidence. Finally, this meta-analysis did not analyze the generalization or maintenance conditions of the included studies. This limitation was due to more than two-thirds of the articles not having generalization or maintenance data to analyze. Therefore, this meta-analyze does not provide information on how PF can be generalized or retains the effects after the conclusion of the intervention.

**Future research.** Two areas warrant future research identified by this meta-analysis. More comprehensive literature search and moderator reporting. This meta-analysis only used peer-reviewed literature thus creating strong evidence for publication bias. Future researchers should conduct literature searches that include gray literature such as book chapters and dissertations, as well as look for innovative methods to increase confidence that all articles that should be included are included. Research on this topic would help the field better understand the true impact PF has on teach implementation fidelity. Secondly, moderator coding is important to help understanding how to use an intervention most effectively and with whom to use it. Addition moderator analysis is needed in determining how to best use PF. Researchers should continue conducting experiments with PF yet begin to manipulate these moderating variables as well as reporting these results consistently.

## **USING PERFORMANCE FEEDBACK TO INCREASE TEACHERS' IMPLEMENTATION OF OPPORTUNITIES TO RESPOND**

The Research-to-Practice Gap (RTPG) is continuously acknowledged by researchers and administrators in the field of education (Brock & Carter, 2017; Cook & Odom, 2013; Kim, Koegel, & Koegel, 2017; Kraft, Blazar, & Hogan, 2018; Massar, 2017; and Wang, 2017). A review of the related literature suggests that the gap in research and practice may partially come from the teachers' inability to hold accurate perceptions of their teaching practices and the frequency with which they use these practices (Guckert, Mastropieri, & Scruggs, 2016; and Jones, 2009). One way to improve perceptions of practices and frequency of using these practices is PF. PF is considered an EBP for improving educator's fidelity of implementation (Fallon, Collier-Meek, Maggin, Sanetti, & Johnson, 2015).

Using PF involves a personal interaction with the oral, written, or gestural communication regarding progress to the desired goal. PF takes place after the behavior is performed (Cooper, Whitney, & Lingo, 2018), as appose to prompting that takes place before or during the behavior (Wang, 2017). PF involves an observation of the teacher to collect data on implementation and sharing the data with the teacher to improve future performance (Brock & Carter, 2017). Denton and Hasbrouck (2009), suggest PF should emphasize the appropriateness of an intervention as well as the implementation fidelity of the intervention.

One can use PF to increase the teacher's use of a specific intervention. One intervention that is supported by research as a classroom management strategy is

Opportunities to Respond (Cavanaugh, 2013; Garcia, 2015; and Katzenbach, Shuster, Shafer, Lloyd, & Carter, 2016). Opportunities to Respond (OTR) is a teacher-implemented strategy in which a teacher offers an instructional question or statement that promotes a student(s) response (Cavanaugh, 2013; and Garcia, 2015). In practice, a teacher gives a statement or question, waits for the student to respond appropriately, then the teacher provides positive feedback to the student. There are two general types of OTRs, verbal responses, and non-verbal responses. Verbal responses can include students answering questions, reading passages aloud, or repeating important statements given by the teacher. Non-verbal responses can include students writing down answers, using thumbs up or down, or using response cards. A teacher can use OTR to prompt responses from an individual student or a group of students (pairs, small workgroups, or whole class). Research states that OTR is most effective when teachers use three to five OTRs per minute for simple responses and as low as one OTR for every 15-30 minutes for more complex responses like a writing prompt (Katzenbach, Shuster, Shafer, Lloyd, & Carter, 2016).

A meta-analytic study showed increased teacher's use of OTR (Cavanaugh, 2013) through the use of PF. A multiple baseline study utilized PF to increase OTR and behavior-specific praise statements with teachers instructing students with emotional behavior disorders (i.e. behavior support teachers). The study found that PF increased behavior-specific praise statements and OTR with the teachers, and increased on-task student behavior (Duchaine, Jolivet, & Fredrick, 2011). However, OTR was a secondary dependent variable and not the focus of the study. This current study

expanded upon Duchaine, Jolivette, and Fredrick (2011) by using the same population (behavior support teachers) and PF on OTR. Yet, the current study used OTR as the primary dependent variable while also collecting on-task classroom behavior data. This study sought to answer two research questions; Does performance feedback increase the teachers' implementation of Opportunities to Respond? What effects do high rates of Opportunities to Respond have on students' on-task behavior?

## **Method**

**Participant recruitment and inclusion criteria.** The current study was approved by the University Internal Review Board (IRB) and procedures were agreed upon with the school district's director of special education. IRB protocol number was IRB2019-0920D reference number 096327. All procedures, materials, consent forms, and data collection was approved and stamped before participants were contacted. The author contacted the special education director for contact information of intermediate schools', middle schools', and high schools' principals. The author met with the principals individually to inform them of the study and ask for contact emails of teachers that worked with students with behavior problems. These teachers were sent a handout. The handout provided the purpose of the study, a brief description, the participant criteria, and the procedures. The author set up an in-person meeting with the participants who met the criteria and showed interest in the study to further review the study and sign informed consent.

Four participants were recruited from one high school and one intermediate school within the same school district. The inclusion criteria for these participants

included the following. 1) Participants were involved in instruction and employed by the school district for the entire school year. 2) Participants worked in a behavior support classroom or classroom that supported students with behavior/emotional disorders and/or Autism. Within this particular district, these participants were identified as SOAR teachers, which stood for "Social Outcomes and Academic Readiness". 3) Participants demonstrated a need for increased levels of OTR. Participants were excluded from the study if 1) participants demonstrated high rates of behavior (a frequency of 4 or more OTRs per minute) or 2) participants were not employed by the school district and/or did not work within a behavior support classroom (i.e. outside consultants or behavior therapists).

**Classroom characteristics.** Four participants were enrolled in the study. Two of the participants were high school SOAR teachers, while the other two participants were intermediate school SOAR teachers. Both schools were located within the same school district in the southern part of the United States. The high school classrooms were outfitted with a teacher desk, a desk for the paraeducators, two additional tables used for student workspaces, and several student desks in various arrangements. These classrooms had access to a projector and student laptops to supplement learning. There was a whiteboard in the front of the class and various visual aids on the walls about current academic concepts (i.e. Multiplication charts, PEMDAS, world maps). These classrooms were self-contained, in that the students stayed in the same classroom the entire day unless a student was assigned an "out class". The teachers and paraeducators also stayed in the same classroom and taught all content throughout the day. The

classrooms had anywhere from one to five students depending on out classes, absences, and behavioral consequences such as suspensions, court-mandated therapy, or students simply walking out of the classrooms.

The two intermediate classrooms were outfitted differently than the high school classrooms. These classrooms had two to three tables that functioned as a teacher desks, a paraeducator desks, and additional workspaces for the students. One of the tables in each classroom was typically used to conduct group work with two to three students at one time. Behind these tables was an off-limits space to the students where supplies, curriculums, and other instructional materials were house and available for the teacher or paraeducators to grab. Each student had a cubby style desk along one wall of the classroom that was assigned to them for independent work. The classrooms had a whiteboard on another wall that was used for instruction, to write to-do lists for academic subjects, and keep track of students' points (class-wide token economy points). The remaining walls contained visual aids for good behavior, proper cell phone usage, class schedules, current math concepts, commonly misspelled words and posters about the benefits of reading. There was no student laptops or projectors in these classrooms. Each teacher had a personal computer that they also used to show academic videos and PowerPoint slides. While the classrooms were self-contained for the students and the paraeducators, the teachers switched classrooms after lunch. There were typically three students in one of the classrooms and four students in the other classroom. The number of students fluctuated ( $\pm 2$ ) depending on absences, suspensions, and temporary alternative placements.

**Participant characteristics.** Participants are described in the order in which the participant entered their intervention phase within the study design. Participants' names were changed to pseudonyms to maintain confidentiality. The first participant to enter the intervention phase was a 22-year-old Caucasian female, Ms. Harrington. Ms. Harrington earned her bachelor's degree in special education. She started as a first-year SOAR teacher at an intermediate classroom. The second participant was a 23-year old Caucasian female, named Ms. Sully. She held a bachelor's degree in special education. Ms. Sully started as a first-year teacher in the second intermediate SOAR classroom.

The third participant to enter the intervention phase was a 35-year old Caucasian female named Ms. Reveille, who worked in a high school classroom. Ms. Reveille was a special education teacher for four years. This was her first year as a SOAR teacher. Ms. Reveille held a bachelor's degree in psychology. The last teacher to enter in the study was a 30-year old Hispanic male named, Mr. Gill. Mr. Gill worked in the other high school classroom. He was a special education teacher for a total of seven years, with the last three years as a SOAR teacher. Mr. Gill has a bachelor's degree in education.

**Procedures.** Before baseline data were collected the teachers were given a handout on OTR. The author used the handout (Katzenbach, Shuster, Shafer, Lloyd, & Carter, 2016), to define OTR, describe how OTR can be used in the classroom, give examples and non-examples of different types of OTR, and research about the use and rate of OTR. The author continued to explain OTR to the participants until each participant was able to create places in their lesson plans to incorporate OTR and give examples of the OTR that they would use during instruction. Ms. Harrington and Ms.



Sully received their handouts and information about OTR together, but Ms. Reveille and Mr. Gill received their handouts and information separately. After this instruction session, the author and each teacher identified a class period where instruction had the greatest chance of occurring. Ms. Reveille chose first-period science as her observation session. Mr. Gill chose third-period math as his observation session. Ms. Harrington and Ms. Sully chose the second period and fourth period. Ms. Harrington taught math second period and fourth period, while Ms. Sully taught language arts second and fourth period.

***Baseline sessions.*** Sessions occurred three to eight times a week depending on the number of observations in which instruction occurred. Instruction did not occur during all observations, in which case data was not collected. During observation sessions the researcher sat in the back of the classroom, unobtrusive to the students, to collect data. Data were collected for 15 minutes during all observations using a data collection sheet created for the study (See *Figure 4.2*). During the observation, the researcher tallied the number of OTR, as well as the types of OTR that the teacher performed. The researcher also collected data on the classroom's on-task behavior during the observation session. After the observation sessions, the researcher told the teacher "thank you" and confirmed the next observation session.

***Performance feedback sessions.*** During the intervention phase, the researcher conducted PF sessions the same day as the observation session. This usually occurred immediately after the observation session, however, sometimes the PF sessions happened after completing a session with a different participant. This small delay in PF sessions (less than three hours) only occurred if the teacher was still teaching long after

the 15-minute observation concluded, and the researcher did not want to interrupt learning to have the PF session. During PF sessions, the researcher showed the teachers their total OTR for that observation session as well as the rate of OTR per minute. The researcher then gave examples and parts of the observation that the teacher could have varied an OTR or performed an additional OTR. The PF session concluded by the researcher showing the teacher a graph of their data and asking the teacher to update their targeted rate of the OTR goal. After the PF sessions, the researcher told the teacher "thank you" and confirmed the next session.

***Procedural fidelity.*** A self-rating form was used to assess procedural fidelity based on a prior study where the social validity of direct observation for the fidelity of implementation was critically low (e.g., teachers became uneasy and less cooperative when an additional observer sat-in on their PF sessions). This lack of social validity in the observation and measurement system harmed the fidelity of implementation and the ability to draw inferences from the data. During the prior study, sessions with a second observer watching, caused the teachers to not buy-in to the suggestions and many times became defensive or stated excuses as to why the observation went as it did. When the researcher removed the second observer the PF sessions became more natural and intimate and the teachers received the PF more constructively.

Procedural fidelity was conducted by a checklist (*Figure 4.1*) that was initialed and dated by both the researcher and the teacher. These checklists were completed twice a week after each PF sessions for a total of 35% of all sessions. The checklist asked six questions to ensure that the PF session went as planned. 1) Did the coach collect data on

my use of the OTR strategy? 2) Did the coach present the data to me the same day as that observation occurred? 3) Did the coach explain the data? 4) Did the coach present an updated graph of my progress? 5) Did the coach provide me with feedback on my performance? 6) My current or updated goal is \_\_\_ OTRs within a 15-minute observation period.

*Social validity.* After the intervention condition, each of the participants was given a 6-question survey to provide the researcher information about how each participant felt about the intervention and if they would continue with the sessions or try coaching again. Each of the questions was open-end. The survey included the following questions. 1. Have you received any combination of observation & coaching before participation in the study? 2. What were the positive aspects (if any) of this coaching and feedback experience? 3. What were the negative aspects (if any) of this coaching and feedback experience? 4. Did the knowledge and/or skills gained (if any) influence your teaching after completion of the study? How so? 5. Is there a return on your investment of time in participating in coaching and feedback sessions? 6. What other aspects of this coaching experience would you like to tell the researcher? The survey was sent to each participant email's address using a google form. The researcher synthesized the answers and identify common concepts.

## Figure 4.1 Procedural Checklist

### Procedural Checklist

At the end of the debrief session please review this checklist with the teacher have the teacher initial and date, stating that each of these items were completed.

Did the coach collect data on my use of the OTR strategy?

Did the coach present the data to me the same day as that the observation occurred?

Did the coach explain the data?

Did the coach present an updated graph of my progress?

Did the coach provide me with feedback on my performance?

My current or updated goal is \_\_\_\_ OTRs within a 15-minute observation period.

\_\_\_\_\_  
Coach's initials

\_\_\_\_\_  
Date

\_\_\_\_\_  
Teacher's initials

\_\_\_\_\_  
Date



IRB NUMBER: IRB2019-0920D  
IRB APPROVAL DATE: 10/02/2019

**Study design and variables.** This study used a multiple baseline across participant design (Horner, et al., 2005). Using a single-subject research design, such as a multiple-baseline across participants, is suitable for this educational setting. Using this type of study design was selected to ensure there were at least three demonstrations of effects. A multiple-baseline design was used because teachers cannot undo the progress and skills they learned through the PF sessions. Therefore, a withdrawal design would not be appropriate for this study. The first participant, Ms. Harrington enter the intervention phase after five data points in the baseline phase. Ms. Sully entered the intervention phase after eight data points and Ms. Reveille after 11 data points in the baseline. The last participant to enter the intervention phase was Mr. Gill after 15 data points in the baseline phase.

**Dependent variables.** There were two dependent variables collected in this study, rate of OTR and classroom on-task behavior. For each participant, the rate of OTR was calculated by dividing the total number of OTR by 15 minutes and multiplying by 100. The researcher also documented the types of OTR that the teacher performed. Ten different types of OTR could be observed: a) verbal choral prompts (VCP), b) verbal individual prompts (VIP), c) verbal choral questions (VCQ), d) verbal independent questions (VIQ), e) non-verbal choral prompts (NCP), f) non-verbal individual prompts (NIP), g) non-verbal choral questions (NCQ), h) non-verbal individual questions (NIQ), i) response cards (RC), and j) other (*See Figure 4.2*). The data collection sheet split the observation session into three five-minute sections. Types of OTR and the different

sections was used to provide more detailed feedback to the participants rather than a data collection method.

**Figure 4.2 Data Collection Sheet**

**Teacher:** \_\_\_\_\_ **Observer:** \_\_\_\_\_ **Time:** \_\_\_\_\_ **Date:** \_\_\_\_\_.

This study defines OTRs as when the teacher provides verbal prompts or questions, to evoke an academic response by the students. OTRs includes the use of response cards, choral responding (Whole class answering), verbal responding, writing responding, or reading aloud to the class. Place a Tally Mark in the appropriate time frame box when you observe the teacher perform an OTR. At the end of the observation record the total number of OTRs for each time frame, the total number of OTRs for the entire session, and the number of OTRs per minute.

# of OTRs	0-5 minutes					5-10 minutes					10-15 minutes				
	Total: _____					Total: _____					Total: _____				
	VCP	VCQ	NCP	NCQ	RC	VCP	VCQ	NCP	NCQ	RC	VCP	VCQ	NCP	NCQ	RC
	VIP	VIQ	NIP	NIQ	Other	VIP	VIQ	NIP	NIQ	Other	VIP	VIQ	NIP	NIQ	Other
	_____ OTRs per minute														
	<b>Session Total:</b> _____														

For this study, on-task is defined as any single or combination of (a) the student looking towards the teacher, (b) the student answering questions during instruction, and/or (c) the student adding alpha-numeric marks with a writing utensil on assignments. At the end of each minute scan the room and write the number of students that are on-task (at that moment).

Classroom Bx															
	1min	2min	3min	4min	5min	6min	7min	8min	9min	10min	11min	12min	13min	14min	15min

Total # of students \_\_\_\_\_

The second dependent variable was the classroom's on-task behavior. These data were collected using a composite of percentages for each 1-minute interval during the 15-minute observation session. At the end of each minute, the researcher counted the number of students that were on-task and recorded it on the data collection sheet (*Figure 4.2*). On-task behavior was operationalized as any single or combination of the students looking towards the teacher, the students answering questions during instruction, and/or the students adding alpha-numeric marks with a writing utensil on a giving assignment. To get a composite score, the researcher first computed the fraction of students that were on-task by dividing the number of on-task students by the total number of students in the

classroom during that interval, this number was left in decimal form. Then the researcher added each decimal and divided by the number of intervals (n=15), then multiplied by 100 to get the percentage of classroom on-task behavior during the 15-minute observation.

**Effect size calculation.** Effect sizes were calculated using TauU (Parker & Vannest, 2012). TauU is a nonparametric, free distribution, bottom-up analysis suitable for small data sets. A bottom-up analysis combines individual phase contrasts to form a single omnibus effect size that presents the entire design. This allows for consistency when interpreting results using visual analysis, because the focus is at each contrasts level instead of a top-down approach such as HLM or a multi-series regression (Parker & Vannest, 2012). Tau uses pairwise comparisons that result in a positive score (i.e. improvement from baseline to intervention), a negative score (i.e. regression from baseline to intervention), or a tie (i.e. no change from baseline to intervention) and then subtracts the number of positives from the number of negatives and divides by the total number of pairs (Parker, Vannest, Davis, & Sauber, 2011). A TauU score uses this formula and then controls for the upward trend in baseline (i.e. monotonic trend). Using an effect size that controls for baseline trends allows for a more sensitive indicator of the treatment effectiveness by removing or accounting for intervention phase values that would most likely occur if baseline would continue along with the same upward trend (Scruggs & Mastropieri, 2001). Currently, there is not a benchmark to help describe TauU effect sizes as small, medium, or large within the literature of PF. However, Vannest and Ninci (2015) do offer general guidance as followed; below 0.20 small

effect, 0.20-0.60 moderate effect, 0.60-0.80 large effect, and above 0.80 very large effects.

**Inter-observer agreement.** Inter-observer agreement (IOA) was collected 21.7% of all sessions during baseline and intervention phases for each of the four participants. IOA was collected by the researcher and two graduate students that were familiar with the research and trained to use the datasheet. IOA was calculated using point-by-point agreement across 49 points on the datasheet. An agreement was determined if both the researcher and the graduate student wrote the same number of tallies for each five-minute session, circled the same types of OTR in each five-minute session, or wrote the same number of students that were on-task at the end of each interval. At the end of the 15-minute observation session, the researcher and the graduate student reviewed their data collection sheets together. If a disagreement occurred, the disagreement was noted, the two of them discussed and decided on an answer. An agreement was determined by adding up the number of agreements, dividing by 45, and then multiplying by 100. *Figure 4.3* shows the IOA percentages for each participant for each phase. The average mean score across all phases and conditions was 95.625%.



**Figure 4.3 IOA Percentages Across conditions and Phases**

Teacher	Baseline	Intervention	Total
Harrington	96	98	97
Sully	93	97	95
Reveille	89	96	92.5
Gill	98	98	98
total	94	97.25	<b>95.625</b>

## Results

**Teacher behavior.** The results for the four participants are displayed graphically in *Figure 4.4*. All four participants increased their use of OTR after being introduced to PF sessions to some degree, with Mr. Gill showing the greatest improvement. Although all participants increased their OTR use, there is variability and several overlapping data points between the baseline and intervention phases. The overall mean score in baseline was  $m=1.6235$  OTR per minute and  $m=2.5915$  OTR per minute during the intervention. Baseline phase scores ranged from 0.4 OTR per minute and 2.6 OTR per minute across all participants. During the intervention phase scores ranged from 1.134 OTR per minute and 4.267 OTR per minute. The TauU describing the omnibus effect size across all participants was 0.7850, 90% CI [0.546, 0.988]. According to Vannest and Ninci (2015), this could be considered a large change between baseline and intervention.

A visual analysis of Ms. Harrington's performance implementing OTR shows a slight upward trend with great variability and several overlapping data points between

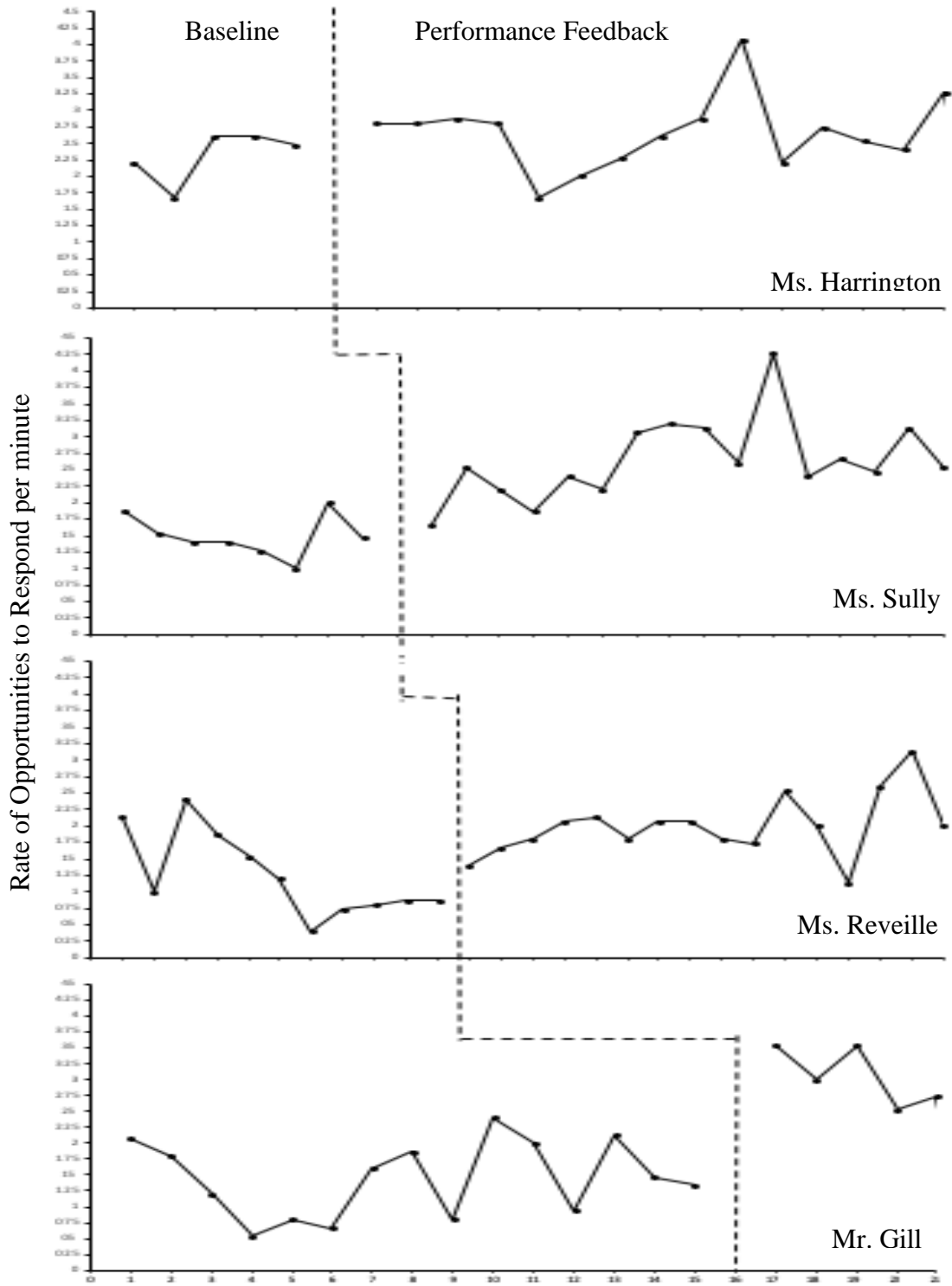
the baseline and intervention phases. This is mostly due to a max baseline score of 2.6 OTR per minute. The overall mean score in baseline was 2.307 OTR per minute and 2.657 OTR per minute during the intervention showing a slight increase. Data ranged from 1.667 to 2.6 OTR per minute in the baseline. For the intervention phase, Ms. Harrington's scores were higher, ranging from 1.667 to 4.067 OTR per minute. Ms. Harrington's TauU score of 0.411, 90% CI [0.103, 0.903] can be interpreted as a moderate change that aligns with the visual analysis of *Figure 4.4*.

*Figure 4.4* shows a visual analysis of the upward trend of Ms. Sully's use of OTR from baseline to intervention. Baseline data ranged from 1 to 2 OTR per minute with a mean score of  $m=1.492$ . During the intervention, Ms. Sully's scores ranged from 1.667 to 4.267 with a mean score of  $m=2.646$  OTR per minute. Calculating the TauU score for Ms. Sully produced a score of 1, with a 90% CI of [0.580, 1]. This TauU score represents a very large change between Ms. Sully's baseline data and her intervention data. However, two data points are overlapping between the baseline and intervention phases.

Ms. Reveille's implementation of OTR immediately increased upon entering the intervention phase, showing a slight upward trend with some variability. The mean score of the baseline values was  $m=1.255$  and the mean score of the intervention values was  $m=1.996$ . Scores ranged from 0.04 to 2.133 OTR per minute in the baseline phase. During the intervention phase, Ms. Reveille's scores ranged from 1.133 to 3.133 OTR per minute. TauU score of 0.7464, 90% CI [0.366, 1], represented a large change in Ms. Reveille's implementation of OTR from baseline to intervention.

Finally, Mr. Gill's implementation of OTR increased immediately after entering the intervention phase, showing a drastic increase in implementing the OTR intervention with some variability in both phases. Baseline data ranged from 0.533 to 2.133 with a mean score of  $m=1.44$  OTR per minute. During the intervention phase, Mr. Gill's scores ranged from 2.533 to 3.533 with a mean score of  $m=3.067$ . Calculating for TauU, produced a score of 0.84, 90% CI [0.337, 1], represented a very large effect size. The visual analysis also shows zero overlapping points between baseline scores and intervention scores.

**Figure 4.4 Effects of PF on Teachers' Use of OTR**



**Classroom on-task behavior.** Classroom on-task behavior was computed for each session during the baseline and intervention phase across all participants. The composite scores showed the percentage of students that were on task during the 15-minute observation periods. This distal measure was used to look at the effects of correctly implementing OTR on the classroom's on-task behavior. Across all participants, an average of 65.22% of the classroom was on-task during the baseline phase. Classroom on-task behavior ranged from 0-100% during the baseline phase. During the intervention phase, an average of 79.729% of the classroom was on-task across all four participants. For the intervention phase, classroom on-task behavior ranged from 33-100%, with all classrooms experiencing 100% of on-task behavior at least once during the intervention.

**Qualitative analysis of social validity.** A six-question survey was administered to all four participants after the intervention condition. While some of the participants stated that they have received coaching or observational feedback in the past, they stated that their previous experiences with coaching were mostly at face value from a person within the district once or twice a year. Responses from the survey expressed several positives from this coaching experience that differed from their experiences in the past. For example, one participant stated that the coaching "was more of a let me see what can be improved instead of what needs changing." The participants expressed that the coaching experience helped them with accountability in preparing for the lesson and continuing to improve on skills that were working. None of the responses expressed negative aspects of the coaching and feedback experience.

The survey also attempted to gain insight on how delivering consistent PF impacted their teaching and if the participants believed there was a return on investment for participating in coaching. Participants conveyed that through this experience they became more aware of how different teaching strategies impacted student engagement and how to more frequently engage with their students. Another participant articulated that the experience helped add skills and strategies to their teaching repertoire to better support the areas where the students were struggling. Furthermore, the participants stated that immediate feedback helped reflect on their teaching and understanding of how their teaching impacted student engagement.

## **Discussion**

This study was a systematic replication of the Duchaine, Jolivette, and Fredrick (2011) by using the same population and PF on OTR. However, further investigation was conducted to use OTR as the primary dependent variable. After a 15-minute observation period, the researcher conducted PF sessions with the SOAR teacher in effect to increase their rate of OTR delivered. This study sought to answer two research questions; Does performance feedback increase the teachers' implementation of Opportunities to Respond? What effects do high rates of Opportunities to Respond have on students' on-task behavior?

A functional relationship was established from a visual analysis of the data between PF and the teachers' increased rates of OTR. Yet the effect sizes, using TauU showed great variability ranging from 0.4 to 1. Mr. Gill was the only teacher that received consistently higher scores during intervention than seen during the baseline

phase. The other three teachers did show improvement in the use of the strategy, but also received scores in their intervention phases that were equal or below scored in their baseline phase. Using visual analysis, an upward trend can be seen in each teachers' data.

The varying success of this intervention may be due to the lack of consistency of instruction and the varying number of students in the classroom. Some days the teachers' ability to implement OTR was greatly affected by the number of students in the classroom and the behavior that the students were performing. For example, there were times when a teacher would present an OTR, give adequate wait time, yet receive no response. After several attempts to elicit a response from the student the teacher would switch from instruction to more of attending and compliance procedures that were outlined in the student's behavior plan. These moments significantly impacted the teachers use of OTR despite the teacher being in the baseline or intervention phase.

Due to classroom on-task behavior being a distal measure the information provided by this study is inconsistent. One hundred percent classroom participation was seen in baseline phases and intervention phases, which could be seen as OTR having little to no effect on classroom on-task behavior. However, the average percent of on-task behavior did increase in each classroom after the rate of OTR began to improve, with a 14 point increase overall. Furthermore, the number of students in the classroom played a significant role in manipulating this data. During observations, a classroom could have as few as one student or as many as five students. Also, a student becoming unresponsive during the observation (e.g. fell asleep, walked out of class, or becoming

uncooperative) had an impact on the classroom on-task behavior and lowered the on-task percentage, as well as having an inverse effect of the teachers use of OTR.

**Limitations.** One major limitation of this study was that datum wasn't able to be collected during each session. Sometimes the researcher would show up for an observation session and the teacher would present the students with individual computer work, only have a single student in the classroom, or the class was taking a break to refocus the classroom after an altercation. During these sessions, data was not collected but the researcher did stay for the observation period to take anecdotal data about how to improve OTR for future observations. Another limitation occurred when the teacher concluded or paused the instruction before the 15-minute observation session was finished. This resulted in both a lower number of OTRs and usually a lower percent of classroom participation. Even in cases where the teacher would return to instruction the researcher notice a change in focus that did impact the teacher's scores.

**Implications for practice.** The use of PF to increase OTR has promising implications for instructional coaches and administrators. By using PF, a coach can relay information to teachers about their performance and give suggestions to increase their use. School personnel should consider redesigning their PD workshops to incorporate in-classroom observation and feedback sessions. These sessions not only show promising results in terms of increasing the fidelity of implementation of the targeted strategy but also allows for continuous communication between the teacher and the coach. This continuous communication allows troubleshooting and working out kinks that are seen when first implementing a new intervention. Additionally, the communication can build



rapport between the teacher and the coach, whereas the coach is seen as an asset or resource to the teacher and not a person entering the classroom to change what the teacher is already doing. Finally, the observation and PF sessions allow for the coach to get a vivid picture of the classroom so that the coach can select an intervention that was tailored to the classroom's needs instead of a cookie-cutter intervention implemented school-wide.

**Implication for research.** Results from this study leads to three avenues for future investigation. First, researchers should replicate this study using different targeted interventions to see if PF is still a viable strategy to increase the fidelity of implementation. Also, researchers should investigate if PF is effective in changing teachers' behavior in larger class sizes (i.e. general education settings). Secondly, researchers should continue to investigate the impact OTR has on on-task behavior and how different types of OTR play a factor for on-task behavior. Finally, researchers should manipulate the dosage and immediacy of the PF sessions. This would be a pivotal factor to understand for cost purposes to administrators and superintendents. Understanding the threshold of how often PF needs to be given to increase implementation, as well as how quickly after the observation session PF should be delivered is an important part of refining and perfecting the use of PF in educational settings.

## CONCLUSION

This dissertation was conducted to deepen the understanding of performance feedback (PF) and evaluate the research available on the subject. Research describes an observation and feedback cycle as a commonality in coaching research (Blazar & Kraft, 2015; Denton & Hasbrouck, 2009; and Joyce & Showers, 1981). Thus, to understand how coaching can impact and redesign PD, an understanding of PF is crucial. Similar to coaching in general, there is no concision in the field defining what PF is and how it should be used in educational settings. However, for this dissertation, PF was defined as a personal interaction with the oral, written, or gestural communication regarding the progress towards the desired outcome. To use PF, first, the teacher must be observed conducting the targeted intervention, then one with knowledge of how to implement the targeted intervention provides the teacher with feedback on their performance. Feedback delivery can vary based on the need of the teacher and access to the observer. PF can be delivered immediately after the observation sessions were conducted or at a delayed time in the future (Burns, Peters, & Noell, 2008; Noell, Witt, Gilberstson, Ranier, & Feeland, 1997; and Sanetti, Luiselli, & Handler, 2007). Additionally, the amount of PF sessions can vary from a daily dosage to about once every two weeks (Coddling, Feinberg, Dunn, & Pace, 2005). The research on PF shows support for both general education instructors as well as special education instructors (Maggin, Fallon, Sanetti, & Ruberto, 2012; McKenney & Bristol, 2015; and Zoder, et al., 2013). Furthermore, researchers have used PF to increase teachers' implementation of various strategies/intervention including both

academic (Maher, 1982; and Zoder, et al., 2013) and behavioral interventions (Leach & Conto, 1999; and Noell, Duhon, Gatti, & Connell, 2002).

This understanding of PF was investigated by conducting a systematic literature review and evaluation of quality using WWC standards for single-case research to answer the research questions: 1) Is PF supported by the literature as an EBP according to the WWC single-case design standards? Also, the literature review allowed for a deeper understanding of how PF has been used to increase teachers' fidelity of implementation within educational settings. Secondly, a meta-analysis using NAP and Tau effect sizes were conducted to answer the research questions 2) What are the effects of PF on the fidelity of implementation of classroom strategies and programs? and 3) What moderator variables of PF produces the best effect for implementation fidelity? Finally, the author conducted a multiple baseline experiment with four behavior support (SOAR) teachers to answer 4) Does PF increase the teachers' implementation of Opportunities to Respond? and 5) What effects do high rates of Opportunities to Respond have on students' on-task behavior?

### **Understanding of Results**

**Literature review.** A systematic literature review identified 15 studies that met the inclusion criteria. While each study supported the success of using PF to increase the fidelity of implementation within the participants across various conditions, some conditions were used more than others. Across the studies, PF was used to implement behavioral strategies such as increasing OTR or implementing self-monitoring (Sanetti, Fallon, & Collier-Meek, 2013) more than academic and curriculum strategies like

DIBELS (Zoder, et al., 2013). Researchers of these studies tended to employ general education instructors as participants, with fewer studies using special education instructors (Coddling, Feinberg, Dunn, & Pace, 2005) or school-based teams (Burns, Peters, & Noell, 2008; and Sanetti, Luiselli, & Handler, 2007). Looking at the PF sessions, researchers typically delivered PF to their participants on a daily schedule with PF sessions happening within 24 hours of the observation sessions.

**Meta-analysis.** The author conducted a meta-analysis on 15 studies that used PF sessions with school personnel to increase implementation fidelity. This meta-analysis included 89 AB contrasts with a mean effect of 0.61 with effect sizes ranging between 0.00 and 1.00. Results from the meta-analysis showed that PF studies produced an omnibus statistically significant mean score of  $\hat{\mu} = 0.6116$ , expressing a large effect (Vannest & Ninci, 2015). To conduct the moderator analysis Tau scores were computed for each of the included studies. All of the scores were 1.00 with confidence intervals between 0.9983, 1.0016. These similarities in effect score show no difference between any of the moderating codes. However, some conclusions can still be made. For instance, moderator analysis of the type of participants indicated that PF can be effective no matter if used with general education instructors or special education instructors. The results indicate no difference in effect despite the change in the immediacy of PF or the dosage of PF. These findings suggest that PF is an effective intervention that can be used with different types of educators while manipulating how often PF is delivered and when it is delivered without significantly lowering the effectiveness of the PF.

**Single-case experiment.** A multiple baseline across four participants was conducted to understand the effect of PF on Opportunities to Respond (OTR). Using visual analysis all four SOAR teachers showed a functional relationship, with improving their frequency of implementing OTR throughout the intervention. TauU effect sizes ranged from 0.4 to 1.0 expressing a large range from small to very large. In terms of understanding how OTR impacts classroom on-task behavior, this study's results are inconsistent. Results ranged from 0-100 percent in baseline and intervention phases across all participants. The average percent of on-task behavior did increase from 62.22% during baseline phases to 79.73% in the intervention. These results suggest that PF is an effective intervention to increase behavior support teachers' use of OTR.

### **Limitations**

Three limitations should be identified throughout this journey of understanding. First, with any systematic literature review, there is a possibility that additional articles that were not identified by the author may meet the inclusion criteria for this dissertation. Another limitation involves the isolation of PF sessions. Because of the importance to understand the effects of PF in isolation, many studies used PF in combination with other coaching components. While these excluded studies could offer insight into how PF is being used in an educational setting, these studies could not confidently express the effects of implementing PF, thus were excluded from the research. Yet it is important to note the significance of using PF in concert with other coaching components to create well design PD opportunities for school personnel. However, that was not the scope of this dissertation. A final limitation, from the single case experiment, was that data were

not collected during each observation. In some cases, this delayed the delivery of PF and possibly could have impacted the effectiveness of the intervention.

### **Future Research**

Continuing to research the use of PF with educators to increase their fidelity of implementation is crucial for coaching to grow as an alternative to traditional PD workshops. Future researchers can continue to expand the literature by utilizing PF sessions in different settings and populations other than with educators within their classrooms. PF research should be expanded to determine success in settings such as vocational settings and in-home sessions. Participants other than educators that still have an impact on students (i.e. parents, community personnel, and counselors) should be sought out and employed in PF studies to see if these populations can have similar success. Another avenue of investigation is identifying interventions and strategies that can be increased by implementing PF. Building a collection of EBP that can be improved through using this implementation strategy could be the answer to lessening the RTPG described throughout this dissertation. Finally, researchers should start investigating the long-term effects of PF. Understanding the fall off of implementation is an impact aspect of any PD. Thus, identifying how long high fidelity of implementation can be maintained after PF sessions are concluded is important for the field of coaching.

### **Implications**

This dissertation offered several points to why using PF in educational settings is important and how PF can impact our teachers' understanding and ability to implement interventions with high fidelity. Coaches can use this information to construct coaching

experiences that incorporate PF sessions. With the understanding of this literature, coaches can manipulate the PF components for different educators and skills set, as well as varying the immediacy and the dosage. This allows for more coaching and in turn higher chances of student success.

Those wanting to implement PF sessions can also use this information to better their use of the observation and feedback cycle. Understanding how to effectively implement PF with educators is vital to incorporating EBP into the daily teachings of our instructors and thus impacting student success. As this dissertation showed PF is a viable implementation strategy to increase implementation of OTR firsthand, as well as provide an overview of the success of PF within the literature. Teachers can engage in PF to increase their use of an intervention or to build fidelity in current practices. Furthermore, understanding how to modify PF without losing its effectiveness can help other teachers that may need less rigorous and more flexible scheduling engage in PF experiences.

## WORKS CITED

*Citations with an asterisk (\*) were included in the literature review and meta-analysis*

- Act, E. S. S. (2015). Every student succeeds act of 2015. *Publ. L*, 114-95.
- Act, N. C. L. B. (2002). No child left behind act of 2001. *Publ. L*, 107-110
- Act, R. E. (1998). Reading excellence act of 1998. *H.R.*, 2614
- Anderson, R., Feldman, S., & Minstrell, J. (2014). Understanding relationship: Maximizing the effects of science coaching. *Education Policy Analysis Archives/Archivos Analíticos de Políticas Educativas*, 22.
- Artman-Meeker, K., Fetting, A., Barton, E. E., Penney, A., & Zeng, S. (2015). Applying an evidence-based framework to the early childhood coaching literature. *Topics in Early Childhood Special Education*, 35, 183-196.
- Baker, R. G. (1983). The contribution of coaching to transfer of training: An extension study (Doctoral dissertation, University of Oregon).
- Blazar, D., & Kraft, M. A. (2015). Exploring mechanisms of effective teacher coaching: a tale of two cohorts from a randomized experiment. *Educational evaluation and policy analysis*, 37, 542-566.
- Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. John Wiley & Sons
- Bowman-Perrott, L., Burke, M.D., Zaini, S., Zhang, N., & Vannest, K. (2016). Promoting positive behavior using the good behavior game: A meta- analysis of single-case research. *Journal of Positive Behavior Interventions*, 18, 1–11.



- Brock, M. E., & Carter, E. W. (2017). A meta-analysis of educator training to improve implementation of interventions for students with disabilities. *Remedial and Special Education, 38*, 131-144.
- Brock, M. E., Cannella-Malone, H. I., Seaman, R. L., Andzik, N. R., Schaefer, J. M., Page, E. J., ... & Dueker, S. A. (2017). Findings across practitioner training studies in special education: A comprehensive review and meta-analysis. *Exceptional Children, 84*, 7-26.
- Burnham, J. F. (2006). Scopus database: a review. *Biomedical digital libraries, 3*, 1.
- \*Burns, M. K., Peters, R., & Noell, G. H. (2008). Using performance feedback to enhance implementation fidelity of the problem-solving team process. *Journal of School Psychology, 46*, 537-550.
- Cavanaugh, B. (2013). Performance feedback and teachers' use of praise and opportunities to respond: A review of the literature. *Education and Treatment of Children, 111-137*.
- Clearinghouse, W. W. (2014). *Procedures and Standards Handbook. Version 3.0. What Works Clearinghouse*. ERIC Clearinghouse.
- \*Coddling, R. S., Feinberg, A. B., Dunn, E. K., & Pace, G. M. (2005). Effects of immediate performance feedback on implementation of behavior support plans. *Journal of applied behavior analysis, 38*, 205-219.
- Connor C. M. (2017) Commentary on the special issue on instructional coaching models: Common elements of effective coaching models. *Theory Into Practice, 56*, 78-83, DOI: 10.1080/00405841.2016.1274575

- Cook, B. G., & Cook, S. C. (2013). Unraveling evidence-based practices in special education. *The Journal of Special Education, 47*, 71-82.
- Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children, 79*, 135-144.
- Cornett, J., & Knight, J. (2009). Research on coaching. *Coaching: Approaches and perspectives, 192-216*.
- Cooper, J. T., Whitney, T., & Lingo, A. S. (2018). Using immediate feedback to increase opportunities to respond in a general education classroom. *Rural Special Education Quarterly, 37*, 52-60.
- Crawford, A. D., Zucker, T. A., Williams, J. M., Bhavsar, V., & Landry, S. H. (2013). Initial validation of the prekindergarten Classroom Observation Tool and goal setting system for data-based coaching. *School Psychology Quarterly, 28*, 277-300.
- Denton, C. A., & Hasbrouck, J. A. N. (2009). A description of instructional coaching and its relationship to consultation. *Journal of Educational and Psychological Consultation, 19*, 150-175.
- DiGennaro, F. D., Martens, B. K., & Kleinmann, A. E. (2007). A comparison of performance feedback procedures on teachers' treatment implementation integrity and students' inappropriate behavior in special education classrooms. *Journal of applied Behavior analysis, 40*, 447-461.

- \*DiGennaro, F., Martens, B. K., & McIntyre, L. L. (2005). Increasing treatment integrity through negative reinforcement: Effects on teacher and student behavior. *School Psychology Review, 34*, 220-231.
- Duchaine, E. L., Jolivette, K., & Fredrick, L. D. (2011). The effect of teacher coaching with performance feedback on behavior-specific praise in inclusion classrooms. *Education and Treatment of Children, 209-227*.
- Duncan, N. G., Dufrene, B. A., Sterling, H. E., & Tingstrom, D. H. (2013). Promoting teachers' generalization of intervention use through goal setting and performance feedback. *Journal of Behavioral Education, 22*, 325-347.
- Downer, J., Brown, J., Herrera, M. J., Stuhlman, M., Bourassa, K., Gologor, B., & Wong, P. (2013). Coaching quality and teachers' implementation of the 4Rs Social-Emotional and Literacy Curriculum: Testing the link between two levels of intervention fidelity. *Society for Research on Educational Effectiveness*.
- Fallon, L. M., Collier-Meek, M. A., Maggin, D. M., Sanetti, L. M., & Johnson, A. H. (2015). Is performance feedback for educators an evidence-based practice? A systematic review and evaluation based on single-case research. *Exceptional Children, 81*, 227-246.
- Foster, M. (2018). Conducting a systematic literature review [PowerPoint].
- Gage, N. A., Grasley-Boy, N. M., & MacSuga-Gage, A. S. (2018). Professional development to increase teacher behavior-specific praise: A single-case design replication. *Psychology in the Schools, 55*, 264-277.

- Garcia, M. (2015). Performance Feedback to Improve Classroom Management Practices (Doctoral dissertation, UC Riverside).
- Gregory, A., Ruzek, E., Hafen, C. A., Mikami, A. Y., Allen, J. P., & Pianta, R. C. (2017). My teaching partner-secondary: A video-based coaching model. *Theory Into Practice, 56*, 38-45, DOI: 10.1080/00405841.2016.1260402
- Guckert, M., Mastropieri, M. A., & Scruggs, T. E. (2016). Personalizing research: Special educators' awareness of evidence-based practice. *Exceptionality, 24*, 63-78.
- Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods, 4*, 324-341.
- Hershfeldt, P. A., Pell, K., Sechrest, R., Pas, E. T., & Bradshaw, C. P. (2012). Lessons learned coaching teachers in behavior management: The PBIS plus coaching model. *Journal of Educational and Psychological Consultation, 22*, 280-299.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional children, 71*, 165-179.
- Institute of Education Sciences. (2014). *What Works Clearinghouse study review standards*. Retrieved from <https://ies.ed.gov/ncee/wwc/FWW>
- Johnson, S. R., Finlon, K. J., Kobak, R., & Izard, C. E. (2017). Promoting student-teacher interactions: Exploring a peer coaching model for teachers in a preschool setting. *Early childhood education journal, 45*, 461-470

- Johnson, S. R., Pas, E. T., & Bradshaw, C. P. (2016). Understanding and measuring coach–teacher alliance: A glimpse inside the ‘black box’. *Prevention Science, 17*, 439-449.
- Jones, M. L. (2009). A study of novice special educators’ views of evidence-based practices. *Teacher Education and Special Education, 32*, 101–120.  
doi:10.1177/0888406409333777
- Joyce, B., & Showers, B. (1980). Improving inservice training: The messages of research. *Educational leadership, 37*(5), 379-385.
- Joyce, B., & Showers, B. (1982). The coaching of teaching. *Educational leadership, 40*(1), 4-10.
- Joyce, B., & Showers, B. (1981). Transfer of training: The contribution of “coaching”. *Journal of Education, 163*(2), 163-172.
- Katzenbach, J. B., Shuster, B. C., Shafer, B. H., Lloyd, B. P., & Carter, E.W. (2016). Opportunities to respond tip sheet. *Tennessee Behavior Supports Project*.
- Kim, S., Koegel, R. L., & Koegel, L. K. (2017). Training paraprofessionals to target socialization in students with ASD: Fidelity of implementation and social validity. *Journal of Positive Behavior Interventions, 19*, 102-114.
- Kraft, M. A., Blazar, D., & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of educational research, 88*, 547-588.

- \*Leach, D. J., & Conto, H. (1999). The additional effects of process and outcome feedback following brief in-service teacher training. *Educational Psychology, 19*, 441-462.
- Licklider, B. L. (1995). The effects of peer coaching cycles on teacher use of a complex teaching skill and teachers' sense of efficacy. *Journal of Personnel Evaluation in Education, 9*, 55-68.
- \*Maggin, D. M., Fallon, L. M., Sanetti, L. M. H., & Ruberto, L. M. (2012). Training paraeducators to implement a group contingency protocol: Direct and collateral effects. *Behavioral Disorders, 38*, 18-37.
- \*Maher, C. A. (1982). Performance feedback to improve the planning and evaluation of instructional programs. *Journal of Organizational Behavior Management, 3*, 33-40.
- Marsalis, S., & Kelly, J. A. (2008). Building a RefWorks database of faculty publications as a liaison and collection development tool.
- Martens, B. K., Hiralall, A. S., & Bradley, T. A. (1997). A note to teacher: Improving student behavior through goal setting and feedback. *School Psychology Quarterly, 12*, 33-41.
- Massar, M. (2018). Effects of coach-delivered prompting and performance feedback on teacher use of evidence-based classroom management practices and student behavior outcomes.

- \*McKenney, E. L., & Bristol, R. M. (2015). Supporting intensive interventions for students with autism spectrum disorder: Performance feedback and discrete trial teaching. *School Psychology Quarterly*, *30*, 8.
- \*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*, 217-234.
- Noell, G. H., Gansle, K. A., Mevers, J. L., Knox, R. M., Mintz, J. C., & Dahir, A. (2014). Improving treatment plan implementation in schools: A meta-analysis of single subject design studies. *Journal of Behavioral Education*, *23*, 168-191.
- \*Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly*, *12*, 77.
- \*O'Handley, R. D., Dufrene, B. A., & Whipple, H. (2018). Tactile prompting and weekly performance feedback for increasing teachers' behavior-specific praise. *Journal of Behavioral Education*, *27*, 324-342.
- Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy*, *40*, 357-367.
- Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy*, *42*, 284-299.
- Parker, R. I., & Vannest, K. J. (2012). Bottom-up analysis of single-case research designs. *Journal of Behavioral Education*, *21*(3), 254-265.

- Pustejovsky, J. E. & Swan, D. M. (2018). Single-case effect size calculator (Version 0.5) Web application. Retrieved from <https://jepusto.shinyapps.io/SCD-effect-sizes/>
- \*Reinke, W. M., Lewis-Palmer, T., & Merrell, K. (2008). The classroom check-up: A classwide teacher consultation model for increasing praise and decreasing disruptive behavior. *School Psychology Review, 37*, 315-332.
- \*Sanetti, L. M., Fallon, L. M., & Collier-Meek, M. A. (2013). Increasing teacher treatment integrity through performance feedback provided by school personnel. *Psychology in the Schools, 50*(2), 134-150.
- \*Sanetti, L. M., Luiselli, J. K., & Handler, M. W. (2007). Effects of verbal and graphic performance feedback on behavior support plan implementation in a public elementary school. *Behavior Modification, 31*, 454-465.
- Scruggs, T. E., & Mastropieri, M. A. (2001). How to summarize single-participant research: Ideas and applications. *Exceptionality, 9*(4), 227-244.
- Showers, B. (1985). Teachers coaching teachers. *Educational leadership, 42*(7), 43-48.
- Solomon, B. G., Klein, S. A., & Politylo, B. C. (2012). The effect of performance feedback on teachers' treatment integrity: A meta-analysis of the single-case literature. *School Psychology Review, 41*, 160-176.
- Rohatgi, A. (2017). WebPlotDigitizer.
- Vannest, K. J., & Ninci, J. (2015). Evaluating intervention effects in single-case research designs. *Journal of Counseling & Development, 93*, 403-411.
- Wang, S. (2017). "Teacher Centered Coaching": An Instructional Coaching Model. *Mid-Western educational researcher, 29*.



- Wang, Q. (2013). Structure and characteristics of effective coaching practice. *The Coaching Psychologist, 9*, 7-17.
- Weston, M. E., & Bain, A. (2015). Bridging the research-to-practice gap in education: A software-mediated approach for improving classroom instruction. *British Journal of Educational Technology, 46*, 608-618.
- Witt, J. C., & Elliott, S. N. (1985). Acceptability of classroom intervention strategies. In T. R. Kratochwill (Ed.), *Advances in school psychology* (4th ed.) (pp. 251-288). Mahwah, NJ: Erlbaum.
- \*Witt, J. C., Noell, G. H., Lafleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis, 30*, 693-696.
- Whalon, K.J., Conroy, M.A., Martinez, J.R., & Werch, B.L. (2015). School-based peer-related social competence interventions for children with autism spectrum disorder: A meta-analysis and descriptive review of single case research design studies. *Journal of Autism and Developmental Disorders, 45*, 1513–1531
- \*Zoder-Martell, K., Dufrene, B., Sterling, H., Tingstrom, D., Blaze, J., Duncan, N., & Harpole, L. (2013). Effects of verbal and graphed feedback on treatment integrity. *Journal of Applied School Psychology, 29*, 328-349.

**APPENDIX A**  
**SYNTAX FOR R STUDIO**

```
# Load metafor library

library(metafor)

# Import data

library(readxl)

PF_data <- read_excel("~/Desktop/PF_data.xlsx")

View(PF_data)

attach(PF_data)

# ES   = unbiased standardized mean difference

# ES.Var = variance of unbiased standardized mean difference

# Create and print forest plot

# yi   = effect size

# vi   = effect-size variance

# sei  = effect-size standard error

# data = specific data set (i.e., dataframe)

#TAU #TAU #TAU #TAU #TAU #TAU #TAU #TAU #TAU

##### Homogeneity

# Create Random effect model

rma.uni(yi = Tau_Est, sei = Tau_SE)

RE.model <- rma.uni(yi = Tau_Est, sei = Tau_SE)

RE.model
```

```
# Create and print forest plot

forest(RE.model, order = order(Tau_Est))

##### Publication Bias

# Funnel Plot

funnel(RE.model, main = "standard error")

#Egger's Regression

regtest(RE.model)

##### RE Model for Moderators

rma(Tau_Est, Tau_SE, subset = (type == 1))

rma(Tau_Est, Tau_SE, subset = (type == 2))

rma(Tau_Est, Tau_SE, subset = (immediacy == 1))

rma(Tau_Est, Tau_SE, subset = (immediacy == 2))

rma(Tau_Est, Tau_SE, subset = (dosage == 1))

rma(Tau_Est, Tau_SE, subset = (dosage == 2))
```

## APPENDIX B

### ARTICLE CHARACTERISTICS

Article	Type	Number	YOT	Highest Degree Earned	Skill Taught	Immediacy of PF	Dosage of PF
Burns	Teams	3	N/A	N/A	Problem solving checklist	BNO	Weekly
Codding	SPED	5	2.5	BA	Behavior support plans	SD	Less than once a week
DiGennaro	GEN	4	30	BA	Reinforcement-based intervention plan	SD	Daily
Leach	GEN	3	10.5	N/A	Academic engagement	IA	Multiple times a week
Maggin	SPED	3	12	HS	Group contingency	IA	Daily
Maher	GEN	2	N/A	N/A	Instructional planning	N/A	Weekly
McKenney	SPED	9	N/A	N/A	STAR DTT	N/A	N/A
Noell (2002)	GEN	4	25	MA	Behavior management intervention	BNO	Daily
Noell (1997)	GEN	3	more than 1	N/A	Reinforcement intervention	BNO	Daily
O'Handley	GEN	3	8	MA	Behavior specific praise	WOO	Weekly
Reinke	GEN	4	25	N/A	Classroom check up	N/A	Daily
Sanetti (2013)	GEN	5	N/A	N/A	Self-monitoring	WOO	Weekly

Article	Type	Number	YOT	Highest Degree Earned	Skill Taught	Immediacy of PF	Dosage of PF
Sanetti (2007)	Teams	1	25	MA	Behavior support plan	IA	Less than once a week
Witt	GEN	4	N/A	N/A	Academic engagement	IA	Daily
Zoder	GEN & SPED	4	N/A	N/A	DIBELS	N/A	N/A

GEN= general education instructor, SPED= special education instructor, N/A=

information not available, BA= bachelor's degree, HS= high school diploma, MA=

master's degree, BNO= before next observation, SD= same day as observation, IA=

immediately after observation, WOO= Week of observation.