

EXAMINING THE USE OF VIDEO ANALYSIS ON TEACHER  
INSTRUCTION AND TEACHER OUTCOMES:  
A META-ANALYSIS

by

MCKENZIE MELINE

A DISSERTATION

Presented to the Special Education and Clinical Sciences  
and the Graduate School of the University of Oregon  
in partial fulfillment of the requirements  
for the degree of  
Doctor of Philosophy

June 2020

## DISSERTATION APPROVAL PAGE

Student: McKenzie Meline

Title: Examining the Use of Video Analysis on Teacher Instruction and Teacher Outcomes: A Meta-Analysis

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Special Education and Clinical Sciences by:

Dr. Beth Harn	Chairperson and Advisor
Dr. Elisa Jamgochian	Core Member
Dr. Sylvia Linan-Thompson	Core Member
Dr. Kathleen Strickland-Cohen	Core Member
Dr. Audrey Lucero	Institutional Representative

and

Kate Mondloch	Interim Vice Provost and Dean of the Graduate School
---------------	--

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2020

© 2020 McKenzie Meline

## DISSERTATION ABSTRACT

McKenzie Meline

Doctor of Philosophy

Special Education and Clinical Sciences

June 2020

Title: Examining the Use of Video Analysis on Teacher Instruction and Teacher Outcomes: A Meta-Analysis

The purpose of this replicated systematic review (SR) and meta-analysis was to examine the literature base of single-case research design studies using video analysis to determine the intervention's effectiveness on teacher outcomes. Using a primary search and an ancestral, citation, and first author searches, this study evaluated participant, student, and setting characteristics in dissertations and peer-reviewed articles published from 2010-2020. A total of 24 included articles were coded for descriptive analysis and design quality.

For the meta-analysis, a total of 16 articles were reviewed for statistical analysis, in which a between-case standardized mean difference was used to calculate effect sizes. Results indicate praise ( $n = 6$ ) and fidelity of implementation ( $n = 6$ ) had the largest effect size that continue to define video analysis as a promising practice. Recommendations for future practice include continued studies using video analysis with diverse educators, students, and settings that meet design quality standards as well as increasing sample size to prove the generalizability of video analysis. Addressing these recommendations will support video analysis becoming an evidence-based practice (EBP) for educator development.

## CURRICULUM VITAE

NAME OF AUTHOR: McKenzie Meline

### GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene  
California Polytechnic, San Luis Obispo  
California Polytechnic, San Luis Obispo

### DEGREES AWARDED:

Doctor of Philosophy, Special Education, 2020, University of Oregon  
Master of Arts, Education with specialization in Special Education, 2011,  
California Polytechnic, San Luis Obispo  
Bachelor of Science, Liberal Studies, 2010, California Polytechnic, San Luis  
Obispo

### AREAS OF SPECIAL INTEREST:

Teacher Preparation  
English Learners  
Education Policy

### PROFESSIONAL EXPERIENCE:

Graduate Teaching Assistant, University of Oregon, 2019-2020  
Practicum Supervisor, University of Oregon, 2016-2019  
Learning Support Coordinator, GEMS World Academy (Singapore), 2014-2016  
Special Education and Secondary Teacher, GEMS The World Academy  
(Saudi Arabia), 2012-2014  
English Language Development Teacher, Beolgyo School District (South Korea),  
2011-2012

### GRANTS, AWARDS, AND HONORS:

Culbertson Scholarship Fund, General Scholarship, University of Oregon, 2019

PUBLICATIONS:

Harn, B. & Meline, M. (2018) Developing critical thinking and reflection in teachers within teacher preparation in G., Mariano (eds.) *Handbook of Research on Critical Thinking Strategies in Pre-Service Learning Environments*. (pp. 126-145) IGI Global

McCroskey, C., Brafford, T., Reardon, K., Meline, M., & Harn, B. (in press). IDEA: History and legal issues. In Fisher, D. & Jung, L. A. (Eds.). *Encyclopedia of Education*. New York, NY: Routledge.

Thier, M., Martinez, C. R., Jr., Al-Resheed, F., Storie, S., Sasaki, A., Meline, M., Rochelle, J., Witherspoon, L. & Yim-Dockery, H. Cultural adaptation of promising, evidence-based, and best practices: A scoping literature review. *Prevention Science*. 21(1), 53-64. doi:10.1007/s11121-019-01042-0.

## ACKNOWLEDGMENTS

I wish to express my gratitude to my doctoral dissertation committee members, Drs. Elisa Jamgochian, Sylvia Linan-Thompson, Audrey Lucero, and Kathleen Strickland-Cohen with a special thank you to my advisor, Dr. Beth Harn. This journey would not have been possible without your continuous effort and genuine support.

It's been a true privilege embarking on this journey with the love and support of my friends and family. To my family, thank you, mom, dad, Brandon, Dustin, and Lindsay for showing unwavering support on this crazy journey and only providing slight judgment whenever I pulled out my computer to do work during family gatherings. Your cheering was heard in Oregon.

To my friends, Sophia, Laura, Jesse, Fran, Zachary, who took me on adventures when I needed it the most, thank you for reminding me that life is too much fun and too precious to spend behind a computer screen. You have reminded me that I am not the sum of this program.

To my academic family, thank you for your wisdom and advice that has helped guide me through the program--specifically, Dr. Angela Ingram, Dr. Kyle Reardon, Tasia Brafford, Stephanie St. Joseph, Aaron Mowery, and Stacy Arbuckle, I appreciate your tireless hours editing this dissertation and coding of articles. You all have been reliable in more than one way.

And finally, to my dog, Bulka, who has tolerated my absence and kept me company on countless late-night work sessions. You're the best dog! I could continue, but then this section would be longer than my dissertation and I figured this paper was long enough.

This dissertation is dedicated to my family and friends who believed in me even when I didn't see it myself. Thanks for being my guiding light through this journey.



## TABLE OF CONTENTS

Chapter	Page
I. INTRODUCTION.....	1
Research Questions.....	6
Literature Review.....	7
Self-Reflective Practices.....	8
Video Analysis.....	14
Application to the Current Study .....	20
Conclusion .....	21
II. METHODOLOGY.....	22
Data Collection Process .....	24
Eligibility Criteria .....	25
Coding Variables .....	26
Title and Abstract Review .....	35
Full-Text Review .....	35
WWC Pilot Single-case Design Standards Review .....	37
Data Analysis .....	40
III. FINDINGS.....	49
Descriptive Analysis .....	49
Research Question 1 .....	49
Research Question 2 .....	52
Statistical Analysis .....	65
Research Question 3 .....	65
Relation to the Parent Study .....	69

Chapter	Page
Research Question 4 .....	70
Comparison of Study Characteristics .....	71
IV. DISCUSSION.....	74
Research Question 1 .....	74
Study and Participant Characteristics .....	75
Student and Setting Characteristics .....	76
Research Question 2 .....	76
Research Question 3 .....	78
ES by Participant Characteristics.....	78
ES by Type of Dependent Variable .....	79
Research Question 4 .....	82
Limitations .....	84
Implications for Future Practice .....	88
Conclusion .....	88
APPENDICES .....	90
A. Qualtrics Form for Coding Study Characteristics.....	90
B. Qualtrics Form for Coding WWC Design Quality Standards.....	112
REFERENCES CITED.....	123

## LIST OF FIGURES

Figure	Page
1. PRISMA Flowchart .....	34
2. Forest Plot Displaying BC-SMD ES .....	67
3. Forest Plot Display BC-SMD ES Based upon DV .....	68
4. Forest Plot Display BC-SMD ES Based for Participant Characteristics .....	87

## LIST OF TABLES

Table	Page
1. Operational Definition of the Coding Variables.....	30
2. WWC Pilot Single-case Design Standards Coding Variables .....	42
3. IRR Across Phases.....	45
4. Study Characteristics of the Included Articles.....	53
5. Participant Characteristics of the Included Articles.....	55
6. Student and Setting Characteristics of the Included Articles.....	58
7. Educator, Student, and Study Characteristics.....	61
8. WWC Design Quality Standards Results .....	63

## CHAPTER I

### INTRODUCTION

Evidence-based practices (EBPs) are scientific and empirically-based approaches shown to be effective, efficient ways to produce desired outcomes (Odom, 2009; Odom et al., 2016). Across the field of education, EBPs demonstrate effective strategies targeting a variety of student skills (Harn, 2017). EBPs are essential for maximizing instructional time and improving student outcomes for the most at-risk students. The use of EBPs in school settings has been adopted into federal policy and is endorsed by The Every Student Succeeds Act (ESSA; 2015), which served as a continuation of the Individuals with Disabilities Education Act's (IDEA; 2004) requirement of "utilizing research-based interventions, curriculum, and practices" (Section §1465(b)(2)(D)), by mandating that academic and behavioral intervention programs targeting at-risk learners be evidence-based. This requirement aims to ensure that intervention programs for the most at-risk learners yield the same results as those established in the research.

Effective, empirically-based programs are also needed when working to improve teacher outcomes, particularly instructional quality (Darling-Hammond, 2017; Yoon et al, 2007). The professional development and preparation educators receive need to be effective and efficient as well as minimize use of school resources (e.g., time, money, materials, etc.). These trainings should target and aim to improve educator behaviors (e.g., instructional quality, data monitoring, classroom management skills, etc.) and ultimately, student outcomes. Through the identification of best practices for both teachers and students, educational practices become more intentional, deliberate, and

effective, thus increasing the efficacy of instructional practices for our most at-risk learners.

To establish the most effective practices, multiple studies implementing a specific intervention or practice need to be examined for effectiveness by determining the consistent significant, positive outcomes. Clearinghouses, such as What Works Clearinghouse (WWC), National Center on Intensive Intervention (NCII), and Evidence for ESSA, are the mechanisms that typically examine numerous studies using the intervention, and then label the practice as evidence-based or in need of further research or evidence. Clearinghouses identify EBPs by first evaluating the quality of the research design and then identifying whether or not the intervention has positive, significant outcomes. Replication studies and systematic reviews (SRs) are important pieces of the process for classifying a practice as evidence-based or in need of additional research. Therefore, each of these fundamental investigations and their contributions' determination of EBPs are discussed below.

### **Importance of Replicating Studies**

One way to identify an EBP is through the replication of studies (Therrien et al., 2016). A replication study is a “study [that] purposefully replicates, extends, further investigates, or uses as its basis one or more previously conducted studies” (Cook et al., 2016, p. 226). Replication studies validate or refute the positive findings of a previous study (Cook, 2014), and determine if findings are generalizable to other participants and/or settings (Schmidt, 2009).

Studies need to be replicated to determine if the practice is effective and can be classified as an EBP. For single-case studies using a small sample size of a minimum of

three participants, positive results cannot be extended to make larger claims that a practice is effective and generalizable to other participants or settings. Therefore, the studies must be replicated multiple times to affirm the findings of a parent study (i.e., the study being replicated). For an intervention used in a single-case research designs (SCRDs), which is a commonly used methodology for special education research, to be considered an EBP the standards include: (a) a minimum of five studies using SCRDS published in peer-reviewed journals, (b) a demonstration of a functional relation to show the efficacy of the treatment for each of the five studies, (c) variation in a minimum of three research groups or settings, and (d) documentation of an effect for a total of 20 participants across all studies (Horner et al., 2005; Horner & Kratochwill, 2012). Replication studies establish EBPs by using similar procedures that include a different and larger number of participants (Cook et al., 2016).

Unfortunately, replication studies are conducted infrequently because they are not valued as highly as other forms of research (Cook et al., 2016). Very few studies replicate previous studies (Makel & Plucker, 2014; Makel et al., 2012), limiting the ability to confirm if a given practice is truly effective or generalizable to other settings or participants. For example, Cook et al. (2016) investigated the prevalence of replication studies in special education from 2013-2014 by conducting a literature search across six journals. The investigation resulted in a total of 83 reviewed articles, of which only 9% ( $n = 26$ ) were identified as replication studies. Of the 26 replication studies, 15 were single-case designs and 11 were group designs. This finding indicates that there is a dearth of replication studies within the special education literature. Replication studies are a

significant contribution to the field; however, replication studies need to be viewed as a valuable contribution.

When studies are being replicated, there are issues often with the methodological procedures that should be noted when interpreting findings. These issues tend to include (a) author overlap (Therrien et al., 2016) or (b) inconsistencies within the replicated study (Ioannidis, 2005; Ioannidis, 2012; Pashler & Wagenmakers, 2012). Author overlap occurs when the author of the parent study is a research team member of the replication study, which may result in author bias and the conflation of results. When different authors conduct the replication study, the lack of interference from the parent author minimizes bias and objectively confirms or refutes the previous findings of the parent study (Makel & Plucker, 2014; Makel et al., 2012). Therefore, when conducting replication studies, it is recommended that other authors conduct the replication to avoid author overlap and reduce the possibility of bias (Cook et al., 2016). Additionally, many replication studies are not regarded as high quality studies, according to the What Works Clearinghouse standards (Therrien et al., 2016). The purpose of replication studies is to verify or refute the findings from the parent study or generalize the results to different settings or populations. If a replicated study does not adhere to design quality standards showing that the study is methodologically weak, the study makes erroneous assumptions regarding the treatment's effects and generalizability of the practice. This undermines the primary purpose of replication studies as being able to determine if a practice is evidence-based (Therrien et al., 2016). In summary, quality replication studies are essential for the validation of parent studies and should be conducted more often and by different authors to support the identification of EBPs.



## **Importance of Systematic Reviews (SRs)**

Systematic reviews are a fundamental part of the of EBP identification process because they provide a consolidated and synthesized review of the literature (Moher et al., 2007). These reviews provide up-to-date information across studies and research groups to give insight on the commonalities of findings regarding a specific topic within the field. SRs typically serve as a starting point for granting agencies, researchers, and even practitioners who want to examine the most recent innovations and the empirical data supporting these practices (Moher et al., 2007).

When reviewing SRs, Moher et al. (2007) discovered that few of the studies (17.7%) reported being an updated version of a previously conducted SR. This indicates that SRs tend not to build upon the current literature, and further demonstrates the lack of replication present in the field. Researchers, therefore, typically conduct SRs without extending previous research to include additional, current publications, which leaves out recent trends within the field and presents discontinuous information. By replicating SRs, researchers augment current understandings within the field, validate findings, and/or refute false positives (Zwaan, 2018), which in turn helps push the field of special education forward.

To improve the field of teacher development in special education, the present study replicated Dr. Morin's (2017) dissertation entitled *The Use of Video Analysis to Change Special Educators' Instructional Practices: A Single-Case Study and Meta-Analysis*. The parent study was a meta-analysis examining the overall and moderator effect of video analysis (VA). The parent study examined the efficacy of VA based upon educator, student, and setting characteristics. This current study extends Dr. Morin's

research by replicating her SR research methods while incorporating the most recent advancements in statistical analysis for meta-analyses using SCRD methodology.

Therefore, the current study consists of the following research questions:

1) What is the status of the literature base on VA regarding study characteristics (i.e., publication type), participant characteristics (i.e., role, education level, experience level, age), student characteristics (i.e., disability type, student outcomes), and setting characteristics (i.e., grade level, group size, type of instruction, setting)?

2. What is the status of the literature base on the research design quality for the included articles as measured by the What Works Clearinghouse (WWC) design quality standards (i.e., meets, meets with reservations, does not meet)?

3) What is the magnitude of effect of VA interventions on the instructional practices of educators?

4) How has the literature base on VA changed since 2016 as reported by Morin's (2017) systematic review?

To maintain the continuity of the progress of VA within the education field, this meta-analysis extends the search of the parent study to include the most recent literature from 2016 to 2020 as well as utilizes the most recent statistical procedures for calculating the treatment effect in studies using SCRD. The proposed SR aims to corroborate and validate the parent meta-analysis findings by searching overlapping years from 2010-2016. To assist in the fidelity to the methodological process of the parent SR, the researcher contacted Dr. Morin for greater specificity and details regarding the SR procedures. The parent author is not an active participant in the proposed study as to avoid author overlap and prevent bias in the current study's findings.

## Literature Review

Not only are EBPs important for providing quality instruction to at-risk students, but the training tools used to improve the instructional quality of the educators working with the most at-risk students need to be empirically based, as well (Darling-Hammond, 2017; Yoon et al, 2007). The process for training educators must be effective and efficient, while using minimal school resources, adhere to time constraints, and demonstrate instructional growth over time. There are a range of methods to improve instructional practices (e.g., professional development, coaching, consultation, etc.), each of which entails providing performance feedback (PF) to educators.

PF is an EBP shown to improve teaching behaviors and instructional quality (Fallon et al., 2015; Scheeler et al., 2004; Sinclair et al., 2019). PF involves the meeting between a consultant (specialist) and a consultee (teacher) to discuss how to improve instructional practices, such as reviewing student data, examining the fidelity of implementation (FOI), and discussing strategies for improvement (Fallon et al., 2015). The traditional way to implement PF requires the consultant and consultee to meet regularly and conduct follow-up observations, which is not time efficient. This makes it difficult to effectively implement in an authentic classroom setting (e.g., general education, special education, intervention, etc.) due to significant time constraints for teachers and interventionists (Reinke et al., 2007). Not implementing PF properly affects the efficacy of the practice and reduces the significance of instructional outcomes. Therefore, complementary and alternative PF practices that are well-suited for an authentic school setting should be considered.

Self-reflection and VA are two practical practices used to provide PF that addresses the time constraints of the observer or person providing the feedback while still providing necessary feedback to improve instructional quality. Self-reflection and VA are considered promising practices, rather than EBPs, due to the limited number of quality studies demonstrating their effectiveness (Morin, Ganz et al., 2019; Morin, Nagro et al., 2019; Nagro & Cornelius, 2013) and the lack of a consistent definition and clear components necessary for implementation (Beauchamp, 2015; Collin et al., 2013). These two promising practices should be considered when trying to improve the instructional practices of educators in an authentic educational setting because both can be implemented despite the complexities encountered in a classroom. A brief review of the research on each strategy is discussed next.

### **Self-Reflective Practices**

Self-reflection is the careful consideration of one's actions to make decisions that informs future practice (Dewey, 1933). Self-reflection is a process that leads to deeper thinking, analysis, and understanding of one's actions and the impact they have on others. To further understand how to apply self-reflection to a school setting, this section includes a description of self-reflection, its utility and implementation, and the limitations of this practice.

#### ***Description of Self-Reflection***

Within education, self-reflection is comprised of four different hierarchal levels of reflective practices: (a) describing the teaching; (b) analyzing the choices and behaviors; (c) judging the outcomes and the instructional decisions made; and (d) applying the analysis for future practice (Nagro et al., 2017; deBettencourt & Nagro, 2018). Teachers

advance through each of these steps, but to achieve the higher levels of reflective thinking, support needs to be scaffolded. This form of additional support is usually provided by personnel (e.g., a specialist, consultant, coach, supervisor, etc.), who help guide teachers through the reflective process. Teachers are eventually able to effectively and efficiently reflect on their own once they are provided with multiple practice opportunities and feedback. The tools needed to support teachers through this process are discussed further in the following sections.

### ***Utility and Implementation of Self-Reflection***

Reflection is a common practice used in teacher preparation programs for preservice teachers or as a professional development strategy to enhance the skills of inservice teachers and paraprofessionals (Benedict et al., 2016; Harn & Meline, 2019). Within these settings, journaling, lesson studies, case-based instruction, and discussions are common tools for reflection. Journals are written reflections used to capture the teacher's thoughts and feelings about his or her instruction (Etscheidt et al., 2012). Lesson studies, which were originally intended as a professional development tool for inservice teachers, are now being used in teacher preparation programs. The lesson study process consists of five steps: "(a) preparation, (b) collaborative planning, (c) teaching the lesson, (d) observation and data collection, and (e) debriefing and data analysis" (Roberts et al., 2018, p. 238). The teachers work together to determine the lesson objective, identify student goals, and create a lesson. This lesson is then taught to the students while the rest of the team observes, takes notes of the interactions, and monitors the achievement of the lesson goals. Afterwards, the team debriefs the lesson and, in some cases, the team may choose to revise and reteach the lesson (Fernandez, 2002).

Case-based instruction uses example narratives (e.g., vignettes, protocols, etc.) to create a scenario that focuses on a specific classroom problem that the teacher analyzes and helps to build a connection between theory and practice (Kagan, 1993). Case studies are versatile, reflective tools that can be used in various content areas and with both inservice and preservice teachers to highlight a diverse range of problematic situations that may be encountered in a classroom setting. Discussions, a common component of many reflective practices, invite professionals to come together to talk about a common topic or challenge. Discussions afford an opportunity for teachers with various perspectives to explore and share ideas that can be put into practice (Borko et al., 2008). These various self-reflective activities demonstrate the versatility, utility, and feasibility of self-reflection that benefits all educators.

### ***Limitations of Self-Reflection***

Although reflection minimizes the time required of an observer or supervisor, some of these self-reflective practices are time consuming for the educators. For example, lesson studies take three to four weeks to complete the entire process, which typically consist of 10-15 hours spent on team meetings. This is a huge time commitment for teachers. They do not have the excess time necessary for the proper implementation of lesson studies, thereby limiting the utility of this practice (Fernandez, 2002). Sims and Walsh (2009) implemented lesson studies across two years as part of a teacher preparation program for early intervention preservice teachers. The preservice teachers were required to analyze lessons, participate in classroom discussions, and think critically about their research lesson. For the program, the primary goal was for preservice teachers to focus on the jointly designed lesson, which is commonly referred to as a research

lesson, and not critique the teacher's instructional behaviors. The preservice teachers met as a group to collaboratively design a research lesson in which one preservice teacher delivered to his/her class. The group then reconvened to review how the lesson went and designed a revised lesson that addressed the challenges encountered in the first lesson delivery. The other preservice teachers in the group taught the lesson to their class. The design of this lesson study proved to be challenging because of the lack of classroom time needed to develop a completed research lesson. As a result, preservice teachers were required to finish the research lesson outside of university class time and it was no longer a collaborative process. This caused the group conversations to revolve around the teacher's delivery of the lesson and the presentation of the adapted research lesson rather than the examination of the lesson objectives and student interactions. Furthermore, the probing questions asked in the discussion groups were broad and the preservice teachers' responses were regarded as superficial, demonstrating that additional guidance is needed to properly self-reflect (Sims & Walsh, 2009). When considering the implementation of reflection, it is important to select the reflective practice that aligns with the needs of the educators and fits the school context, taking into account resource and time constraints. Even if reflective practices are practical for the school context, a teacher's reflective capabilities need to be considered because this can impact and consume school resources.

Reflection is not an inherent trait of educators and studies show educators experience difficulties when self-reflecting on their own (Sims & Walsh, 2009; Tracz et al., 2005, van Es & Sherin, 2002). Without providing direct instruction on how to reflect, reflective activities (e.g., journaling, lesson studies, case-based instruction, discussions, etc.) are ineffective and requires frequent opportunities to practice ( Nagro &

deBettencourt, 2018). Therefore, teachers need additional support and multiple opportunities to practice and learn how to reflect better. This takes time and additional resources upfront until the teacher is able to reflect independently. For example, Spalding and Wilson (2002), as part of the teacher preparation program described previously, had preservice teachers submit journal entries reflecting on their experiences in their practicum sites. The preservice teachers' initial journal entries focused on a descriptive analysis of what happened in the classroom, which is the rudimentary level of reflection (Nagro et al., 2017; deBettencourt & Nagro, 2018). The preservice teachers had to learn how to progress from descriptive reflective practices to higher level critical thinking that extended beyond stating what one is doing and instead involved a reflection of one's actions. To achieve this, the researchers explicitly taught the preservice teachers how to self-reflect by having them review the narratives of other teachers and identify reflective components. The preservice teachers were then able to transfer these reflective components into their own journal entries. Additionally, the preservice teachers reported that instructor feedback helped develop their reflective practices. When following-up with these teachers in their second year of teaching, they continued to use self-reflection to enhance their instructional practices (Spalding & Wilson, 2002), indicating that self-reflection is practical for teachers and transforms them into lifelong learners (Harn & Meline, 2019; Tripp & Rich, 2012). In order to sustain the efficacy of self-reflection as it is applied to the field to improve instructional quality, reflection needs to be modeled and taught, and requires additional feedback from others in order to achieve higher levels of self-reflection.



Finally, self-reflection is considered a promising practice, not an EBP (Beauchamp, 2015; Collin et al., 2013). The effectiveness of self-reflection needs to be further examined for its impact on student and teacher outcomes. In one instance, Richards et al., (2012) had physical education (PE) preservice teachers write a case study based upon their experiences in the classroom. The case study topics included classroom management, adapted PE, collaboration, ethical decisions, and other common issues experienced by beginning PE teachers. As part of the coursework, the preservice teachers had peers and instructors provide feedback which encouraged the preservice teachers to deepen their reflective practices. In the end, the preservice teachers gained a deeper understanding of how to resolve issues they might encounter in the school setting because they were provided with different perspectives on how to overcome challenges. The process encouraged the preservice teachers to think critically and reflect on the complexities of teaching (Richards et al, 2012). Regretfully, the study neglected to measure the change in instructional skills and its impact on student outcomes. This is a common issue for studies using self-reflection which indicates a greater need for a SR to examine the effectiveness of reflective practices within the literature base. To be considered as an EBP, self-reflection needs a consistent definition (Beauchamp, 2015; Collin et al., 2013) and demonstrates its efficacy by analyzing educator and student outcomes.

In summary, there are various types of reflection (i.e., journaling, lesson studies, case-based instruction, and discussions) that can be implemented in diverse classroom settings, and self-reflection is a promising practice that has demonstrated positive results (Borko et al. 2008; Richards et al., 2012; Spalding & Wilson, 2002). Although teachers

report that they like the practice of self-reflection, it is evident that the ability to effectively self-reflect is not an inherent skill (Tracz et al., 2005; van Es & Sherin, 2002). Studies using self-reflection found preservice teachers provided basic descriptions of their teaching behaviors without using other sophisticated skills to analyze, judge, and apply to their teaching (deBettencourt & Nagro, 2018; Richards, et al., 2012; Spalding & Wilson, 2002; Sims & Walsh, 2009). Educators need to be explicitly taught how to be self-reflective through scaffolded procedures to develop higher-order reflective skills. This reflects the need for specialized personnel (e.g., consultants, coaches, etc.) and additional time for teachers to work continuously with the specialist to promote effective self-reflective practices. However, with the use of a rubric or framework to guide reflection, self-reflection could develop beyond the descriptive level and progress towards higher reflective practices such as judgement of teaching practices and application of changes to future practice (deBettencourt & Nagro, 2018). Finally, to better understand the effectiveness of self-reflection, more research needs to be conducted to demonstrate that teachers who engage in these practices show improvements associated to teacher and student outcomes (Harn & Meline, 2019).

### **Video Analysis (VA)**

Video analysis is a promising practice commonly used for developing teacher skills and reflective practice (Morin, Ganz et al., 2019; Morin, Nagro et al., 2019; Nagro & Cornelius, 2013) by incorporating video technology that gives teachers the ability to review their own instruction or the instruction of their peers. This provides teachers the opportunity to think critically about the interaction between their teaching behaviors and students (Tripp & Rich, 2012). To further understand how to implement VA, this section

provides a description of VA, its utility and implementation, a comparison to self-reflection, and limitations of VA.

### ***Description of VA***

VA consists of three main components: videorecording, video review, and analysis of teaching strategies (Mosley, Wetzel et al., 2017). The teachers record a lesson, watch the video footage, and analyze teaching behaviors through reflection or discussion about what they observed. Because of the use of video, teachers view actions they may have forgotten, are made aware of their behaviors, and/or notice students' responses (Knight et al., 2012; Sherin & van Es, 2005). This gives teachers a more complete and accurate perception of the classroom instruction and student interactions (Nagro & Cornelius, 2013; Rich & Hannafin, 2009). Additionally, teachers can replay the footage and review the event multiple times (Tripp & Rich, 2012) to promote a more objective view of teaching behaviors. These specific teaching behaviors can be referenced and used as examples to validate the feedback being provided. Finally, since VA requires the analysis of teaching behaviors through reflective practices and is used in tandem with reflective thinking, educators need to be supported through the VA process to reach the higher levels of reflection.

### ***Utility and Implementation of VA***

VA is a versatile reflective tool that can be implemented in a variety of ways in terms of (a) the reflection process, (b) content being viewed, and (c) who provides feedback. VA is used simultaneously with self-reflection. Self-reflection can be incorporated in VA through video editing, video annotations, or other self-reflection practices (Osipova et al., 2011). Video editing involves video recording a teacher

providing instruction and editing it to highlight key incidents (Calandra et al., 2008).

Video annotation gives the teachers the ability to self-reflect on their teaching by adding a caption to a video segment (Rich & Hannafin, 2009). This provides video evidence by linking the commentary with the teaching behavior. The third type of VA is video self-reflection, which involves teachers analyzing and making connections about their teaching behaviors while reviewing classroom footage (Nagro & Cornelius, 2013).

Additionally, the educator can view a variety of video footage content. Educators can view: published videos, peer videos, and personal videos (Zhang et al., 2011).

Published videos or video cases are targeted videos that show the teaching environment, student behaviors, instructional content that happen in an authentic classroom setting.

This allows the observer of the publish videos to pinpoint potential areas of concern or demonstrate a variety of teaching behaviors allowing the observer the opportunity to make decisions prior to entering the classroom (Olson et al., 2016; Zhang et al., 2011).

Peer videos are videos of colleagues providing instruction and give the observer the opportunity to learn new teaching techniques (Zhang et al., 2011). Personal videos act as a mirror as the observer views him/herself providing instruction, which gives them the advantage of seeing things that may not have been noticed while teaching the class (Zhang et al., 2011).

Finally, VA can also differ in who observes and provides feedback on the instruction. Feedback can be provided in a group setting (Tripp & Rich, 2012; Hong & Van Riper, 2016; McDuffie et al., 2014), by an expert reviewer (Weber, et al., 2018; Lee & Wu, 2006), or by the individual themselves (Nagro et al., 2017). These variations and

flexibility in the implementation of VA allows for a range of possibilities for how to use this practice and makes it feasible for complex classroom settings.

VA has also shown to be an effective, promising practice for both preservice and inservice teachers in various grade levels and content areas (Morin, Ganz et al., 2019). As part of a yearlong professional development seminar, Sherin and van Es (2005) had mathematics teachers attend monthly meetings where they watched video clips of each other's classroom teaching. The facilitator drove the group dialogue by prompting teachers with open-ended discussion questions. The researchers recorded, transcribed, and analyzed the group sessions. The results indicated the teachers' focus throughout the course shifted from teacher behaviors and pedagogy to student thinking. This growth in reflective skills demonstrates the teachers' ability to improve their reflective practices over time by using video and guided discussion questions.

### ***Comparison to Reflective Practices***

Since VA includes a reflective component, VA is ineffective unless teachers are guided through the process to meet the higher-level self-reflective practices. For instance, van Es and Sherin (2002) had six preservice teachers write essays before and after watching their instruction. The preservice teachers participated in three one-hour long Video Analysis Support Tool (VAST) sessions where the teacher reviewed video from their classroom as well as their peer's classroom. After each video, the preservice teachers were provided with open-ended prompts focusing on student thinking, teacher's roles, and classroom discourse. The preintervention essays were descriptive in content and only discussed what was occurring in the classroom. After the intervention, the preservice teacher's essays included a deeper analysis of the classroom events. Through

scaffolded support and guidance, teachers began using higher level critical thinking when analyzing their own videos. Without prompts, they focused on pedagogy and descriptive comments showing that VA on its own is not an effective practice for teachers. But, with guided VA, teachers can improve their reflective practices and increase the use of instructional strategies (Nagro, et al. 2017).

To support self-reflective practices, VA extends beyond the traditional methods of reflection that rely on memory (Nagro, 2020) by providing a more accurate and objective descriptions of the teacher and student behaviors (Nagro & Cornelius, 2013; Rich & Hannafin, 2009). Robinson and Kelley (2007) compared standalone reflective practices and value-added reflective practices in combination with VA. Preservice teachers practiced role-playing teaching interactions and reflected on their performance. The preservice teachers that engaged in VA acquired higher levels of reflection in comparison to the control group which only reflected on their performance. Tripp and Rich (2012) further confirm VA's contribution to reflection and change in instructional practices. When interviewing teachers who participated in VA and peer feedback, the teachers reported VA (a) provided them an opportunity to view their teaching from an alternative perspective, (b) gave them greater confidence in the feedback provided, (c) inspired teachers to take action and be held accountable for making a change in their instructional behavior, (d) made them more inclined to implement the recommended changes, and (e) teachers perceived that they had improved their instructional skills. Overall, VA complements reflective practice by providing an accurate portrayal of the instruction that leads to improved instructional growth and teacher development.

### ***Limitations of VA***

Some of the major limitations of VA are the amount of time required for teachers participating in VA, the scaffolding of activities for novice users of VA, concerns about the generalizability of VA, and its relation to student outcomes. As previously discussed, VA is a versatile practice that can vary in terms of self-reflective components, video content, and who provides feedback which contributes to its use in complex classroom settings. However, educators need to learn how to properly conduct VA. To effectively observe their instruction, novice users of VA need guidance on what to observe to help them properly reflect on their instruction. Hong and Van Riper (2016) found that using instructional videos where the teacher viewed new instructional practices modeled in an authentic classroom setting was a useful professional development tool for inservice teachers and paraprofessionals. This process helped teachers in reviewing the videos critically, and the teachers discovered new instructional strategies. Regrettably, there was no indication that these skills were applied to their own classroom setting or impacted student outcomes. As a result, studies examining VA as a behavior change strategy should also examine its generalizability and its relation to student outcomes (Morin, Nagro et al., 2019).

### ***Summarizing VA***

VA is an effective behavior change strategy (Sherin & van Es, 2005) and, with recent technological advancements, has become a feasible practice within the classroom (Morin, Ganz et al., 2019; Morin, Nagro et al., 2019, Nagro et al., 2017) and can be applied to a diverse range of instructional behaviors for teachers. Through the use of video technology, the teacher or supervisor reviews the classroom instruction at any time resulting in less time and fewer resources spent on the actual observation. Instead, more

time and resources are allocated to essential components of an observation such as reviewing the video and providing feedback (Weber, et al., 2018).

Given that self-reflection is used simultaneously with VA, VA is ineffective unless interventionists or teachers are guided through the process. Teachers need to be told which instructional items to observe to make the practice more efficacious.

Therefore, studies with guided VA, which use a rubric or observation tool to support reflection, are more effective in improving reflective practices and use of instructional strategies for inservice and preservice teachers (Nagro, et al. 2017).

Although VA is a supported and liked practice by teachers (Tracz et al., 2005), there is a limited amount of studies conducted that empirically support explore the efficacy of the practice (Nagro & Cornelius, 2013) and its impact on student outcomes (Morin, Nagro et al., 2019). There is also a dearth of research measuring VA's impact on teacher effectiveness which prohibits it from being classified as an EBP (Morin, Ganz et al., 2019; Morin, Nagro et al., 2019; Nagro & Cornelius, 2013). Therefore, the current study intends to extend the field of research by examining the impact of VA on teacher instructional quality as measured by teacher outcomes (e.g., behavior specific praise, FOI, opportunities to respond, etc.).

### **Application of VA to the Current Study**

The current study examines studies that have used VA as a form of self-reflection and performance feedback in the hopes of providing evidence of VA as an EBP. This study builds upon Morin's (2017) SR and meta-analysis that examined the effect sizes of SCRD using VA as a treatment as well as examining study characteristics (i.e., participants, students, and setting). The current study extends Morin's (2017) study by



employing the most recent research-validated meta-analysis practices to calculate study effect sizes using between-case standardized mean difference (BC-SMD) in addition to incorporating the most recent literature base from 2016-2020.

Another component of the current study is to examine the relation between VA and teacher outcomes. Most recent VA studies are qualitative in nature and examine the educators' perspective about the feasibility and utility of VA (Hong & Van Riper, 2016; Mosley Wetzel et al., 2017; Trip & Rich, 2012), so the current study examines the impact of VA on various teacher outcomes (e.g., praise, fidelity of implementation, opportunities to respond, etc.). Furthermore, to be considered an EBP, VA needs to be included in more studies that meet quality design standards, so the methodological rigor is also analyzed in the current study.

## **Conclusion**

In conclusion, schools need effective and efficient ways to monitor and promote instructional quality, especially for teachers providing interventions to our most at-risk students. Given that VA is a practice that incorporates self-reflection, the focus of this current study is the use of VA as teacher-preferred practices to improve instructional skills. As an teacher development tool, VA has strengths (e.g., feasibility, less observer time, etc.) and weaknesses (e.g., time allocation, scaffolding of reflective practice, etc.), but VA requires further examination to demonstrate its effectiveness in improving student learning as well as identifying specific methods to make it more readily used in schools. This study examines the literature base involving VA and its impact on teacher outcomes.

## CHAPTER II

### METHODOLOGY

Meta-analyses enhance systematic literature reviews by synthesizing the findings across multiple studies using the same intervention treatment (Borenstein et al., 2009). This provides greater statistical power of the intervention package than the analysis of an individual study. Additionally, it provides more robust information about the generalizability of the intervention across settings, participants, students, and other variables (Tanner-Smith et al., 2016). The purpose of this SR replication and meta-analysis is to gather and systematically review SCRD using VA as a treatment to improve educator teaching behaviors. The studies meeting the inclusionary criteria were descriptively analyzed for study characteristics (i.e., design, participant, student, and setting characteristics) and statistically analyzed for treatment effectiveness.

The present study replicated the SR procedures used in a recent meta-analysis: *The Use of Video Analysis to Change Special Educators' Instructional Practices: A Single-Case Study and Meta-Analysis* (Morin, 2017). Replication studies consist of two different types of replications: direct replications and conceptual replications. A direct replication is the recreation of core components of a parent study (Schmidt, 2009; Zwaan et al., 2018). For the purpose of this study, the SR of the literature base was a direct replication of Morin's (2017) dissertation. The current study adheres to the exact SR methods of the parent study, discussed later.

A conceptual replication deviates from the parent study by intentionally altering a component (Makel & Plucker, 2014; Makel et al., 2012; Schmidt, 2009) and possible effect size (Zwaan et al., 2018). Due to statistical analysis advancements for meta-

analyses that analyze the effect sizes for SCRD, the current study departs from the parent study by calculating the BC-SMD effect sizes rather than the parent study's Tau-U average effect size. A Tau-U is a non-parametric effect size that combines the non-overlapping data points and trend between the baseline and intervention phases to determine the percentage of non-overlapping data between two phases (Parker et al., 2011). Although preferred over other non-parametric measures because it (a) includes the use of all data points, (b) controls for baseline trend, (c) uses simplified calculation procedures that include trend and non-overlap between phases, (d) is sensitive, and (e) has greater statistical power than other single-case analysis methods (Parker et al., 2011), Tau-U is only intended for an individual participant, and it is inappropriate to use a Tau-U estimated effect size when calculating an overall study omnibus effect size due to multiple dependent variables (DVs) being measured within one study. Recent advancements have enabled an overall estimated effect size to be calculated across participants within a study that measures the same DV using a BC-SMD (Hedges et al., 2012; 2013). Therefore, the statistical analysis portion of the dissertation is a conceptual replication of Morin's (2017) study. Instead of a Tau-U average effect size, the present study reports multiple BC-SMD estimated effect sizes per study and is a conceptual replication of the parent study's statistical methods.

This chapter introduces the SR and meta-analysis methodology used in this study to investigate the following research questions:

- 1) What is the status of the literature base on VA regarding study characteristics (i.e., publication type), participant characteristics (i.e., role, education level, experience level, age), student characteristics (i.e., disability type,

student outcomes), and setting characteristics (i.e., grade level, group size, type of instruction, setting)?

2) What is the status of the literature base on the research design quality for the included articles as measured by the What Works Clearinghouse (WWC) design quality standards (i.e., meets, meets with reservations, does not meet)?

3) What is the magnitude of effect of VA interventions on the instructional practices of educators?

4) How has the literature base on VA changed since 2016 as reported by Morin's (2017) systematic review?

This section discusses the (a) data collection process, (b) eligibility criteria, (c) coding variables, and (d) data analysis for the meta-analysis component of the SR. To further authenticate the methods of the replicated SR, the researcher reached out to the primary investigator of the parent meta-analysis for additional support and clarification.

### **Data Collection Process**

With SRs, there is variation in the search and data collection procedures that result in inconsistencies across SR studies which can cause a decrease quality and confidence in the results (Moher et al., 2007). To maintain the study's integrity, the proposed meta-analysis follows the standards and procedures outline by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009). The PRISMA guidelines require a flowchart to report the number of included and excluded at each phase of the study (See Figure 1.1).

The study also replicates Morin's (2017) SR methods with minor adjustments made to the article collection part of the study. These adjustments were made due to the

limited access to resources that the parent study author used. The direct replication of the SR portion of the meta-analysis required that multiple types of searches be conducted to collect relevant articles using VA as a treatment. The researcher conducted two types of searches: (a) a primary search of research databases using predetermined search terms and (b) a forward and backward search that included ancestral, citation, and first author searches to gather any additional documents that were not identified by the research databases in the primary search.

### **Eligibility Criteria**

Once the documents were gathered, they needed to be reviewed to determine if they met the eligibility criteria to be included in the study. In this section, eligibility criteria used to identify articles included in the SR are discussed.

### ***Inclusionary/Exclusionary Criteria***

When conducting an SR, the Cochrane Collaboration, an organization that establishes protocols for conducting SRs and publishes SRs, recommends using the Participants, Intervention, Comparison, Outcomes, and Study design (PICOS) acronym when determining the inclusionary criteria (Methley et al., 2014). The inclusionary criteria used in this study were predefined by the parent study, but also met the PICOS guidelines for selection criteria. To be included in the study, the document needed to (a) use single-case research methodology, (b) have a minimum of one educator (a teacher or preservice teacher) as a participant, (c) take place in an early intervention to grade 12 setting, (d) require the analysis of the preservice teacher's or teacher's video, (e) use an evaluation or feedback component, (f) have comparative data (e.g., pre/posttest, baseline/intervention phases, graphs with data points, etc.), (g) measure a dependent

variable related to teacher outcomes, (h) have behavioral or observable teacher outcomes, (i) be conducted in the United States and in English, and (j) be as a peer-reviewed journal article or dissertation. If some of the criteria were not present in the title and/or abstract, the researcher and coders did not code the criteria to help adhere to objective coding process. For example, if the title and abstract did not specifically state that the study included SCRD methodology, but met the other criteria (e.g., used video analysis, included an educator, had observable teacher outcomes, etc.), the coder was instructed not to code the methodology and proceed to move the article to full-text review to determine the type of methodology used.

Documents were excluded if the study (a) used qualitative or quantitative methods, (b) did not have a minimum of one teacher or preservice teacher as a participant, (c) had professionals working in non-school based facilities (e.g., home, clinical setting, direct care facilities), (d) included videos of other professionals (e.g., exemplar videos of other people), (e) lacked an evaluation or feedback component, (f) had no comparative data, (g) had no dependent variable related to teacher outcomes, (h) included unobservable or non-behavioral outcomes (e.g., surveys, reflections, ability to reflect, content knowledge tests, etc.), (i) were conducted in another language or not in the United States, and (j) were review and/or discussion articles.

### **Coding Variables**

In alignment with the parent study, the selected documents were coded for the following study characteristics: (a) type of study design (e.g., multiple probe, multiple baseline, reversal, AB designs, etc.), (b) publication type (i.e., dissertation or peer-reviewed article), and (c) design quality (i.e., meets WWC design quality standards,

meets WWC design quality standards with reservations, or does not meet WWC design quality standards).

The documents were examined for the following participant variables (a) role (i.e., inservice; preservice; paraprofessional; other; not reported), (b) education level (i.e., high school/general education development diploma; some college, associate's degree, or specialized training; complete Bachelor's degree; Master's degree; not reported), (c) experience level (i.e., 0 years; 1-2 years; 3+ years; not reported), and (d) age (i.e., 18-29 years; 30-39 years; 40-49 years; 50 years and over; not reported). The researcher coded the documents for the following student and setting characteristics: (a) group size (i.e., one-to-one; small group; large group; other; not reported), (b) type of instruction (i.e., academic; communication or language; life skills; other; not reported), (c) grade level (i.e., preschool, elementary; middle school; high school; not reported), (d) setting (i.e., self-contained; inclusion; resource classroom; general education; other; not reported), and (e) disability (i.e., developmental disability; physical disability; mental disability; emotional or behavioral disorders; learning disabilities; cognitive disabilities; other; not reported). Table 1.1 operationally defines each of these variables. All variables were coded across the included studies with the option of "not reported" for each variable.

### ***Primary Search***

The primary search was conducted on May 4-5<sup>th</sup>, 2020 to identify peer-reviewed articles and dissertations completed between 2010 and 2020. The following research databases were systematically searched: (a) ERIC ( $n = 4,106$ ), (b) APA PsycNET ( $n = 1,968$ ), (c) Teacher Reference Center ( $n = 817$ ), and (d) Academic Search Premier ( $n = 2,954$ ). The parent meta-analysis also included Education Source and Education Full

Text; however, due to the researcher not having access to these databases, they were not included in this study. Additionally, the parent study used PsycInfo, PsycArticles, Academic Search Complete, which were substituted for similar databases (i.e., APA PsycNET and Academic Search Premier) that the researcher had access to.

Search terms were inputted into these databases to find articles that would be relevant to the study. For the primary search, the parent study used a combination of terms from three word sets: (a) educator terms (i.e., teacher\*, "teach\* assistant\*", paraprofessional\*, and "instructional assistant\*"), (b) video\*, and (c) components of VA terms (i.e., analy\*, evaluat\*, reflect\*, and feedback\*). In the parent study, a combination of the terms from each set were searched systematically within the database using one search bar. For example, in ERIC, the first search included the terms teacher\* AND video\* AND analy\* in one search bar. The next search in ERIC included teacher\* AND video\* AND evaluat\* in one search bar, and so forth.

Under the guidance of the University of Oregon education librarian, the present study modified the search terms and procedures of the parent study to decrease the amount of irrelevant hits and minimize duplicate articles. For the present study, a Boolean search containing all of the terms from the three sets in individual search bars. The terms for different types of educators (i.e., teacher OR paraeducator OR “teacher assistant” OR “instructional assistant”) were searched in the first search bar. The term “video” was searched in the second search bar. Then, the type of analysis (i.e., analy\* OR evaluat\* OR reflect\* OR feedback) was searched in the third search bar. After the documents were collected from the educational databases, the identified documents were transferred to Zotero, a reference management software, where duplicate files were



removed ( $n = 866$ ). Then, these documents were uploaded to Covidence (covidence.org) an online systematic review management system, where articles were compared across databases and additional duplicate articles were excluded ( $n = 2,339$ ). Within Covidence, the researcher organized and managed the coding procedures between the coding team. A final total of 6,640 articles were identified for the title and abstract review (See Figure 1.1).

### ***Ancestral, Citation, and First Author Search***

An SR is a collection of relevant research that is synthesized and analyzed (Cooper et al., 2019; Levy & Ellis, 2006), and therefore, must include a population of research studies that meet the inclusionary criteria. To do this, the search needs to extend beyond the parameters of a reference database search to guarantee that the SR includes all of the relevant articles possible. Therefore, a backward and a forward search can be utilized for retrieving articles outside of the reference databases. A backward search involves reviewing the published articles that precede the original article. This type of search includes a backward author search, a backward reference search, and previously used keywords search (Levy & Ellis, 2006). A forward search looks for publications proceeding from the original article. This includes a forward reference search in which articles citing the original article are identified or forward author search in which there is a search for articles published by the same author of the original article. Conducting these types of searches expands the search process by identifying articles outside of the reference databases and other electronic sources (Levy & Ellis, 2006).

**Table 1.1.** Operational Definitions of the Coding Variables

Variable	Operational Definition
<i>Role</i>	
Inservice teacher	An inservice teacher is a lead classroom teacher or the primary individual responsible for delivering instruction. An inservice teacher may also be referred to as the lead teacher, special education teacher, general education teacher, credentialed teacher, teacher in-charge, etc.
Paraprofessional	A paraprofessional provides support to students and is supervised by a credentialed or lead teacher. A paraprofessional may also be referred to an aide, educational assistant, instructional aide, 1:1 aide, etc.
Preservice teacher	A preservice teacher is an individual currently enrolled in a teacher preparation program. A preservice teacher may also be referred to as a teacher candidate.
<i>Group Size</i>	
One-to-one	One-to-one group size is the ratio of one student to one educator (i.e., inservice teacher, paraprofessional, or preservice teacher).
Small group	A small group is a subset of students from a larger group who receive instruction. A small group could include centers, reading groups, etc.
Large group	A large group is all students in a classroom who receive instruction at the same time. A large group could include whole group reading instruction, morning circle time, etc.
<i>Type of Instruction</i>	
Academics	Academic skills are tools students need to complete intellectual tasks. Academic skills focus on math, reading, language arts, science, writing, etc. Within each of these categories, there is a subset of skills. For example, reading could include phonics, fluency, reading comprehension, etc.
Communication	Communication skills are tools students need to be able to relay information. Communication skills may include asking questions, making requests, using AAC, responding to questions, etc.
Life skills	Life skills are tools students need to accomplish tasks in their daily lives. Life skills include toileting, cooking, grocery shopping, dressing, eating, hygiene, etc.

**Table 1.1. (continued).**

Variable	Operational Definition
<i>Grade Level</i>	
Preschool	Preschool includes students who are younger than 6 years of age OR are in a grades K and below.
Elementary school	Elementary school includes students who are less than 12 years of age OR are in grades 1-5.
Middle school	Middle school includes students who are less than 14 years of age OR are in grades 6-8.
High school	High school includes students who are 14 years of age OR in grades 9-12.
<i>Setting</i>	
General education	General education is the typical classroom. General education is determined if none of the students in the class had a disability or if there is no mention of students with a disability.
Self-contained	A self-contained classroom is where students with a disability spend all or a majority of their school time. A self-contained classroom includes a special education classroom, separate school, or specialized school for students with disabilities.
Resource	A resource classroom is where students with disabilities spend some of their time in a separate classroom receiving instruction. Students in this setting also spend time in a general education classroom setting.
Inclusion	An inclusion setting is classroom with students with and without disabilities receiving instruction.
<i>Student Disability</i>	
Developmental disability	A development disability is a disability that is present before adulthood. Developmental disabilities include autism spectrum disorder, intellectual disability, Down syndrome, or other developmental disorder.
Physical disability	A physical disability is a condition that impairs mobility. A physical disability may include cerebral palsy.
Mental disability	A mental disability is a condition that affects emotions, thinking, and/or behavior. A mental disability may include anxiety disorder, conduct disorder, bipolar disorder, depression, schizophrenia, and attention deficit hyperactivity disorder

**Table 1.1. (continued).**

Variable	Operational Definition
<i>Student Disability</i>	
Emotional or behavioral disability	An emotional or behavioral disability interferes with a person's ability to sustain relationships and results in frequent use of inappropriate behavior. An emotional or behavioral disability may include oppositional defiant disorder.
Learning disability	A learning disability is a condition that impairs a student from acquiring a skill or knowledge as similar same-aged peers. A learning disability may include dyslexia or a specific learning disability.
Cognitive disability	A cognitive disability impairs mental functioning. A cognitive disability may include a brain injury or cognitive impairment.
Other disabilities	Other disabilities may include multiple disabilities or other health impairments.
Disability not reported	Disabilities not reported may include a developmental delay (e.g., fine motor, literacy, language, cognitive, etc.), general challenging behavior, or no disability identified.

*Note.* Education level, experience level, and age are concrete descriptions and, therefore, are not include in the table.

Following the primary search of the databases, the researcher conducted a backward search which included an ancestral, citation, and first author search to identify any additional documents that may have been omitted. An ancestral search examines the reference lists of the included articles to locate potential articles that may meet the eligibility criteria (Levy & Ellis, 2006). For the present study, the researcher examined the reference list of the included articles ( $n = 1,494$ ). Articles that did not meet the 2010-2020 year and publication type inclusionary criteria were immediately excluded ( $n = 1,313$ ). After duplicates were removed ( $n = 45$ ), 136 articles were included for review (See Figure 1.1).

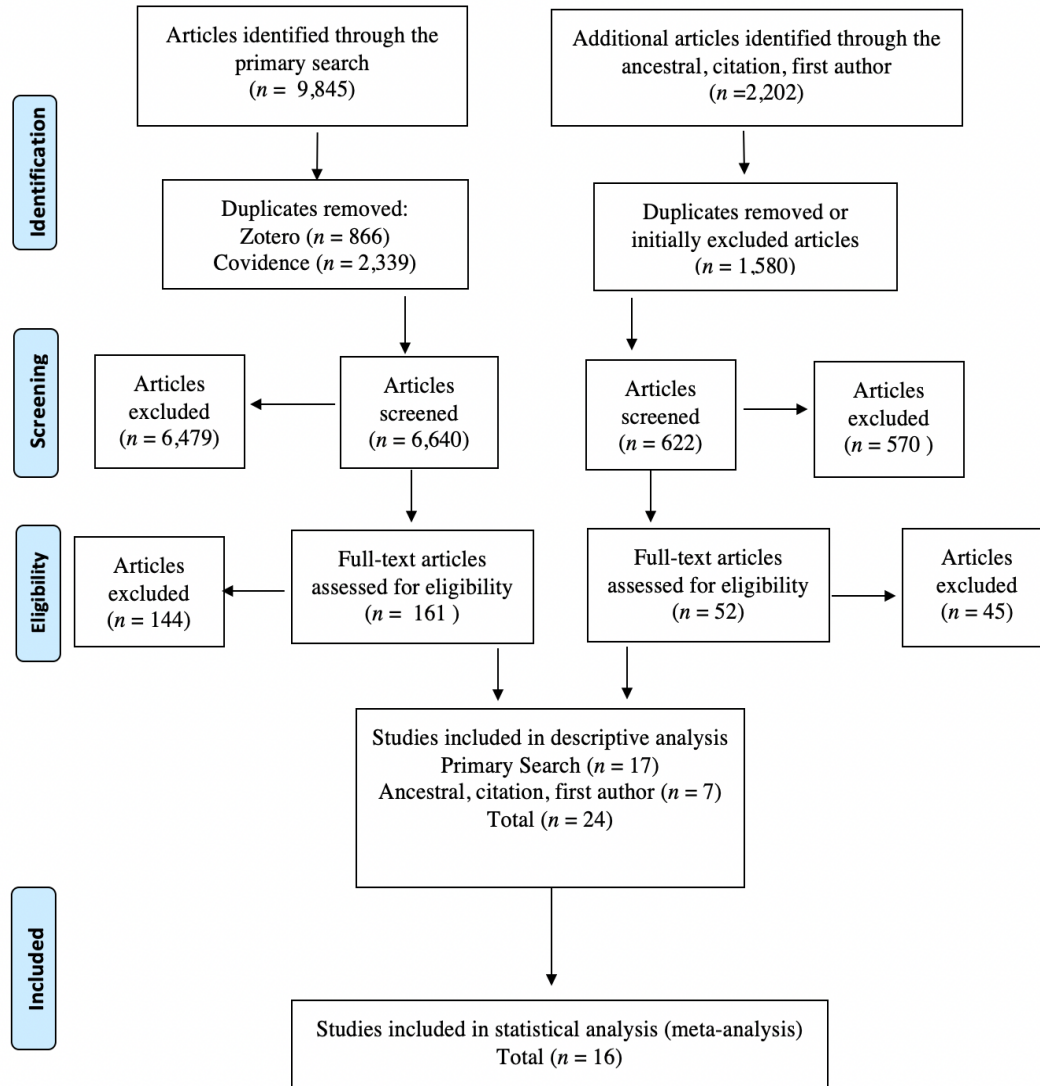
A citation search identifies sources that referenced the original article (Cooper et al., 2019; Levy & Ellis, 2006). Google Scholar was used to find articles that cited the

original included article by using the “cited by” feature ( $n = 544$ ). Articles that did not meet the 2010-2020 year and publication type inclusionary criteria were immediately excluded ( $n = 44$ ). After duplicates were removed ( $n = 120$ ), 380 articles were included for review (See Figure 1.1).

Finally, the researcher completed a first author search (Levy & Ellis, 2006). In the parent study, Morin (2017) used Scopus, an online abstract and citation database, to identify additional articles written by the first author, but due to the researcher’s inability to access this program, these searches were conducted within a similar program called Web of Science, a subscription-based citation database. The researcher conducted a first author search by inputting the author’s first and last name into the Web of Science search bar. When multiple authors with the same first and last name were identified in the search, the researcher used the university affiliation to ensure the correct author was chosen. Then, Web of Science identified articles associated with the author. Through this process, 164 additional articles were collected and identified. Articles that did not meet the 2010-2020 year and publication type inclusionary criteria were immediately excluded ( $n = 40$ ). After duplicates were removed ( $n = 18$ ), 106 articles were included for review (See Figure 1.1).

Across the forward and backward search, any article not within the inclusionary years of 2010-2020 was immediately excluded, as well as any non-peer reviewed or discussion articles (e.g., books, book chapters, review, etc.). After the documents were collected from the research databases, the same procedures used in the primary search were followed. The identified documents were stored in Zotero and then uploaded to Covidence where duplicates were removed ( $n = 183$ ). The researcher and coders then

coded the articles ( $n = 622$ ) during the title and abstract phase and the full-text review phase (See Figure 1.1).



**Figure 1.1.** The PRISMA Flowchart for search results from the SR of studies using VA.

Adapted from Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009).

*Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA*

*Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.*

## **Title and Abstract Review**

In the title and abstract review phase, articles were excluded if they met the exclusionary criteria of: (a) using qualitative or quantitative methods, (b) not having a minimum of one teacher or preservice teacher as a participant, (c) having professionals working in non-school based facilities (e.g., home, clinical setting, direct care facilities), (d) including videos of other professionals (e.g., exemplar videos of other people), (e) lacking an evaluation or feedback component, (f) having no comparative data, (g) having no dependent variable related to teacher outcomes, (h) including unobservable or non-behavioral outcomes (e.g., surveys, reflections, ability to reflect, content knowledge tests, etc.), (i) being conducted in another language or not in the United States, and (j) being a review and/or discussion article.

The researcher reviewed and coded the titles and abstracts of all documents gathered from the primary search ( $n = 6,642$ ) and the ancestral, citation, and first author search ( $n = 622$ ) to determine if it met the inclusionary criteria. Four additional coders coded 20% of the documents. (Coder training and reliability are discussed later).

## **Full-Text Review**

After the documents were identified as not meeting any of the exclusionary criteria in the title and abstract phase, the documents advanced to the full-text review where the articles were evaluated to ensure that they met all of the inclusionary criteria. Each article was examined for the following inclusionary criteria: (a) use of single-case research methodology, (b) a minimum of one participant (i.e., a teacher, paraprofessional, or preservice teacher), (c) a teacher in an early intervention to grade 12 setting, (d) the analysis of the preservice teacher's or teacher's video, (e) an evaluation or

feedback component, (f) comparative data (e.g., pre/posttest, treatment/control groups, graphs with data points, etc.), (g) measurement of a dependent variable related to teacher outcomes, (h) behavioral or observable teacher outcomes, (i) conducted in the United States and in English, and (j) published as a peer-reviewed journal article or dissertation.

Studies that met these criteria were then coded for study characteristics. The researcher and coders used a Qualtrics Survey form to determine participant variables: (a) role (i.e., inservice; preservice; paraprofessional; other; not reported), (b) education level (i.e., high school/general education development diploma; some college, associate's degree, or specialized training; complete Bachelor's degree; Master's degree; not reported), (c) experience level (i.e., 0 years; 1-2 years; 3+ years; not reported), and (d) age (i.e., 18-29 years; 30-39 years; 40-49 years; 50 years and over; not reported). The researcher and coding team also coded the documents for student and setting characteristics: (a) group size (i.e., one-to-one; small group; large group; other; not reported), (b) type of instruction (i.e., academic; communication or language; life skills; other; not reported), (c) grade level (i.e., preschool; elementary; middle school; high school; not reported), (d) setting (i.e., self-contained, inclusion, resource classroom, general education, other, not reported), and (e) disability (i.e., developmental disability; physical disability; mental disability; emotional or behavioral disability; learning disability; cognitive disability; other; not reported).

From the primary search's title and abstract review, 161 articles were included in the full-text review. Through the review process, 144 articles were excluded resulting in 17 articles included in the SR. From the ancestral, citation, and first author search title and abstract review, 52 articles were included in the full-text review. Through the review



process, 45 articles were excluded resulting in seven articles included in the SR. After completing the review process for both the primary search and the ancestral, citation, and first author search, a total of 24 articles were included in the descriptive analysis of the SR. Due to statistical limitations, not all studies were included in the meta-analytic portion of the study ( $n = 8$ ) and the reasons for this are discussed in greater detail later in this chapter.

### **What Works Clearinghouse (WWC) Pilot Single-Case Design Standards Review**

After narrowing the documents to only those that met the inclusionary criteria, each individual study's methods were examined for their adherence to the *What Works Clearinghouse Standards Handbook Version 4.0* (What Works Clearinghouse, 2020, WWC). This information was used to answer research question (RQ) two, which investigated the design of the study. Studies that did not meet the design quality standards were included in the descriptive analysis, but were excluded from the statistical analysis.

Additionally, studies were examined using the WWC pilot single-case design standards, which include the following: (a) manipulation of the variable (Standard 1), (b) inter-assessor agreement (IAA; Standard 2), (c) demonstration of effect (Standard 3), (d) number of data points per phase (Standard 4), and (e) multiple-probe design only standards (Standard 5), along with an Overall Design Rating (See Table 1.2). *Design Standard 1: Manipulation of the variable* is coded as either reporting the manipulation of the independent variable (1) or not reporting the manipulation of the independent variable (0). *Design Standard 2: Inter-assessor agreement (IAA)* consists of three sub standards: IAA reporting (Standard 2A); IAA Frequency (Standard 2B); and IAA Quality (Standard 2C). The IAA is either reported (1) or not reported (2). The IAA frequency is coded as

reporting IAA for a minimum of 20% the sessions within each condition (2), reporting IAA for a minimum of 20% the sessions without disaggregating by treatment or phase (1), or not reporting IAA for a minimum of 20% of the sessions (0). The quality of IAA is coded as the study meeting the minimum of 80% for percent agreement or 60% for Kappa (1) or the study not meeting the minimum of 80% for percent agreement or 60% for Kappa (0). *Design Standard 3: Demonstration of effect* is determined by demonstration of the intervention effect by having three attempts over three points of time (1) or not demonstrating intervention effect by having three attempts over three points of time (0). For alternating treatment designs, the study needs to demonstrate the intervention effect by having three attempts over three points of time with a minimum of two conditions (1) or not demonstrating an intervention effect by having three attempts over three points of time and not including a minimum of two treatment conditions (0). *Design Standard 4: Number of data points per phase* is determined as consisting of a minimum five data points in the baseline and treatment phases (2), a minimum three data points in the baseline and treatment phases (1), and fewer than three data points in the baseline and treatment phases (0). For alternating treatment designs the number of data points per phase is determined by a minimum five data points in the baseline and treatment phases (2), a minimum four data points in the baseline and treatment phases (1), and fewer than four data points in the baseline and treatment phases (0). *Design Standard 5: Multiple-probe designs* consists of three sub standards: Initial baseline (Standard 5A); probe points before the intervention (Standard 5B); and considerations for additional probe points (Standard 5C). The initial baseline is coded as a minimum of three consecutive data points within the first three sessions of baseline for each level (2),

minimum of one data point within the first session of baseline for each level (1), and does not include a minimum of one data point within the first session of baseline for each level (0). Probe points before intervention is coded as a minimum of three consecutive data points within the first three sessions before introducing the intervention for each level (2), minimum of one data point within the first session before introducing the intervention for each level (1), and does not include a minimum of one data point within the first session before introducing the intervention for each level (0). Consideration for other probe points is coded as each unit of analysis (e.g., participant, behavior, etc.) that was still in baseline when intervention is introduced for the previous unit of analysis (e.g., participant, behavior, etc.), had a data point when the previous unit(s) first received the intervention or when the previous unit(s) reached the prespecified intervention criterion (i.e., 3 out of 5 correct before entering intervention), AND this data point is consistent in level and trend with the previous baseline data points in that unit (1); or, each unit of analysis (e.g., participant, behavior, etc.) that was still in baseline when intervention is introduced for the previous unit of analysis (e.g., participant, behavior, etc.), did not have a data point when the previous unit(s) first received the intervention or when the previous unit(s) reached the prespecified intervention criterion (i.e., 3 out of 5 correct before entering intervention), AND this data point is not consistent in level and trend with the previous baseline data points in that unit (0). Finally, the Overall Design Rating is reported as obtaining the highest score possible across all standards (2), receiving a score of 1 on Standards 2 or 4 and not receiving a 0 on any of the Design Standards (1), or receiving a score of 0 on one or more of the Design Standards. These standards are defined in Table 1.2.

## **Data Analysis**

Data analysis consists of inter-rater reliability (IRR), synthesis of the descriptive data, and analysis of the quantitative data. Each of these components are discussed in the following section.

### ***Inter-rater Reliability (IRR)***

IRR was calculated for both the identification of documents as well as coding of included studies. To obtain reliability in the identification phase, the researcher was the primary coder across all phases and four additional coders (three doctoral students and one undergraduate) were used to double code 20% of the documents. Each coder attended a training session where they learned about the eligibility criteria and the coding variables for included documents. During the training session, coders coded a practice article as a group by identifying the inclusionary criteria within the article. Next, the coders coded a second article independently and then the group discussed the discrepancies and resolved any issues. Once a .81 percent agreement (Landis & Koch, 1977) was obtained on two consecutive articles, over 20% ( $n = 1,337$ ) of the primary search articles were double coded. If the two coders disagreed, a third coder reviewed the article to determine if the article would be included in the study. A reliability of .94 percent agreement was achieved across all coders (Table 1.3).

For the title and abstract review of the ancestral, citation, and first author search, a doctoral student who participated in the coding of the primary search articles double coded 20% ( $n = 124$ ) of the articles. Similar to the primary search, if the two coders disagreed, the disagreement was discussed until a consensus was made. A reliability of .96 percent agreement was achieved across coders (See Table 1.3).

Similar IRR training procedures were applied for the full-text review and scoring of WWC single-case quality design standards. For the full-text review of the primary search, the researcher and a doctoral student served as the primary coders ( $n = 81$  and  $80$ ) and another doctoral student coded 20% of the articles ( $n = 32$ ). A reliability of .96 percent agreement was achieved across coders (See Table 1.3).

For the full-text review of the ancestral, citation, and first author search, the researcher and a doctoral student served as the primary coders ( $n = 26$ ) while another doctoral student coded 20% of the articles ( $n = 11$ ). All documents were identified and coded using an online Qualtrics form (Appendix A). A reliability of .99 percent agreement was achieved across coders (See Table 1.3).

For the WWC single-case quality design standards coding, the researcher served as the primary coder ( $n = 24$ ) and a doctoral student coded 20% of the articles ( $n = 5$ ). A Qualtrics form was used when determining the studies' design quality (Appendix B). A reliability of .94 percent agreement was achieved across coders (See Table 1.3).

For all phases, percent agreement was used to determine IRR. Percent agreement is calculated by taking total agreements and dividing by agreements plus disagreements multiplied by 100 (Cooper et al., 2007; Watkins & Pacheco, 2000). Each phase had a minimum of 20% of the articles double-coded. Inter-rater reliability was calculated across all phases with the average percent agreement for the coders ranging from 94% to 99%. The average percent agreement across all phases was 96%. Table 1.3 displays the phase, number of coders, total articles coded, number of articles coded, and average percent agreement for each phase of the study.

**Table 1.2.** What Works Clearinghouse Pilot Single-Case Design Standards Coding Variables. Shown are the scores and criteria definitions for each WWC design standard.

Score	Criteria
<i>Design Standard 1: Manipulation of the Independent Variable</i>	
1	Reports the manipulation of the independent variable
0	Does not report the manipulation of the independent variable
<i>Design Standard 2: Reporting Inter-Assessor Agreement (IAA)</i>	
Reporting IAA (Standard 2A)	
1	Reports IAA
0	Does not report IAA
IAA Frequency (Standard 2B)	
2	A minimum of 20% the sessions within each condition
1	A minimum of 20% the sessions without disaggregating by treatment or phase
0	No reporting of IAA for a minimum of 20% the sessions
IAA Quality (Standard 2C)	
1	Meets the minimum agreement of 80% for percent agreement or 60% for Kappa
0	Does not meet the minimum agreement of 80% for percent agreement or 60% for Kappa
<i>Design Standard 3: Demonstration of Treatment Effects</i>	
1	Intervention effect shown by having three attempts over three points of time.  OR  Intervention effect shown by having three attempts over three points of time with a minimum of two conditions (alternating treatment design)
0	Intervention effect not shown in three attempts over three points of time  OR  Intervention effect not shown in three attempts over three points of time with a minimum of two conditions (alternating treatment design)

**Table 1.2. (continued)**

Score	Criteria
<i>Design Standard 4: Number of Data Points Per Phase</i>	
2	A minimum of five data points in the baseline and treatment phases  OR  A minimum five data points in the baseline and treatment phases (alternating treatment design)
1	A minimum of three data points in the baseline and treatment phases  OR  A minimum four data points in the baseline and treatment phases (alternating treatment design)
0	Less than three data points in the baseline and treatment phases  OR  Less than four data points in the baseline and treatment phases (alternating treatment design)
<i>Design Standard 5: Multiple Probe Designs</i>	
Initial baseline (Standard 5A)	
2	A minimum of three consecutive data points within the first three sessions of baseline for each level
1	A minimum of one data point within the first session of baseline for each level
0	Does not include a minimum of one data point within the first session of baseline for each level
Probe Points Before the Intervention (Standard 5B)	
2	A minimum of three consecutive data points within the first three sessions before introducing the intervention for each level.
1	A minimum of one data point within the first sessions before introducing the intervention for each level.
0	Does not include a minimum of one data point within the first sessions before introducing the intervention for each level

**Table 1.2. (continued).**

Score	Criteria
<i>Consideration of Additional Probe Points (Standard 5C)</i>	
1	Each unit of analysis (e.g., participant, behavior, etc.) that was still in baseline when intervention is introduced for the previous unit of analysis (e.g., participant, behavior, etc.) had a data point when the previous unit(s) first received the intervention or when the previous unit(s) reached the prespecified intervention criterion (i.e., 3 out of 5 correct before entering intervention) AND this data point is consistent in level and trend with the previous baseline data points in that unit.
0	Each unit of analysis (e.g., participant, behavior, etc.) that was still in baseline when intervention is introduced for the previous unit of analysis (e.g., participant, behavior, etc.) did not have a data point when the previous unit(s) first received the intervention or when the previous unit(s) reached the prespecified intervention criterion (i.e., 3 out of 5 correct before entering intervention) AND this data point is not consistent in level and trend with the previous baseline data points in that unit.
<i>Overall Design Quality</i>	
2	The highest score possible across all standards
1	A score of 1 on Standards 2 or 4 and not receiving a 0 on any of the Design Standards
0	Receiving a score of 0 on one or more of the Design Standards



**Table 1.3.** Inter-rater reliability across phases

Phase	Number of coders	Number of articles	Number of articles double-coded	Average percent agreement
<b>Primary search</b>				
Title/abstract	4 double-coders	$n = 6,640$	$n = 1,337$	94%
Full-text review	3; 2 primary coder and 1 double coder	$n = 161$	$n = 32$	96%
<b>Ancestral, citation, and first-author search</b>				
Title/abstract	1 double coder	$n = 622$	$n = 124$	96%
Full-text review	3; 2 primary coder and 1 double coder	$n = 52$	$n = 11$	99%
<b>WWC quality design standards</b>	1 double-coder	$n = 24$	$n = 5$	94%

***Synthesis of the Data***

Data was extracted using the GetData Graph Digitizer (<http://getdata-graph-digitizer.com>), a free, online software that helps retrieve the coordinate points from digital graphs to obtain an estimate of the data. The software requires the researcher to input a JPEG photo of the graphs from the included single-case studies. This obtains an estimate of the data points for the baseline and intervention phases. Generalization and maintenance phases were excluded from the data as these phases do not demonstrate an immediate effect of the intervention, which is the focus of the current study.

Then, for the purpose of this meta-analysis, further analysis of the data was conducted to calculate an effect size for each study's dependent variable(s). The BC-

SMD was calculated to determine the average effect across multiple participants. A standardized mean difference is the “effect size obtained by subtracting the mean outcome of the comparison group from the mean outcome of the treatment group and dividing that difference by an estimate of its standard deviation” (Shadish et al., 2015 p. 101). In the case of SCRd, the comparison is made between the mean intervention and the mean baseline (Shadish et al., 2008). Participant data for each of the included study’s DV was input into Pustejovsky’s (2020) single-case design hierarchical linear model (scdhlm) calculator, which is a free, online R-program web application. This program allows the synthesis of single-case studies by providing a parametric average effect size of data from different cases by calculating the BC-SMD (Shadish et al., 2015).

BC-SMD was chosen to calculate the average effect size because of its ability to account for trend and dependency within an SCRd (Shadish et al., 2015). In the case of meta-analyses, the BC-SMD allows for the statistical analysis of the average effect size of multiple participants within a study. Unlike non-parametric measures such as Tau-U that calculate overlapping data at the individual participant level, BC-SMD allows for average effect size calculation at the study level while still accounting for variability between the cases (Pustejovsky, 2018; Shadish et al., 2015). This allows for an individual study’s results to be compared with a larger body of literature. Having comparable results makes BC-SMD ideal for meta-analyses because multiple studies with different variables can be analyzed and compared (Pustejovsky, 2018; Shadish et al., 2015).

Additionally, BC-SMD calculates a *d* statistic not a *p*-value. A *d* statistic accounts for the variability (e.g., sample size, study design, length of phases, outcome measure scales, etc.) that may impact the magnitude of the effect size (Pustejovsky, 2018). For

instance, within single-case studies, the type of SCRD can vary along with the length of the baseline phase and how the DV is being measured. This impacts the effect size of studies using the same treatment. One fault of the  $d$  statistic is that it overestimates the effect size when there is a small sample size, which is typically the case for SCRD. Pustejovsky's (2020) *scdhl*m calculator automatically corrects for this (Shadish et al., 2015) by reporting a Hedges  $g$  (Hedges, 1981), which does a small sampling bias adjustment and allows for a valid comparison between SCRD studies. Therefore, for the purpose of this research study, a BC-SMD was used to calculate the Hedges'  $g$  average effect size of the studies included in the meta-analysis.

According to Valentine et al. (2016), the BC-SMD does have limitations. First, a functional relation needs to be confirmed by conducting a visual analysis. Then, only multiple-baseline, multiple-probe, and reversal designs can be calculated. Finally, all studies must have a minimum of three participants (Valentine et al., 2016). Therefore, once the SR identifies all included studies, only the studies that meet the BC-SMD requirements are analyzed during the meta-analytic component of the study.

Finally, a randomized effect size accounts for the variability within a study (e.g., sampling error, intervention characteristics, etc.; Borenstein et al., 2009). The participants in the studies are not representative of the population and there is variability in the interventions using VA. To correct for this variability, a randomized effect size was calculated.

Additionally, across the single-case studies, there are various measures for dependent variables using different scales. For example, teacher quality for studies using VA as a treatment used many outcome measures that encompass opportunities to respond

(OTRs; Smith, 2015; Westover, 2010), FOI (Capizzi et al., 2010; Fedders, 2011; Murphy et al., 2015), instructional quality (Coogle, 2019; Knight et al., 2018) and/or praise (Capizzi et al., 2010; Pinter et al., 2015; Smith, 2015; Starling, 2015; Thompson et al., 2012; Westover, 2010) with some studies measuring multiple outcomes while others just measuring one.

Also, the same teacher outcome (e.g., praise, FOI, opportunities to respond, negative response, etc.) were often measured differently across studies. For example, praise was measured as rate of behavior specific praise per minute (Capizzi et al., 2010), frequency of praise per 15 minutes (Pinter et al., 2015), and percent interval of specific praise (Smith, 2015). To account for this, the BC-SMD randomized estimated effect size is interpreted as a small effect (0.2-0.49), medium effect (0.5-0.79), and large effect (equal to or greater than 0.8; Cohen, 1988) using the absolute value of the effect sizes. Negative effect sizes demonstrate that the target behavior decreased after the introduction of the intervention. For example, Hawkins (2011) conducted a study measuring both behavior-specific praise statements (BSPS) and non-behavior-specific praise statements (NBSPS). For the NBSPS, the implementation of VA as an intervention decreased the behavior from the baseline to intervention phase demonstrating a negative effect. The calculation of these effect sizes identifies if there is a functional relation between the intervention and dependent variable. For studies measuring more than one DV, there are multiple effect sizes per study. Therefore, the meta-analysis in the current study includes the statistical analysis of BC-SMD for effect size of individual DVs within the included studies.

## CHAPTER III

### FINDINGS

This study utilized a SR and a meta-analysis to identify the effectiveness of VA within the literature base. The purpose of this review was to gain insights to the different characteristics of studies and to uncover the effectiveness of VA as treatment for educators. This chapter reports the results of the SR and meta-analysis on VA. The specific procedures for searching, identifying, and coding articles, including IRR, are reported in Chapter II. After the articles were identified, the graphical data were extracted from the studies and analyzed using meta-analytic methods, described in the previous chapter. The results of the descriptive analysis (i.e., Research Question 1 and Research Question 2), statistical analysis (i.e., Research Question 3), and relation to the parent study (i.e., Research Question 4) are discussed below.

#### **Descriptive Analysis of Studies Using VA**

The SR of the literature resulted in 24 articles that met the inclusionary criteria as described in Chapter III and the descriptive characteristics of those studies are reported related to Research Questions 1 and 2 of this study.

***Research Question 1 (RQ 1): What is the status of the literature base on VA regarding study characteristics (i.e., publication type), participant characteristics (i.e., role, education level, experience level, age), student characteristics (i.e., disability type, student outcomes), and setting characteristics (i.e., grade level, group size, type of instruction, setting)?***

RQ 1 examines the characteristics, described in Table 1.1, most apparent within the literature of SCRD studies using VA as a treatment. The study, educator, student, and

setting characteristics were coded at the study level (See Tables 1.4, 1.5, and 1.6). For example, Smith (2015) reported that the study took place in resource and self-contained classrooms within elementary ( $n = 3$ ), middle ( $n = 2$ ), and high schools ( $n = 2$ ). Therefore, Smith (2015) was coded as taking place in a resource and self-contained classroom in elementary, middle, and high schools. These coding procedures were consistent across similar articles that reported aggregated descriptive information. The findings for (a) study characteristics, (b) educator characteristics, (c) student characteristics, and (d) setting characteristics are reported below.

**Analysis of Study Characteristics.** The SR included both peer-reviewed articles and dissertations using VA as a treatment. Westover (2010) is a dissertation and Westover and Martin (2014) is a peer-reviewed article gathered in the collection process, but are identical studies using the same data and reporting outcomes; and therefore, is coded as both a dissertation and a peer-reviewed article. Of the included articles, 15 articles (63%) were peer-reviewed and 10 were dissertations (42%). Table 1.4 shows the publication type of the included articles. The design quality of the articles is discussed in the following research question.

**Analysis of Educator Characteristics.** Across all of the articles, the studies included various participant characteristics including role (i.e., inservice, paraprofessional, or preservice), age (i.e., 18-29 years, 30-39 years, 40-49 years, 50+ years, or not reported), educational level (i.e., high school/GED, some college, bachelor's degree, master's degree, or not reported), and teaching experience (i.e., 0 years, 1-2 years, or 3+ years). The findings indicate that a majority of the studies 63% ( $n = 15$ ) reported that the participants were inservice teachers, 50% ( $n = 12$ ) studies reported that

the participants ranged from 18-29 years of age, and 71% ( $n = 17$ ) studies included participants having three or more years of experience. Of the included articles, most of the studies included participants 54% ( $n = 13$ ) that held a bachelor's degree. Table 1.5 displays all educator characteristics across studies and Table 1.7 displays a synthesized version of the study characteristics.

**Analysis of Student Characteristics.** Across all of the articles, the studies included students with various disabilities (i.e., developmental disability, physical disability, mental disability, emotional or behavioral disability, learning disability, cognitive disability, other disability, or disability not reported) in various grade levels (i.e., preschool, elementary school, middle school, high school, post-secondary, or not reported). The findings indicate 50% ( $n = 12$ ) of the studies included participants that worked with students classified as having a developmental disability which includes autism spectrum disorder, intellectual disability, Down syndrome, or other developmental disorders. Additionally, the studies had students in different grade levels. It was reported that forty-six percent ( $n = 11$ ) of the studies took place at the elementary school level and eight (33%) were at the preschool level. Table 1.6 shows the complete list of the disabilities and grade level of the students in the included studies, and Table 1.7 displays a synthesized version of the study characteristics.

**Analysis of Setting Characteristics.** Finally, each article was examined for setting characteristics categorizing group size (i.e., one-to-one, small group, large group, or not reported), type of instruction (i.e., academic, communication, life skills, or not reported), and instructional setting (i.e., general education, self-contained, resource, inclusion, not reported). The studies primarily took place in small groups (46%,  $n = 11$ )

and inclusion classrooms (50%,  $n = 12$ ). In these settings, 54% ( $n = 13$ ) studies focused on academic skill development, as defined in Chapter II. Table 1.6 shows the different types of setting characteristics and the total number of studies taking place in each setting. Table 1.7 displays a synthesized version of the study characteristics.

***Research Question 2 (RQ 2): What is the status of the literature base on the research design quality for the included articles as measured by the What Works Clearinghouse (WWC) design quality standards (i.e., meets, meets with reservations, does not meet)?***

RQ 2 examined the research design quality of the literature base of single-case studies using VA as a treatment. The design quality is measured by evaluating the studies using the criteria included in the WWC design quality standards (See Table 1.2). Studies that met all of the criteria and received an overall study score of two were identified as *Meets Standards* (see description in Chapter II). Studies that met a portion of the criteria and received an overall study score of one were identified as *Meets with Reservations*. Studies that did not meet the standards and received an overall study score of zero were identified as *Does Not Meet*. Of the included articles, 13% ( $n = 3$ ) met the WWC standards, 58% ( $n = 14$ ) met the WWC standards with reservations, and 29% ( $n = 7$ ) did not meet the WWC standards. Table 1.8 shows each study's adherence to the individual WWC single-case design quality standards along with an overall study rating.



**Table 1.4.** Study characteristics of the included articles.

Article	Publication Type	Design	Design Quality	Number of Participants
Alexander et al. (2012)	PR	AB	Doesn't meet standards (0)	2
Bishop et al. (2015)	PR	MPD	Meets standards with reservations (1)	3
Capizzi et al. (2010)	PR	MBD	Meets standards with reservations (1)	3
Coogle et al. (2019)	PR	MPD	Meets standards with reservations (1)	3
D'Agostino et al. (2020)	PR	MPD	Meets standards with reservations (1)	6
Englund (2010)	Diss.	MBD	Meets standards with reservations (1)	6
Fedders (2011)	Diss.	MBD	Meets standards with reservations (1)	3
Hager (2012)	PR	MBD	Doesn't meet standards (0)	1
Hawkins, & Heflin (2011)	PR	MBD	Meets standards with reservations (1)	3
Knight et al. (2018)	PR	MBD	Meets standards (2)	8
Leins Dvorchak (2015)	Diss.	MBD	Meets standards with reservations (1)	5
Lynes (2012)	Diss.	ABCD	Doesn't meet standards (0)	6
MacVittie (2018)	Diss.	ABC	Meets standards (2)	3
McLeod et al. (2019)	PR	MBD	Doesn't meet standards (0)	2

**Table 1.4. (continued).**

Article	Publication Type	Design	Design Quality	Number of Participants
Morin (2017)	Diss.	MBD	Meets standards (2)	5
Murphy et al. (2015)	PR	AB	Doesn't meet standards (0)	2
Pelletier et al. (2010)	PR	MBD	Meets standards with reservations (1)	3
Pinter et al. (2015)	PR	MBD	Meets standards with reservations (1)	4
Robinson (2011)	PR	MBD	Doesn't meet standards (0)	4
Smith (2015)	Diss.	MBD	Meets standards with reservations (1)	6
Snyder (2013)	Diss.	MBD	Doesn't meet standards (0)	4
Starling (2015)	Diss.	MBD	Meets standards with reservations (1)	4
Thompson et al. (2012)	PR	MPD	Meets standards with reservations (1)	3
Westover (2010)	Diss. and PR	MBD	Meets standards with reservations (1)	3

*Note.* Diss. = dissertation; PR = peer-reviewed; MBD = multiple-based line design; MPD = multiple probe design.

**Table 1.5.** Participant characteristics of the included articles.

Article	Participants	Age	Role	Education	Experience
Alexander et al. (2012)	Susan, Rachel	NR	Preservice	NR	0 years, 3+ years
Bishop et al. (2015)	Natalie, Rhonda, Brenda	18-29 years old, 30-39 years old	Inservice	Some college or specialized training, Bachelor's, Master's	1-2 years, 3+ years
Capizzi et al. (2010)	Amy, Sarah, Scott	18-29 years old, 30-39 years old	Preservice	NR	1-2 years, NR
Coogle et al. (2019)	Andreia, Hadi, Abigail	30-39 years old, 50+ years old	Inservice	Bachelor's degree	1-2 years, 3+ years
D'Agostino et al. (2020)	Amy, Betty, Carey, Danielle, Emily, Fae	18-29 years old, 30-39 years old, 40-49 years old, 50+ years old	Inservice	Some college or specialized training, Bachelor's degree, Master's degree	1-2 years, 3+ years
Englund (2010)	Center 1 (PA, PB, PC) Center 2 (PD, PE, PF)	18-29 years old, 30-39 years old, 40-49 years old, 50+ years old	Inservice	High school or GED, Some college or specialized training, Bachelor's degree, Master's degree	1-2 years, 3+ years
Fedders (2011)	Teacher 1-3	18-29 years old	Inservice	NR	0 years, 1-2 years, 3+ years
Hager (2012)	Jennifer	18-29 years old	Preservice	Some college or specialized training	0 years

**Table 1.5. (continued).**

Article	Participants	Age	Role	Education	Experience
Hawkins, & Heflin (2011)	Cantelli, Thomas, Williams	18-29 years old, 30-39 years old	Inservice	Master's degree	1-2 years, 3+ years
Knight et al. (2018)	Teachers 1-8,	NR	Inservice	NR	1-2 years, 3+ years
Leins Dvorchak (2015)	Davis, Kate, Rover, Rita, Moss	NR	Inservice	Bachelor's degree, Master's degree	3+ years
Lynes (2012)	Teachers 1-6	NR	Inservice	Some college or specialized training, Bachelor's degree	1-2 years, 3+ years
MacVittie (2018)	Katie, Cassie, Mary	30-39 years old	Inservice	Bachelor's degree, Master's degree	3+ years, NR
McLeod et al. (2019)	Kelly, Mimi	NR	Preservice	Bachelor's degree	0 years
Morin (2017)	Stephanie, Crystal, Mary Anne, Pamela, Angela	18-29 years old, 30-39 years old	Inservice, paraprofessional	Bachelor's degree	0 years, 3+ years
Murphy et al. (2015)	Hannah, Lydia	18-29 years old	paraprofessional	High school or GED, Some college or specialized training	1-2 years
Pelletier et al. (2010)	Layla, Bob, Sam	NR	Inservice	NR	NR

**Table 1.5. (continued).**

Article	Participants	Age	Role	Education	Experience
Pinter et al. (2015)	Linda, Ava, Leeza, Mick	NR	Inservice	Master's degree	3+ years
Robinson (2011)	Anna, Deborah, Sandra, Mary	18-29 years old, 50+ years old	Paraprofessional	High school or GED, Bachelor's degree	1-2 years, 3+ years
Smith (2015)	Beth, Julia, Kat, Chelsey, Mary, Katie	18-29 years old	Preservice	Some college or specialized training	0 years
Snyder (2013)	Amanda, Leah, Kristin, Tricia	18-29 years old, 40-49 years old	Paraprofessional	High school or GED, Some college or specialized training, Bachelor's degree	1-2 years, 3+ years
Starling (2015)	Participants 1-4	NR	Inservice	NR	NR
Thompson et al. (2012)	Anna, Jane, Gail	40-49 years old, 50+ years old	Inservice	Bachelor's degree, NR	3+ years, NR
Westover (2010)	Dyads A, B, C	40-49 years old, 50+ years old	Paraprofessional	High school or GED, Bachelor's degree	0 years, 3+ years

*Note.* NR = not reported.

**Table 1.6.** Student and setting characteristics of the included articles.

Article	Disability Type	Grade Level	Group Size	Instruction	Setting
Alexander et al. (2012)	NR	Elementary	Small group	Academic skills	Resource classroom
Bishop et al. (2015)	NR	Preschool	One-to-one	NR	Inclusion
Capizzi et al. (2010)	Developmental disability, emotional or behavioral disability, learning disability	Elementary	NR	Academic skills	Resource classroom
Coogle et al. (2019)	Developmental disability	Preschool	One-to-one	Communication skills	Inclusion
D’Agostino et al. (2020)	Developmental disability	Preschool	One-to-one	Communication skills	Inclusion
Englund (2010)	NR	Preschool	NR	Communication skills	Inclusion
Fedders (2011)	Developmental disability	Elementary	One-to-one	Academic skills	Self-contained classroom
Hager (2012)	Cognitive disability	Elementary	Small group	Academic skills	NR
Hawkins, & Heflin (2011)	Mental disability, Emotional or behavioral disorders	High	Small group, large group	Academic skills	Self-contained classroom
Knight et al. (2018)	NR	Middle	NR	NR	NR
Leins Dvorchak (2015)	NR	Middle	Large group	Academic skills	Inclusion
Lynes (2012)	NR	Preschool	Small group	Communication skills	Inclusion

**Table 1.6. (continued).**

Article	Disability Type	Grade Level	Group Size	Instruction	Setting
MacVittie (2018)	Developmental disability, emotional or behavioral disability, learning disability	Elementary	Small group	Academic skills	Inclusion
McLeod et al. (2019)	Developmental disability, physical disability, emotional or behavioral disability	Preschool	Small group	NR	Inclusion
Morin (2017)	Developmental disability, physical disability, mental disability, learning disability, NR	Elementary, post-secondary	One-to-one, small group, large group	Academic skills	Inclusion
Murphy et al. (2015)	Developmental disability, physical disability	Elementary	One-to-one	Communication skills	Inclusion
Pelletier et al. (2010)	Emotional or behavioral disability	NR	One-to-one	Communication skills	NR
Pinter et al. (2015)	Developmental disability, emotional or behavioral disability, learning disability, cognitive disability, other disability	Middle, high	Small group	Academic skills, life skills	Self-contained classroom
Robinson (2011)	Developmental disability	Preschool	One-to-one	Communication skills	Inclusion

**Table 1.6. (continued).**

Article	Disability Type	Grade Level	Group Size	Instruction	Setting
Smith (2015)	Developmental disability, emotional or behavioral disability, learning disability, other disability, NR	Elementary	Small group, large group	Academic skills	Self-contained classroom
Snyder (2013)	NR	Preschool	Small group	Academic skills	Inclusion
Starling (2015)	NR	Elementary	Small group	Academic skills	Self-contained classroom
Thompson et al. (2012)	NR	Elementary	Large group	NR	General education classroom
Westover (2010)	Developmental disability	Elementary	One-to-one	Academic skills	Self-contained classroom

*Note.* NR = not reported.



**Table 1.7.** Educator, student, and setting characteristics across the included articles.

Study Characteristics	Total ( <i>n</i> )
<b>Educator characteristics</b>	
Role	
Inservice	15
Paraprofessional	5
Preservice	5
Age	
18-29 years old	12
30-39 years old	8
40-49 years old	5
50+ years old	6
Not reported	8
Education level	
High school/GED	7
Some college	6
Bachelor's degree	13
Master's degree	7
Not reported	7
Teaching experience	
0 years	7
1-2 years	11
3+ years	17
Not reported	4
<b>Student characteristics</b>	
Student disability	
Developmental disability	12
Physical disability	3
Mental disability	3
Emotional or behavioral disability	7
Learning disability	5
Cognitive disability	2
Other disability	2
Disability not reported	11
Grade Level	
Preschool	8
Elementary school	11
Middle school	3

**Table 1.7. (continued).**

Study Characteristics	Total ( <i>n</i> )
Grade Level	
High school	2
Post-secondary	1
Not reported	1
<b>Setting characteristics</b>	
Group Size	
One-to-one	9
Small group	11
Large group	5
Not reported	3
Type of instruction	
Academic	13
Communication	7
Life skills	1
Not reported	4
Instructional setting	
General education	1
Self-contained	6
Resource	2
Inclusion	12
Not reported	3

*Note.* Participant, student, and setting characteristics are reported at the study level.

**Table 1.8** What Works Clearinghouse design quality standards results for included articles.

Article	Standard 1		Standard 2		Standard 3	Standard 4	Standard 5 (Probe)			Overall Design Quality
Alexander et al. (2012)	1	1	2	1	0	1	N/A	N/A	N/A	0
Bishop et al. (2015)	1	1	1	1	1	2	N/A	N/A	N/A	1
Capizzi et al. (2010)	1	1	2	1	1	1	N/A	N/A	N/A	1
Coogle et al. (2019)	1	1	1	1	N/A	N/A	2	1	1	1
D’Agostino et al. (2020)	1	1	2	1	N/A	N/A	2	1	1	1
Englund (2010)	1	1	1	1	1	1	N/A	N/A	N/A	1
Fedders (2011)	1	1	1	1	1	1	N/A	N/A	N/A	1
Hager (2012)	1	0	0	0	0	1	N/A	N/A	N/A	0
Hawkins & Heflin (2011)	1	1	1	1	1	2	N/A	N/A	N/A	1
Knight et al. (2018)	1	1	2	1	1	2	N/A	N/A	N/A	2
Leins Dvorchak (2015)	1	1	1	1	1	2	N/A	N/A	N/A	1
Lynes (2012)	1	1	1	1	1	0	N/A	N/A	N/A	0
MacVittie (2018)	1	1	2	1	1	2	N/A	N/A	N/A	2
McLeod et al. (2019)	1	1	2	0	1	2	N/A	N/A	N/A	0
Morin (2017)	1	1	2	1	1	2	N/A	N/A	N/A	2

**Table 1.8 (continued).**

Article	Standard 1	Standard 2	Standard 3	Standard 4	Standard 5 (Probe)	Overall Design Quality				
Murphy et al. (2015)	0	0	0	0	0	1	N/A	N/A	N/A	0
Pelletier et al. (2010)	1	1	1	1	1	1	N/A	N/A	N/A	1
Pinter et al. (2015)	1	1	1	1	1	2	N/A	N/A	N/A	1
Robinson (2011)	1	1	0	1	1	N/A	1	1	1	0
Smith (2015)	1	1	1	1	1	1	N/A	N/A	N/A	1
Snyder (2013)	1	1	2	0	1	1	N/A	N/A	N/A	0
Starling (2015)	1	1	2	1	1	1	N/A	N/A	N/A	1
Thompson et al. (2012)	1	1	1	1	1	2	N/A	N/A	N/A	1
Westover (2010)	1	1	1	1	1	2	N/A	N/A	N/A	1

*Note.* Standard 1 includes manipulation of the independent variable. Standard 2 includes reporting on inter assessor agreement (IAA), and frequency and quality of inter-assessor agreement. Standard 3 includes treatment effects. Standard 4 includes points per phase. Standard 5 (probe design only) includes initial baseline points, points before intervention, and additional probe points. N/A = not applicable.

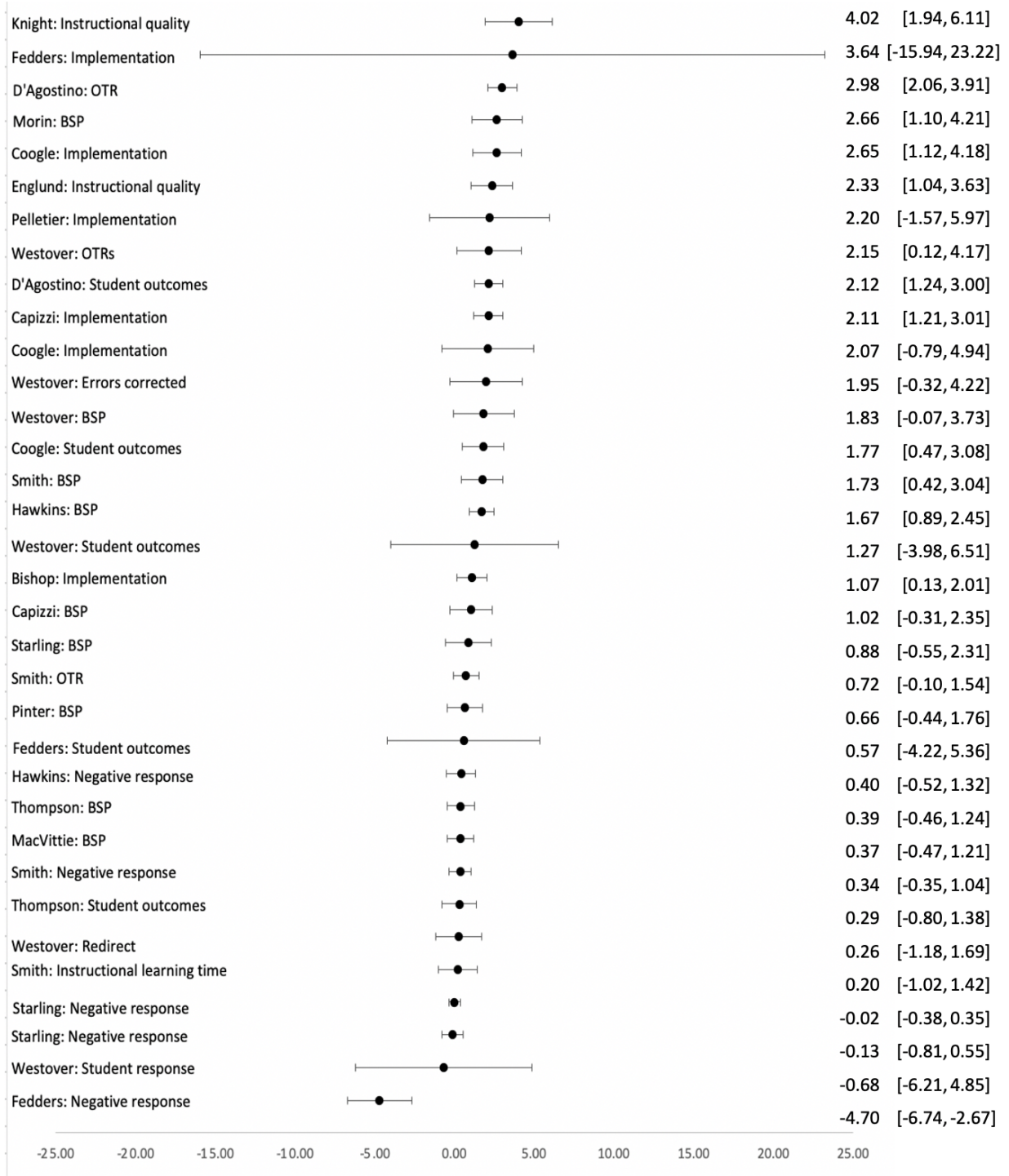
## **Statistical Analysis of Studies Using VA**

For the meta-analysis portion of the study, articles were analyzed for treatment effectiveness by calculating the BC-SMD for participants within a study. To run the statistical analyses, individual participant data was extracted from the graphs within each study using the Getgraph data's software. During this process, one study (Leins Dvorchak, 2015) was excluded from the statistical analysis portion of the meta-analysis because the data used a celeration graph in which the data were unable to be extracted using the Getgraph data's software nor could the data be visually extracted. Additionally, due to the limitations of the BC-SMD calculator, only studies (a) demonstrating a functional relation, (b) using multiple-baseline, multiple-probe, or reversal designs, and (c) have a minimum of three participants were included in the analysis (Pustejovsky et al., 2014; Shadish et al., 2015; Valentine et al., 2016). Each included article's methods were read to determine if there was a functional relation and to identify which SCRD design type (i.e., multiple-baseline, multiple-probe, or reversal designs) was used along with the number of participants. This criterion eliminated the following four studies: Alexander et al. (2012); Hager (2012); McLeod et al. (2019); and Murphy et al. (2015). Additional studies were excluded from analysis because they did not meet the WWC design quality standards (Lynes, 2012; Robinson, 2011; Snyder, 2013). Lynes (2012) did not have the minimum number of data points per phase; and Robinson (2011) and Snyder (2013) did not report IAA resulting in an overall study score of zero. As a result, of the total 24 articles identified, 16 were included in the meta-analysis.

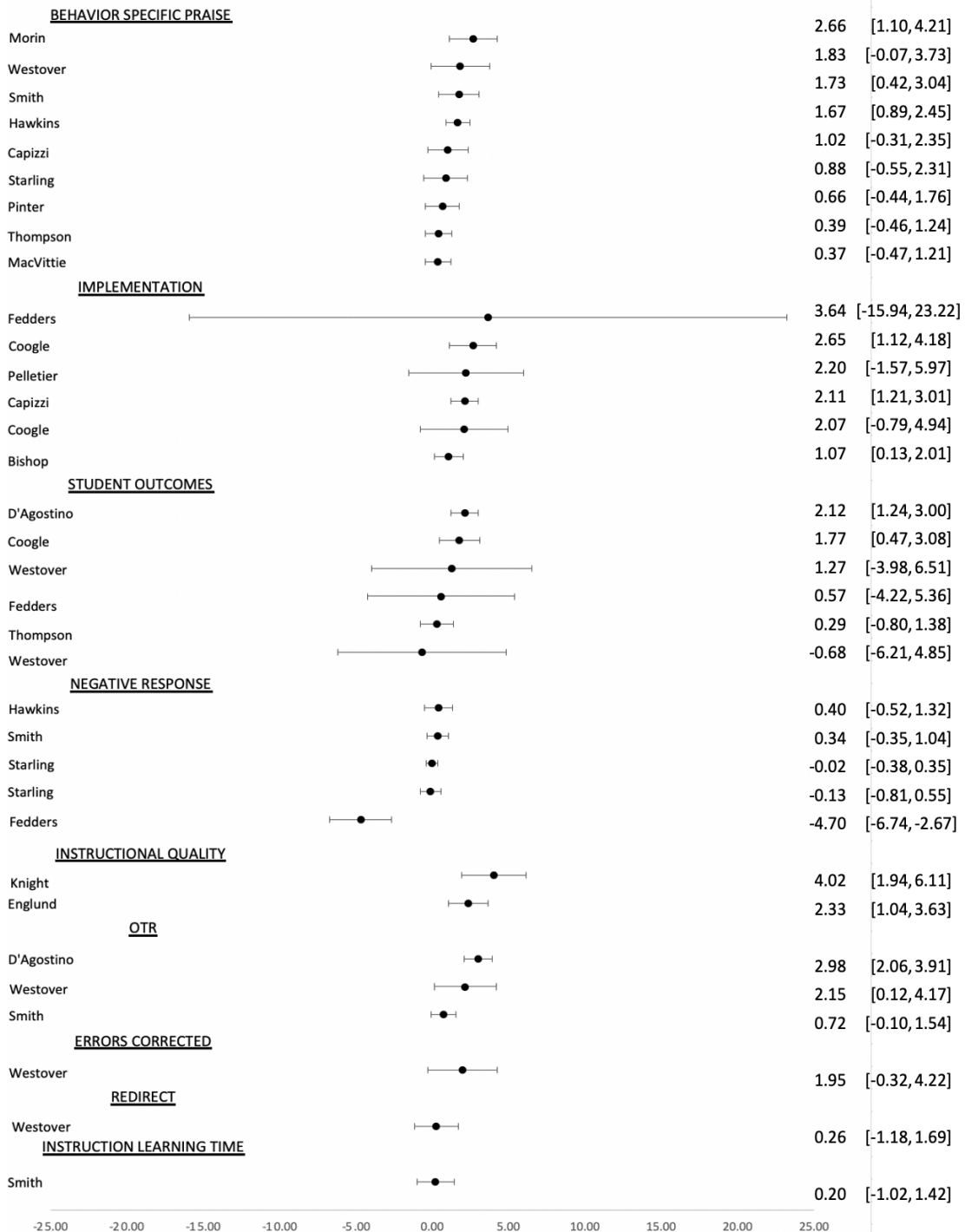
***Research Question 3 (RQ 3): What is the magnitude of effect of VA interventions on the instructional practices of educators?***

RQ 3 analyzes the magnitude of effect of the use of VA as an intervention for educator instructional practices. Figure 1.2 displays a forest plot for each included study. The forest plot shows the effect size (ES) and confidence interval for the individual DVs for each included study (Shadish et al., 2015). The BC-SMD ES across the studies range from -4.70 to 4.02.

**Effect Size by DV.** The included studies measured praise ( $n = 9$ ), implementation ( $n = 6$ ), student outcomes ( $n = 6$ ), negative response ( $n = 5$ ), opportunities to respond (OTR;  $n = 3$ ), instructional quality ( $n = 2$ ), error correction ( $n = 1$ ); redirect ( $n = 1$ ); and instructional time ( $n = 1$ ). Across the DVs, the largest ES were measuring praise ( $n = 6$ ; Capizzi et al., 2010; Hawkins & Heflin, 2011; Morin, 2017; Smith, 2015; Starling, 2015; Westover, 2010), FOI ( $n = 5$ ; Bishop et al., 2015; Capizzi et al., 2010; Coogle et al., 2019; Fedders, 2011; Pelletier et al., 2010), student outcomes ( $n = 3$ ; Coogle et al., 2019; D'Agostino et al., 2020; Westover, 2010), instructional quality ( $n = 2$ , Englund, 2010; Knight et al., 2018), OTR ( $n = 2$ ; D'Agostino et al., 2020, Westover, 2010), and errors ( $n = 1$ ; Westover, 2010). Praise ( $n = 1$ ; Pinter et al., 2015), student outcomes ( $n = 1$ ; Fedders, 2011) and OTR ( $n = 1$ ; Smith,) had a medium effect size (see Figure 1.3). Confidence intervals are reported because they are important when analyzing the ES as it demonstrates precision and the stability of the effect (Borenstein, 1994; Borenstein et al., 2009). Although these ES show a wide range, with a number of them being very large, other studies using BC-SMD also report similar findings (Barton et al., 2017; Maggin et al., 2017). Figure 1.3 displays a forest plot showing the BC-SMD ES for individual studies based upon the DV.



**Figure 1.2.** A forest plot displaying the BC-SMD ES for individual studies and the DVs.



**Figure 1.3.** A forest plot displaying the BC-SMD ES for individual studies based upon the DV.



## **Relation to the Parent Study**

The current study is a direct replication of Morin's (2017) SR and a conceptual replication of her meta-analysis used to examine treatment effects of VA for studies using SCRD methods. The parent study used a Tau-U to calculate the omnibus ES of each study and moderator effects, which Shadish et al. (2008; 2015) states is not recommended practice given multiple DVs. With new statistical developments for analyzing data for meta-analyses using SCRD, the current study calculated BC-SMD ES. Since the ES are not comparable across meta-analyses, the comparison only examined the SR process involving the descriptive characteristics across both studies.

Additionally, with differing accessibility to the research databases and reference software, there were varying results in the included articles. Morin's (2017) SR gathered articles from 1976-2016 with a total of 28 included articles. The current study overlaps and extends Morin's (2017) study conducting a search between 2010-2020. The current study's SR included a total of 24 articles; 13 articles were originally included in Morin's (2017) SR (i.e., Alexander, 2012; Bishop et al., 2015; Capizzi et al., 2010; Englund, 2010; Fedders, 2011; Hager, 2012; Hawkins & Heflin, 2011; Lynes, 2013; Pelletier et al., 2010; Pinter et al., 2015; Robinson, 2011; Snyder, 2013; Westover, 2010) and 11 were newly identified in the current study. Of these 11 articles, six articles were published in the proceeding years of Morin's (2017) SR (i.e., Coogle et al., 2019; D'Agostino et al., 2020; Knight et al., 2018; MacVittie, 2018; McLeod et al., 2019; Morin, 2017) and five were identified within the same search years as Morin's (2017) SR (i.e., Leins Dvorchak, 2015; Murphy et al., 2015; Smith, 2015; Starling, 2015; Thompson et al., 2012). These

were identified due to differences in the access of library research databases and reference databases.

Finally, the current study coded participant, student, and setting characteristics at the study level while Morin (2017) disaggregated the data and examined each characteristic at the participant level. For example, Capizzi et al. (2010) had two participants. Participant 1 was an undergraduate with five years of teaching experience completing her practicum in an elementary resource classroom teaching academics to students with moderate disabilities. Participant 1's group size and specific student disabilities were not reported. Participant 2 was an undergraduate with no teaching experience completing her practicum in a middle school teaching academics in a small group setting. Participant 2's classroom setting and specific student disabilities were not reported. The current study reports characteristics at the study level. The unreported information was coded as not reported (NR). The current study reports overall study data because the BC-SMD ES are reported at the study level and not at the participant level as Morin's (2017) Tau-U ES. As a result, there is not a direct comparison between Morin's (2017) findings and the current study. The following section looks at the comparison of the overlapping years with only the identical articles as well as the extension of the literature database, which included the six articles published between 2016-2020.

***Research Question 4 (RQ 4): How has the literature base on VA changed since 2016 as reported by Morin's (2017) systematic review?***

RQ 4 looks at how the literature base on VA has changed since 2016 as reported by Morin's (2017) SR. The following section compares the (a) study characteristics; (b) design quality standards, and (c) DVs measured. When reviewing these descriptive

results, it should be known that the current study's findings are reported at the study-level and Morin's (2017) findings are reported at the participant level.

### **Comparison of Study Characteristics**

As previously mentioned in RQ 1, a majority (63%;  $n = 15$ ) of the included studies reported that of the participants were inservice teachers; 50% ( $n = 12$ ) of studies reported that the participants ranged from 18-29 years of age; and 67% ( $n = 16$ ) of studies included participants having three or more years of experience. Of these studies, 13 (54%) had participants with a bachelor's degree. In comparison, Morin's (2017) findings were based on the 105 participants within the 28 included articles. Of these participants, 52% ( $n = 55$ ) were inservice teachers; 21% ( $n = 22$ ) of participants were between the age of 18-29 years; and 41% ( $n = 43$ ) of participants with four or more years of teaching. Of the reported educational background, 24% ( $n = 25$ ) of the participants had a bachelor's degree. From 2016-2020, the majority of studies reported that the participants were inservice teachers ( $n = 5$ ), were between 30-39 years of age ( $n = 4$ ), held a bachelor's degree ( $n = 5$ ), and had three or more years of experience ( $n = 5$ ). This indicates that studies continue to include participants who are inservice teachers, hold a bachelor's degree, and have three or more years of experience. The only difference is the increased age of the participants.

In the current study, the majority of the included studies took place in small group settings (46%,  $n = 11$ ) and in inclusion classrooms (50%,  $n = 12$ ). In these settings, 54% ( $n = 13$ ) of the studies had participants providing academic skills development. In comparison, Morin's (2017) findings indicate that 39% ( $n = 41$ ) of the participants provided one-to-one instruction; 32% ( $n = 33$ ) of the participants taught in a small group

setting; 39% ( $n = 41$ ) of the participants focused on academic skills development. A majority of the instruction took place in self-contained (39%,  $n = 41$ ) and inclusion (31%;  $n = 33$ ) classrooms. An extension of the study years indicates that the most recent studies took place in inclusion classrooms ( $n = 5$ ) in a small group ( $n = 3$ ) or one-to-one setting ( $n = 3$ ). In these settings, teachers provided academic instruction ( $n = 2$ ), communication instruction ( $n = 2$ ), or the instruction was not reported ( $n = 2$ ).

Across the current study, students had various disabilities. Twelve (50%) studies included educators who worked with students with a developmental disability (i.e., autism spectrum disorder, intellectual disability, Down syndrome, or other developmental disorders). Eleven (46%) studies reported that the student disability as “Not Reported”, which means that the study did not state the student’s disability or the student had a developmental delay (e.g., fine motor, literacy, language, cognitive, etc.), general challenging behavior, or no disability identified. Similarly, Morin (2017) found that the majority of students (38%,  $n = 15$ ) had developmental disabilities. Of the recently published articles, the most commonly reported disability were developmental disabilities ( $n = 5$ ), which indicates a continued trend of VA being implemented with participants who provide instruction to students with developmental disabilities.

Additionally, the studies included students in different grade levels. Forty-six percent ( $n = 11$ ) of the studies included students in elementary schools and 33% ( $n = 8$ ) studies included students in preschools. In contrast, Morin (2017) found that 34% ( $n = 36$ ) of the participants provided instruction in a preschool setting and 33% ( $n = 35$ ) of participants provided instruction in an elementary school setting. These findings are similar to the current study. In the years since Morin’s (2017) review, the newly

identified studies took place in preschools ( $n = 3$ ) and elementary schools ( $n = 2$ ), indicating the past and most recent studies continue to focus in these settings.

**Comparison of Publication Type and Design Quality Standards.** The current study identified 15 peer reviewed articles and 10 dissertations, with Westover (2010) being coded as both a dissertation and a peer-reviewed article. Three studies (13%) met the WWC standards, 14 (58%) studies met with reservations, and seven (29%) studies did not meet WWC standards. Morin (2017) identified 61% ( $n = 17$ ) peer-reviewed articles and 39% ( $n = 11$ ) dissertations. Of these, 50% ( $n = 14$ ) of studies met the standards with reservations and 39% ( $n = 11$ ) did not meet. In the most recent studies (i.e., Coogle et al., 2019; D'Agostino et al., 2020; Knight et al., 2018; McVittie, 2018; McLeod et al., 2019; Morin, 2017), four articles were peer reviewed and three met the WWC quality standards indicating the design of the studies methodologically adhere to the standards.

**Comparison of DVs Measured.** The current study reported teacher outcomes while Morin's (2017) meta-analysis reported student outcomes. Given this limitation, no comparison of DVs can be made between studies. The analysis of the teacher outcomes relies on the findings of this meta-analysis and the trend in study DVs preceding Morin's publication date. From 2016-2020, the current study identified teacher outcomes in the following categories: praise ( $n = 2$ ); student outcomes ( $n = 2$ ); OTR ( $n = 2$ ); implementation ( $n = 2$ ); instructional quality ( $n = 1$ ). This indicates that the most recent literature base focused on measuring praise, student outcomes, OTR, and implementation

## CHAPTER IV

### DISCUSSION

Through recent improvements in digital technology, VA has become a more commonly used tool for improving educator instructional quality (Knight et al., 2012). Currently, the VA literature base provides evidence on how to implement VA as either a teacher preparation tool for preservice teachers or as a professional development tool for inservice teachers and paraprofessionals. The purpose of this SR and meta-analysis was to understand the contribution that VA has made to the field of educator development. After completing a thorough SR of articles between 2010-2020, a total of 24 articles were identified that matched the inclusion criteria discussed in Chapter III. This chapter (a) summarizes the findings of each research question, (b) addresses the limitations of the SR and meta-analysis, (c) provides implications for future practice, and (d) draws a conclusion about the current use of VA.

**RQ 1: What is the status of the literature base on VA regarding study characteristics (i.e., publication type), participant characteristics (i.e., role, education level, experience level, age), student characteristics (i.e., disability type, student outcomes), and setting characteristics (i.e., grade level, group size, type of instruction, setting)?**

To better understand the current literature base of SCRD using VA, RQ 1 descriptively analyzes the (a) study and participant characteristics and (b) student and setting characteristics. Each of these characteristics and subcategories are addressed in greater detail below.

### ***Study and Participant Characteristics***

The current SR identified 15 (63%) peer-reviewed articles and 10 (42%) dissertations meeting inclusionary criteria. Additionally, the majority of the studies included inservice teachers ( $n = 15$ ; 63%), participants with three or more years of experience ( $n = 17$ ; 71%), participants who were 18-29 years of age ( $n = 12$ ; 50%), and participants who had bachelor's degrees ( $n = 13$ ; 54%). These findings suggest that participants in studies using VA are inservice teachers and have a minimum of three years of teaching experience. This result is consistent with previous research studies. For example, Webster et al. (2012) found similar results using video self-reflection as part of a treatment package with 51 Head Start teachers who had an average of 10 years of teaching experience. The Head Start teachers were assigned randomly to an experiment group (i.e., immediate video self-reflection or delayed video self-reflection) or the control group. The participants in the immediate and delayed video self-reflection groups increased the number of praise statements given, demonstrating that inservice teachers with experience improved their instructional skills by participating in VA as a professional practice. With more evidence demonstrating its effectiveness for inservice teachers, VA could be used as a professional development tool to help support educators.

Importantly, inservice teachers are not the only educators that interact with students and provide targeted supports. Preservice teachers and paraprofessionals both serve instructional roles and could potentially benefit the most from VA; however, they were less frequently studied. In comparison to inservice teachers, only five studies (21%) included preservice teachers and five studies (21%) included paraprofessionals. Given

these low study numbers, more studies using VA need to include both preservice teachers and paraprofessionals to determine VA treatment effects with all types of educators.

### ***Student and Setting Characteristics***

The findings also indicate the majority of studies ( $n = 12$ ; 50%) included students that were identified as having a developmental disability and about half of the studies ( $n = 11$ ; 46%) took place in an elementary school. In terms of classroom setting, half of the studies ( $n = 12$ ; 50%) took place in inclusion classrooms; most studies provided instruction in a small group setting ( $n = 11$ ; 46%); and over half of the studies ( $n = 13$ ; 54%) focused on teaching academic skills. Because students in these studies are receiving intervention or special education services and are the most at-risk students, instruction needs to be provided by a highly qualified and trained interventionist who has strong content and instructional knowledge (Johnson et al., 2013). To address this need, VA could be used as a professional development or teacher preparation tool to help support educators with little or no training in education. As a result, studies using VA need to be more inclusive of preservice teachers and paraprofessionals. Furthermore, studies need to include student outcomes to determine the efficacy of VA for at-risk students requiring individualized support.

**RQ 2: What is the status of the literature base on the research design quality for the included articles as measured by the What Works Clearinghouse (WWC) design quality standards (i.e., meets, meets with reservations, does not meet)?**

Of the 24 included studies, three (13%) studies met the WWC design quality standards, 14 (58%) studies met the WWC design quality standards with reservations, and seven (29%) studies did not meet the WWC design quality standards. These findings



demonstrate that approximately one-third of the studies did not meet the WWC standards. These rates are similar to other meta-analyses using SCRD (Barton et al., 2017; Barton, et al., 2020; Maggin et al., 2017). For example, Barton, et al., (2020) conducted a review of SCRD focused on student play interventions. As part of the study, the authors reviewed the rigor of the methods of the 27 included articles using the WWC standards and found that seven (26%) met the design quality standards, eleven (41%) met them with reservations, and nine (33%) did not meet WWC standards. The WWC standards were designed to address concerns about the reliability and interpretation of visual analysis of SCRD (Horner & Kratochwill, 2012). These WWC findings indicate that the standards are not being implemented regularly and point to the adolescence of the methodology. Due to the lack of studies adhering to high quality design standards in current study's SR, it can be concluded that more rigorous methods are required in this area of research to support its use as a potential EBP (Horner & Kratochwill, 2012; Odom, 2009).

For an approach to be recognized as an EBP, the practice used as a treatment in SCRD must have (a) a minimum of five studies using single case research methodology published in peer-reviewed journals, (b) a demonstration of a functional relation for each study, (c) variation in a minimum of three different research groups or settings, and (d) documentation of an effect for a total of 20 participants across all studies (Horner et al., 2005; Horner & Kratochwill, 2012). This meta-analysis includes SCRDs, which inherently have small sample sizes; thus, multiple studies are necessary to meet the requirement of an adequate sample size (Horner et al., 2005). When using these criteria within the current review, studies using behavior specific praise ( $n = 9$ ) as the dependent

variable met the requirements for an EBP. Praise showed promise by having large effect size (ES;  $g = 0.88-2.66$ ), but given the wide confidence intervals, the results should be interpreted with caution. The need to further replicate studies measuring praise and its ES is discussed in greater detail below.

### **RQ 3: What is the magnitude of effect of VA interventions on the instructional practices of educators?**

Although practices using SCRD have a standard of demonstrating a functional relation to be identified as an EBP using visual analysis standards, Horner and Kratochwill (2012) also urge the field to calculate a standardized ES. Having a standardized ES would allow results from SCRDs to be compared across research design methodologies (Pustejovsky, 2018; Shadish et al., 2015) and further validates the efficacy of the practice. To determine the ES of the included studies, a BC-SMD was used, which allows for a comparison across study designs (Pustejovsky, 2018). This research question examines the magnitude of effect by (a) participant characteristics and (b) type of dependent variable.

#### ***ES by Participant Characteristics***

When analyzing the participant characteristics (i.e., educator role, age, education level and teaching experience), findings show that studies with a large ES ( $g > 0.80$ ) included the following: Bishop et al., 2015; Capizzi et al., 2010; Coogle et al., 2019; D'Agostino et al., 2020; Englund, 2010; Fedders, 2011; Hawkins & Heflin, 2011; Knight et al., 2018; Morin, 2017; Pelletier et al., 2010; Smith, 2015; Starling, 2015; Westover, 2010). Across the studies with large effect sizes, ten studies included inservice teachers. Of the studies demonstrating a large ES that were not with inservice teachers, two studies

(Capizzi et al., 2010; Smith, 2015) included preservice teachers and two studies (Morin, 2017; Westover, 2010) included paraprofessionals as participants. Of these studies, nine had participants with three or more years of teaching experience, six had participants holding a bachelor's degree or higher, and eight had participants between the ages of 18-29.

The studies with a medium effect ( $g = 0.50-0.79$ ) included the following three studies: Fedders, 2011; Pinter et al., 2015; Smith, 2015. Two studies (Pinter et al., 2015;  $g = 0.66$ , Fedders, 2011 ;  $g = 0.57$ ) with a medium effect size included inservice teachers, one study (Smith, 2015;  $g = 0.72$ ) including preservice teachers had a medium effect size, and one study (Westover, 2010;  $g = -0.68$ ) including paraprofessionals had a negative medium effect size. Three studies included participants with no teaching experience and three studies included participants with three or more years of teaching experience. One study had participants with a high school or GED and bachelor's degree, one study had participants with some college experience, and one study had participants who held a master's degree. Figure 1.4 shows the ES based on participant characteristics.

Although no conclusions can be made as to why more studies that included teachers with more experience had a large ES, one reason could be that more experienced teachers, who may have more confidence, were more likely to participate in such a study than teachers with less experience. The large ES could also be simply a function of having more actual studies (i.e. numerically) that included this population as well.

### ***ES by Type of Dependent Variable***

When examining the ES by dependent variable (DV), results indicated that the effect of VA as an intervention varied by the outcome measure being used. A total of

nine DVs (i.e., praise, FOI, student outcomes, instructional quality, negative response, opportunities to respond, errors corrected, redirect, and instructional learning time) were measured across the studies. When examining the large effect sizes ( $g > 0.80$ ), behavior specific praise and FOI had six studies, student outcomes had three studies, instructional quality had two studies, opportunities to respond had two studies, and errors corrected had one study. Fedders (2011) measured negative response that had a large negative effect ( $g = -4.70$ ).

When examining the medium effect sizes ( $g = 0.50-0.79$ ), behavior specific praise, student outcomes, and opportunities to respond had one study each. Additionally, there was one study with a medium negative effect (Westover, 2010;  $g = -0.68$ ), which measured student outcomes (i.e., no response). No response was defined as the student not responding to the paraprofessional within 10-seconds (Westover, 2010). This student behavior had a negative effect, meaning that this behavior decreased or students responded more quickly, which is the expected trend for a no response behavior.

Finally, of the remaining studies, results demonstrated that one study had no effect ( $g < .2$ ; Starling, 2015) and five had a small effect ( $g = 0.20-0.49$ ; Hawkins & Heflin, 2011; MacVittie, 2018; Smith, 2015; Thompson et al., 2012; Westover, 2010). It is important to note that there was a study that measured two DVs (Starling, 2015) and had small negative effects. These DVs were focused on negative specific praise statements ( $g = -0.02$ ) and reprimands ( $g = -0.13$ ), so it would be expected that these behaviors would decrease once the intervention was introduced resulting in a negative trend in ES.

The DVs with the largest ES were praise and FOI. The six studies that measured praise had a moderate to large ES ( $g = -0.88-2.66$ ). FOI was the DV in six studies with moderate to large ES ( $g = 1.07-3.64$ ). Interestingly, the confidence intervals for both were quite wide, which is common in SCRD meta-analyses and is present in other studies (Barton et al., 2017; Maggin et al., 2017). However, it does demonstrate considerable variability of the effect.

The reason why VA may impact these DVs so strongly is because praise and FOI are discrete teaching behaviors that are easily identifiable and measurable making them ideal DVs for studies. Additionally, these procedural behaviors make the actions easier to implement in comparison to less discrete behaviors such as redirection and instructional time, which were also DVs of some of the included studies but had smaller ES.

Interestingly, there was variability of the effectiveness of VA within studies that measured multiple DVs, indicating that VA may impact some teacher behaviors differently than others. For example, Smith (2015) examined the use of VA and measured OTR, negative response, instructional learning time, and praise and found ES of  $g = 0.72$ ,  $g = 0.34$ ,  $g = 0.20$ , and  $g = 1.73$ , respectively. The effect sizes ranged from small (instructional learning) to large (praise). Similarly, Westover (2010) also measured multiple DVs and obtained ES's that ranged from small effect (redirect;  $g = 0.26$ ), negative moderate effect (i.e., student outcomes;  $g = -0.68$ ), to large (i.e., student outcomes,  $g = 1.27$ ; praise,  $g = 1.83$ ; error correction,  $g = 1.95$ ; OTR,  $g = 2.15$ ). These two studies further demonstrate the different impact of VA on various DVs. Figure 1.3 shows the ES based on all the DVs across studies.

Given that the study characteristics (i.e., participants) and interventions using VA were consistent in these studies, yet yielded different impact by DV makes an important claim that the field should carefully consider the role that the type of DV may impact the efficacy of using VA. Nagro et al. (2020) advocate for the field of VA to advance its understanding of the practice to include more challenging teaching behaviors, such as classroom management skills, and discuss the challenges of how to feasibly implement studies with more complex teaching behaviors. To implement VA with classroom management, Nagro et al. (2020) suggest the following procedures: (a) recording the lesson; (b) reviewing the video while using an observation tool to help focus attention to the targeted instructional components; (c) reflecting using a structured graphic organizer, (d) revising instruction for the betterment of students, and (e) then repeating the process. It is noted that discrete and less complex teaching behaviors, such as praise and FOI, are easily observable and measurable which may increase the reliability as well as the utility in those studies compared to using more complex instructional behaviors. One way to mitigate this measurement challenge is through the use of a standardized observation tool in which participants and coders are trained to reliability. Observation tools such as the Quality Intervention Delivery and Receipt (QIDR; Harn et al., 2011), Classroom Assessment Scoring System (CLASS; Pianta et al., 2008), and the Framework for Teaching (FFT; Danielson, 2011) are observation tools that may assist in evaluating more complex teaching behaviors by using them as a graphic organizer to guide reflection as suggested by Nagro et al. (2020).

**RQ 4: How has the literature base on VA changed since 2016 as reported by Morin's (2017) systematic review?**

Comparison of the current meta-analysis to the parent study (Morin, 2017) SR indicate that there has been little change regarding the type of articles and study characteristics (e.g., samples, research design, setting characteristics) examining VA. From 2016-2020, six VA-related articles (i.e., Coogle et al., 2019; D'Agostino et al., 2020; Knight et al., 2018; MacVittie, 2018; McLeod et al., 2019; Morin, 2017) were published. Similar to the parent study, the more recent studies reported that the participants were primarily inservice teachers ( $n = 5$ ), between 30-39 years of age ( $n = 4$ ), held a bachelor's degree ( $n = 5$ ), and had three or more years of experience ( $n = 5$ ). The most recent studies also primarily took place in inclusion classrooms ( $n = 5$ ), delivered in small groups ( $n = 3$ ), or in one-to-one environments ( $n = 3$ ). In these settings, teachers provided academic instruction ( $n = 2$ ), communication instruction ( $n = 2$ ), or the instruction was not reported ( $n = 2$ ). The student characteristics also focused on students with developmental disabilities ( $n = 5$ ). Finally, the most recent studies measured praise ( $n = 2$ ), student outcomes ( $n = 2$ ), OTR ( $n = 1$ ), implementation ( $n = 1$ ), and instructional quality ( $n = 1$ ).

Although technological advancements have been made making VA a more feasible tool for teacher development (Knight et al., 2012), these findings show a slight stagnation in the development of the field and indicate the need to increase and extend VA research to address the current identified gaps. One consideration would be to examine the reasons why the use of VA has not increased over the years. A plausible explanation is that teachers feel uncomfortable viewing their instruction (Mosley Wetzel et al., 2017). However, as educators watch videos of themselves they become more comfortable and accustomed to watching themselves teach (Hong & Van Riper, 2016).

This exposure to VA and self-reflective practices transforms teachers into lifelong learners (Benedict, et al., 2016; Harn & Meline, 2019; Tripp & Rich, 2012) who analyze and adapt their teaching to better support their students. Teachers can become more responsive to their students through the use of VA.

### **Limitations**

There were multiple limitations within this study. At the SR level, limitations occurred regarding the access to resources needed to replicate Morin's (2017) dissertation. Therefore, the current study used similar, but not the exact research and reference databases used in the parent study. This altered the articles collected in the primary search and the ancestral, citation, and first author searches. For example, Snyder (2013) was included in the parent study but was not identified in the current SR's collection process. Therefore, this study included the article in the full-text review because it met all the inclusionary criteria. Snyder (2013) was included in the descriptive data but was ultimately excluded in the statistical analysis because the study had fewer than three participants. Additionally, Morin (2017) included Lindsey (2013), which could not be obtained for the current study due to database and website restrictions. Finally, two articles were excluded because the video component used was an exemplar teacher, not the participant (i.e., Digennaro-Reed et al., 2010) or took place in a setting outside of the US (i.e., Stephenson et al., 2011). These restrictions made it challenging to conduct a direct replication of Morin's (2017) meta-analysis and highlights an important issue in attempting to replicate SR: the process and procedure of replication studies needs to be reproducible (Zwaan et al., 2018).



Additionally, this study used different research and reference databases as well as search engines than the parent study, which resulted in a difference between the collected articles. Even if the articles were identified in the searches and coded in the title and abstract review, it was challenging to obtain access to the full-text of some of the articles. Articles that were identified from the search but could not be obtained through inter-library loan ( $n = 7$ ) were not included; all of these were dissertations.

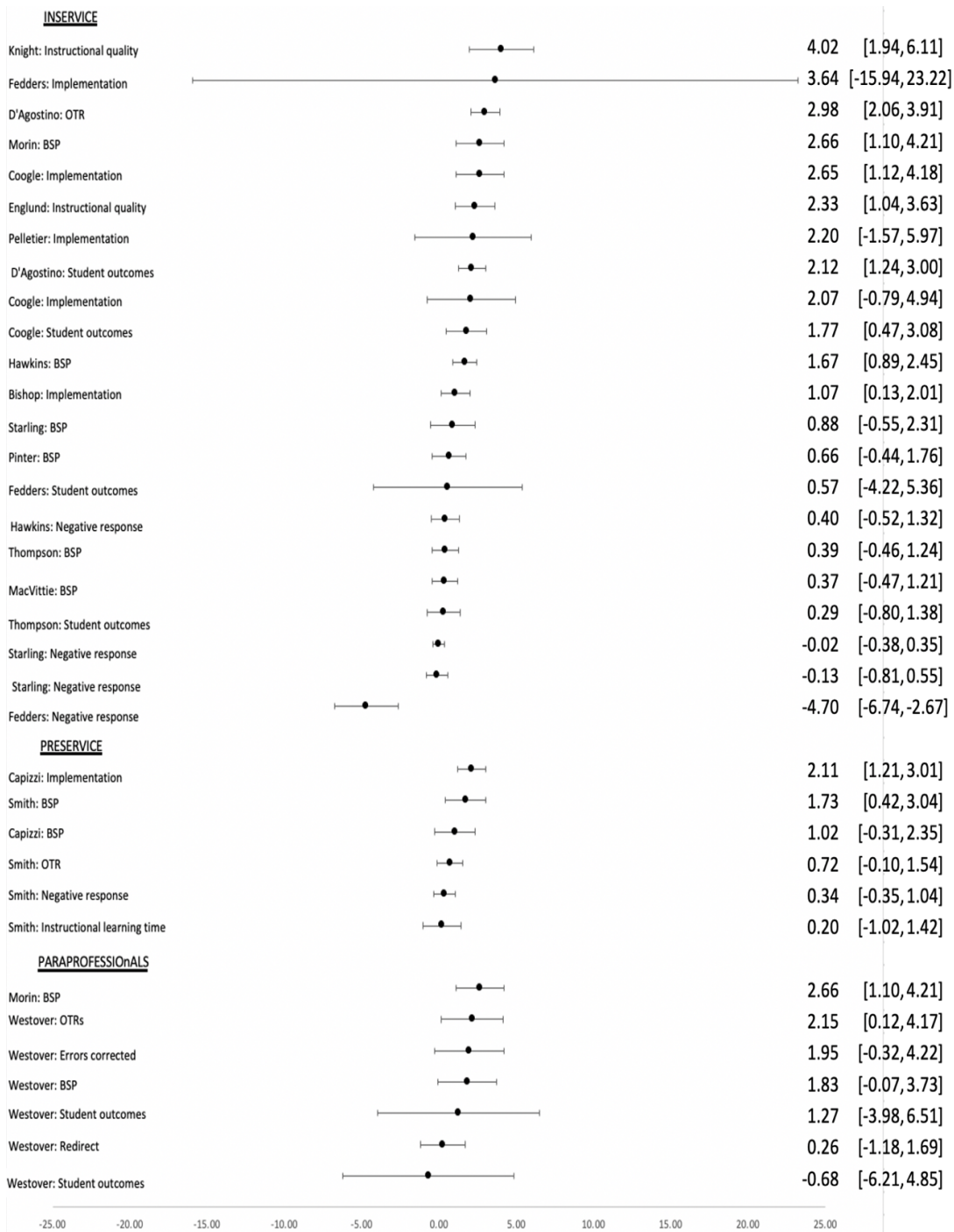
The inclusionary criteria also limited the types of studies that could be examined. Self-reflection is an essential piece of VA and the growth of teachers and reflective practices should be examined, yet the SR inclusionary criteria required that the teacher outcomes be observable and measurable. This restricted the ability to determine if the teacher's reflective ability as a component of VA had resulted in higher-levels of self-reflection.

Due to the limitations of the statistical analysis, I was unable to (a) isolate VA from other treatment packages, (b) examine moderator effects, and (c) calculate the robust variance estimation (RVE). When looking at the studies using VA as an intervention, it may have been included as part of a treatment package. For example, Coogle et al.'s (2019) study used a treatment package that combined both bug-in-ear and VA reflection. Educators received real-time coaching through bug-in-ear and also received an email with their instructional video, which they were to review and reflect upon. The use of two interventions used simultaneously made it challenging to determine if VA or a treatment package (i.e., VA and bug-in-ear coaching) was effective.

Relatedly, due to the lack of sufficient data and the small number of studies, Borenstein et al., (2009) recommends not statistically summarizing the moderators and

questions the reliability of the calculations if it were to be conducted. Given the current statistical procedures for meta-analyses, the current study did not calculate the moderator effects for SCRD studies using VA. Moderator analysis can be calculated using a t-test, analysis of variance, or regression model to determine the moderator of variables such as participants, setting, and student characteristics (Shadish et al., 2014). Regrettably, due to the small sample size and variability in the DVs across the included studies, a moderator analysis could not be conducted.

Finally, newly recommended meta-analytic methods propose calculating the omnibus ES using robust variance estimation (RVE), which accounts for unknown covariance and sampling distributions (Hedges et al., 2010). RVE is used for dependent ES (Fisher & Tipton, 2015; Hedges et al., 2010; Tanner et al., 2016), which occurs in SCRD, to determine the effect of a treatment on different outcomes (Hedges et al., 2010). Typical procedures for a meta-analysis includes first calculating the BC-SMD ES of individual DVs within a study and then calculating the RVE for the effects size of a DV across studies. Unfortunately, due to the multiple DVs (i.e., praise, implementation, student outcomes, instructional quality, error correction, instructional learning time, negative response, OTR, and redirect) across the included articles and low number of studies examining similar DVs, the RVE could not be calculated.



**Figure 1.4.** A forest plot displaying the BC-SMD ES for individual studies based upon the participant characteristics.

## **Implications for Future Practice**

The results of this SR and meta-analysis identify the issue of a small sample size and lack of methodological rigor for SCRD. These two concerns limit the generalizability of VA and impact it from being classified as an EBP. One way to overcome the issue of a small sample size and lack of rigorous methods is to use alternative research designs as recommended by Odom (2009) and Odom et al. (2005). The most common approach when studying VA is the use of SCRD; but by using group design or mixed method approaches with VA, it would help the field better understand its actual utility both by participant type (i.e., inservice, preservice, paraprofessional) and dependent variable. By increasing the number of participants in studies using VA, results can be more broadly generalized. These larger and more rigorous studies need to have diverse study characteristics and measure more complex classroom teaching behaviors to determine the effectiveness of VA. Group designs are more appropriate for better understanding the impact of an intervention (Odom, 2009; Odom et al., 2005). This could be applied to VA to better understand the role that study characteristics (i.e., participants) have on different dependent variables (e.g., OTR, praise, instructional quality, etc.)

## **Conclusion**

While the field continues to frequently use VA, the nature of the studies, primarily using single-case methodological approaches, minimizes our ability to call it an EBP. The current study highlights some of the challenges encountered in order for VA to be considered an EBP. First to become an EBP, studies using VA as a treatment must diversify the participant, student, and setting characteristics to increase the generalizability of the practice. Future studies should adhere to WWC standards for

SCRD to improve our understanding of the utility of VA. Another option is to consider alternative methodologies (e.g., quasi-experimental, group design, etc.) with larger sample sizes and varied study settings to enable different types of analyses that can be used to determine potential moderating variables related to the effectiveness of VA (e.g., type of DV, participant, etc.).

Additionally, this study highlights a challenge in truly “replicating” a study because of differential access/use of search engines yielding variable access to studies (e.g. dissertations). Finally, the advancements in the statistical analysis for completing a meta-analysis using SCRD is still evolving, so comparing these results to the outdated practice in the parent study is ill advised.

In relation to the use of VA as an intervention, the measured outcomes (DV) also need to be linked to student outcomes (Morin, Ganz, et al., 2019). This is particularly critical to professional development tools such as VA that aim to improve instructional skills that result in increased student outcomes. Across the included studies, five different studies (28%) measured student outcomes. More research needs to be conducted to understand under what conditions and in what manner VA can be used to more effectively improve instructional practices and impact student outcomes. In conclusion, VA continues to be a promising practice. Once the previously mentioned challenges are addressed and advancements in statistical analysis are made, VA has the potential to be identified an effective EBP.

**APPENDIX A**

**QUALTRICS FORM FOR CODING STUDY CHARACTERISTICS**



Coder Initials

Article number:

Article Citation: (Author, year)

Is the study conducted in the US?

### Research design

MBD: multiple baseline design

MPD: multiple probe design

Other: (e.g. ABA)

NOT SINGLE CASE

### Publication Tyle

Dissertation

Peer-reviewed article

None of the above

### Video Analysis:

Participants watch a video of themselves as the primary subject

Participants watch an exemplar video (e.g., video of a peer, video of another teacher)

No video analysis

Dependent variable:

Observable/behavioral teacher outcomes

Non-observable or non-behavioral teacher outcomes: reflection

## TEACHER DEMOGRAPHICS

Number of participants

Teacher role: (select all that apply)

**Inservice:** An inservice teacher is a lead classroom teacher or the primary individual responsible for delivering instruction. An in-service teacher may also be referred to as the lead teacher, special education teacher, general education teacher, credentialed teacher, teacher in-charge, etc.

**Preservice:** A preservice teacher is an individual currently enrolled in a teacher preparation program. A preservice teacher may also be referred to as a teacher candidate.

**Paraprofessional:** A paraprofessional provides support to students and is supervised by a credentialed or lead teacher. A paraprofessional may also be referred to as an aide, educational assistants, instructional aide, 1:1 aide, etc.

not reported

NO TEACHER or PARAPROFESSIONAL



Education level: (select all that apply)

High school or GED
Some college or specialized training (e.g., associate's degree)
Completed Bachelor's degree
Master's degree
Not reported

Experience level: (include all experience in a classroom; select all that apply)

0 years
1-2 years
3+ years
not reported

Age: (select all that apply)

18-29 years old

30-39 year old

40-49 years old

50+

not reported

## SETTING CHARACTERISTICS

---

Grade level: (select all that apply)

**Preschool:** Preschool includes students who are younger than 6 years of age OR are in grades K and below.

**Elementary school:** Elementary school includes students who are less than 12 years of age OR are in grades 1-5.

**Middle school:** Middle school includes students who are less than 14 years of age OR are in grades 6-8.

**High school:** High school includes students who are 14 years of age OR in grades 9-12.

Not reported

NO PRE-K to 12 SETTING

Group size: (select all that apply)

**One-to-one:** One-to-one group size is the ratio of one student to one educator (i.e., inservice teacher, paraprofessional, or preservice teacher).

**Small group:** A small group is a subset of students from a larger group who receive instruction. A small group could include centers, reading groups, etc. (2-9 students)

**Large group:** A large group is all students in a classroom who receive instruction at the same time. A large group could include whole group reading instruction, morning circle time, etc. (10+ students)

Not reported

Type of instruction: (select all that apply)

**Academic skills:** Academic skills are tools students need to complete intellectual tasks. Academic skills focus on math, reading, language arts, science, writing, etc. Within each of these categories, there is a subset of skills. For example, reading could include phonics, fluency, reading comprehension, etc.

**Communication or language skills:** Communication skills are tools students need to be able to relay information. Communication skills may include asking questions, making requests, using AAC, responding to questions, etc.

**Life skills:** Life skills are tools students need to accomplish tasks in their daily lives. Life skills include toileting, cooking, grocery shopping, dressing, eating, hygiene, etc.

Not reported

Setting: (select all that apply)

**Self-contained:** A self-contained classroom is where students with a disability spend all or a majority of their school time. A self-contained classroom includes a special education classroom, separate school, or specialized school for students with disabilities.

**Inclusion:** An inclusion setting is a classroom with students with and without disabilities receiving instruction.

**Resource classroom:** A resource classroom is where students with disabilities spend some of their time in a separate classroom receiving instruction. Students in this setting also spend time in a general education classroom setting.

**General education:** General education is the typical classroom. A general education is determined if none of the students in the class had a disability or if there is no mention of students with a disability.

Not reported

## STUDENT DEMOGRAPHICS

---

Disability: (select all that apply)

**Developmental disability:** A development disability is a disability that is present before adulthood. Developmental disabilities include autism spectrum disorder, intellectual disability, Down syndrome, or other developmental disorder.

**Physical disability:** A physical disability is a condition that impairs mobility. A physical disability may include cerebral palsy.

**Mental disability:** A mental disability is a condition that affects emotions, thinking, and/or behavior. A mental disability may include anxiety disorder, conduct disorder, bipolar disorder, depression, schizophrenia, and attention deficit hyperactivity disorder.

**Emotional or behavioral disorders:** An emotional or behavioral disability interferes with a person's ability to sustain relationships and results in the frequent use of inappropriate behavior. An emotional or behavioral disability may include oppositional defiant disorder.

**Learning disabilities:** A learning disability is a condition that impairs a student from acquiring a skill or knowledge as similar same-aged peers. A learning disability may include dyslexia or a specific learning disability.

**Cognitive disabilities:** A cognitive disability impairs mental functioning. A cognitive disability may include a brain injury or cognitive impairment.

**Not reported:** Disabilities not reported may include a developmental delay (e.g., fine motor, literacy, language, cognitive, etc.), general challenging behavior, or no disability identified.

Coder Initials

Article number:

Article Citation: (Author, year)

Is the study conducted in the US?

Yes

No

Not reported

## Research design

MBD: multiple baseline design

MPD: multiple probe design

Other: (e.g. ABA)

NOT SINGLE CASE

## Publication Tyle

Dissertation

Peer-reviewed article

None of the above

## Video Analysis:

Participants watch a video of themselves as the primary subject

Participants watch an exemplar video (e.g., video of a peer, video of another teacher)

No video analysis

Dependent variable:

Observable/behavioral teacher outcomes

Non-observable or non-behavioral teacher outcomes: reflection

## TEACHER DEMOGRAPHICS

Number of participants

Teacher role: (select all that apply)

**Inservice:** An inservice teacher is a lead classroom teacher or the primary individual responsible for delivering instruction. An in-service teacher may also be referred to as the lead teacher, special education teacher, general education teacher, credentialed teacher, teacher in-charge, etc.

**Preservice:** A preservice teacher is an individual currently enrolled in a teacher preparation program. A preservice teacher may also be referred to as a teacher candidate.

**Paraprofessional:** A paraprofessional provides support to students and is supervised by a credentialed or lead teacher. A paraprofessional may also be referred to as an aide, educational assistants, instructional aide, 1:1 aide, etc.

not reported

NO TEACHER or PARAPROFESSIONAL



Education level: (select all that apply)

High school or GED

Some college or specialized training (e.g., associate's degree)

Completed Bachelor's degree

Master's degree

Not reported

Experience level: (include all experience in a classroom; select all that apply)

0 years

1-2 years

3+ years

not reported

Age: (select all that apply)

18-29 years old

30-39 year old

40-49 years old

50+

not reported

## SETTING CHARACTERISTICS

---

Grade level: (select all that apply)

**Preschool:** Preschool includes students who are younger than 6 years of age OR are in grades K and below.

**Elementary school:** Elementary school includes students who are less than 12 years of age OR are in grades 1-5.

**Middle school:** Middle school includes students who are less than 14 years of age OR are in grades 6-8.

**High school:** High school includes students who are 14 years of age OR in grades 9-12.

Not reported

NO PRE-K to 12 SETTING

Group size: (select all that apply)

**One-to-one:** One-to-one group size is the ratio of one student to one educator (i.e., inservice teacher, paraprofessional, or preservice teacher).

**Small group:** A small group is a subset of students from a larger group who receive instruction. A small group could include centers, reading groups, etc. (2-9 students)

**Large group:** A large group is all students in a classroom who receive instruction at the same time. A large group could include whole group reading instruction, morning circle time, etc. (10+ students)

Not reported

Type of instruction: (select all that apply)

**Academic skills:** Academic skills are tools students need to complete intellectual tasks. Academic skills focus on math, reading, language arts, science, writing, etc. Within each of these categories, there is a subset of skills. For example, reading could include phonics, fluency, reading comprehension, etc.

**Communication or language skills:** Communication skills are tools students need to be able to relay information. Communication skills may include asking questions, making requests, using AAC, responding to questions, etc.

**Life skills:** Life skills are tools students need to accomplish tasks in their daily lives. Life skills include toileting, cooking, grocery shopping, dressing, eating, hygiene, etc.

Not reported

Setting: (select all that apply)

**Self-contained:** A self-contained classroom is where students with a disability spend all or a majority of their school time. A self-contained classroom includes a special education classroom, separate school, or specialized school for students with disabilities.

**Inclusion:** An inclusion setting is a classroom with students with and without disabilities receiving instruction.

**Resource classroom:** A resource classroom is where students with disabilities spend some of their time in a separate classroom receiving instruction. Students in this setting also spend time in a general education classroom setting.

**General education:** General education is the typical classroom. A general education is determined if none of the students in the class had a disability or if there is no mention of students with a disability.

Not reported

### Research design

MBD: multiple baseline design

MPD: multiple probe design

Other: (e.g. ABA)

NOT SINGLE CASE

### Publication Tyle

Dissertation

Peer-reviewed article

None of the above

### Video Analysis:

Participants watch a video of themselves as the primary subject

Participants watch an exemplar video (e.g., video of a peer, video of another teacher)

No video analysis

Dependent variable:

Observable/behavioral teacher outcomes

Non-observable or non-behavioral teacher outcomes: reflection

## TEACHER DEMOGRAPHICS

Number of participants

Teacher role: (select all that apply)

**Inservice:** An inservice teacher is a lead classroom teacher or the primary individual responsible for delivering instruction. An in-service teacher may also be referred to as the lead teacher, special education teacher, general education teacher, credentialed teacher, teacher in-charge, etc.

**Preservice:** A preservice teacher is an individual currently enrolled in a teacher preparation program. A preservice teacher may also be referred to as a teacher candidate.

**Paraprofessional:** A paraprofessional provides support to students and is supervised by a credentialed or lead teacher. A paraprofessional may also be referred to as an aide, educational assistants, instructional aide, 1:1 aide, etc.

not reported

NO TEACHER or PARAPROFESSIONAL

Education level: (select all that apply)

High school or GED
Some college or specialized training (e.g., associate's degree)
Completed Bachelor's degree
Master's degree
Not reported

Experience level: (include all experience in a classroom; select all that apply)

0 years
1-2 years
3+ years
not reported

Age: (select all that apply)

18-29 years old

30-39 year old

40-49 years old

50+

not reported

## SETTING CHARACTERISTICS

---

Grade level: (select all that apply)

**Preschool:** Preschool includes students who are younger than 6 years of age OR are in grades K and below.

**Elementary school:** Elementary school includes students who are less than 12 years of age OR are in grades 1-5.

**Middle school:** Middle school includes students who are less than 14 years of age OR are in grades 6-8.

**High school:** High school includes students who are 14 years of age OR in grades 9-12.

Not reported

NO PRE-K to 12 SETTING



Group size: (select all that apply)

**One-to-one:** One-to-one group size is the ratio of one student to one educator (i.e., inservice teacher, paraprofessional, or preservice teacher).

**Small group:** A small group is a subset of students from a larger group who receive instruction. A small group could include centers, reading groups, etc. (2-9 students)

**Large group:** A large group is all students in a classroom who receive instruction at the same time. A large group could include whole group reading instruction, morning circle time, etc. (10+ students)

Not reported

Type of instruction: (select all that apply)

**Academic skills:** Academic skills are tools students need to complete intellectual tasks. Academic skills focus on math, reading, language arts, science, writing, etc. Within each of these categories, there is a subset of skills. For example, reading could include phonics, fluency, reading comprehension, etc.

**Communication or language skills:** Communication skills are tools students need to be able to relay information. Communication skills may include asking questions, making requests, using AAC, responding to questions, etc.

**Life skills:** Life skills are tools students need to accomplish tasks in their daily lives. Life skills include toileting, cooking, grocery shopping, dressing, eating, hygiene, etc.

Not reported

Setting: (select all that apply)

**Self-contained:** A self-contained classroom is where students with a disability spend all or a majority of their school time. A self-contained classroom includes a special education classroom, separate school, or specialized school for students with disabilities.

**Inclusion:** An inclusion setting is a classroom with students with and without disabilities receiving instruction.

**Resource classroom:** A resource classroom is where students with disabilities spend some of their time in a separate classroom receiving instruction. Students in this setting also spend time in a general education classroom setting.

**General education:** General education is the typical classroom. A general education is determined if none of the students in the class had a disability or if there is no mention of students with a disability.

Not reported

## STUDENT DEMOGRAPHICS

---

Disability: (select all that apply)

**Developmental disability:** A development disability is a disability that is present before adulthood. Developmental disabilities include autism spectrum disorder, intellectual disability, Down syndrome, or other developmental disorder.

**Physical disability:** A physical disability is a condition that impairs mobility. A physical disability may include cerebral palsy.

**Mental disability:** A mental disability is a condition that affects emotions, thinking, and/or behavior. A mental disability may include anxiety disorder, conduct disorder, bipolar disorder, depression, schizophrenia, and attention deficit hyperactivity disorder.

**Emotional or behavioral disorders:** An emotional or behavioral disability interferes with a person's ability to sustain relationships and results in the frequent use of inappropriate behavior. An emotional or behavioral disability may include oppositional defiant disorder.

**Learning disabilities:** A learning disability is a condition that impairs a student from acquiring a skill or knowledge as similar same-aged peers. A learning disability may include dyslexia or a specific learning disability.

**Cognitive disabilities:** A cognitive disability impairs mental functioning. A cognitive disability may include a brain injury or cognitive impairment.

**Not reported:** Disabilities not reported may include a developmental delay (e.g., fine motor, literacy, language, cognitive, etc.), general challenging behavior, or no disability identified.

**APPENDIX B**

**QUALTRICS FORM FOR CODING WHAT WORKS CLEARINGHOUSE**

**DESIGN QUALITY STANDARDS**



Reviewer Initials

Article Number

Author, Year

**Design Standard 1: Manipulation of the Independent Variable (p. 16)**

1: Reports the manipulation of the independent variable

0: Does not report the manipulation of the independent variable

**Design Standard 2: Reporting Inter-Assessor Agreement (IAA) 9pg. 11-13)**

**Reporting IAA (Standard 2A)**

1: Reports IAA

0: Does not report IAA

Reporting IAA (Standard 2A)

1: Reports IAA

0: Does not report IAA

IAA Frequency (Standard 2B) (pg 13)

2: A minimum of 20% the sessions within each condition

1: A minimum of 20% the sessions without disaggregating by treatment or phase

0: No reporting of IAA for a minimum of 20% the sessions

IAA Quality (Standard 2C) (pg. 14)

1: Meets the minimum agreement of 80% for percent agreement or 60% for Kappa

0: Does not meet the minimum agreement of 80% for percent agreement or 60% for Kappa

Is the study a multiple baseline or alternating treatment design?

Yes

No

### Design Standard 3: Demonstration of Treatment Effects (p. 17)

1: Intervention effect shown by having three attempts over three points of time. OR Intervention effect shown by having three attempts over three points of time with a minimum of two condition (alternating treatment design)

0: Intervention effect not shown in three attempts over three points of time OR Intervention effect not shown in three attempts over three points of time with a minimum of two condition (alternating treatment design)

### Design Standard 4: Number of Data Points Per Phase

2: A minimum of five data points in the baseline and treatment phases OR A minimum five data points in the baseline and treatment phases (alternating treatment design)

1: A minimum of three data points in the baseline and treatment phases OR A minimum four data points in the baseline and treatment phases (alternating treatment design)

0: Less than three data points in the baseline and treatment phases OR Less than four data points in the baseline and treatment phases (alternating treatment design)

Is the study a multiple probe design?

Yes

No

## Design Standard 5: Multiple Probe Designs (pg. 23)

---

### Initial baseline (Standard 5A)

2: A minimum of three consecutive data points within the first three sessions of baseline for each level

1: A minimum of one data point within the first sessions of baseline for each level

0: Does not include a minimum of one data point within the first sessions of baseline for each level

---

### Probe Points Before the Intervention (Standard 5B)

2: A minimum of three consecutive data points within the first three sessions before introducing the intervention for each level.

1: A minimum of one data point within the first sessions before introducing the intervention for each level.

0: Does not include a minimum of one data point within the first sessions before introducing the intervention for each level

---

## Consideration of Additional Probe Points (Standard 5C)

1: Each unit of analysis (e.g., participant, behavior, etc.) that was still in baseline when intervention is introduced for the previous unit of analysis (e.g., participant, behavior, etc.) had a data point when the previous unit(s) first received the intervention or when the previous unit(s) reached the prespecified intervention criterion (i.e., 3 out of 5 correct before entering intervention) AND this data point is consistent in level and trend with the previous baseline data points in that unit.

0: Each unit of analysis (e.g., participant, behavior, etc.) that was still in baseline when intervention is introduced for the previous unit of analysis (e.g., participant, behavior, etc.) did not have a data point when the previous unit(s) first received the intervention or when the previous unit(s) reached the prespecified intervention criterion (i.e., 3 out of 5 correct before entering intervention) AND this data point is not consistent in level and trend with the previous baseline data points in that unit.

## *Overall Design Quality*

2: The highest score possible across all standards

1: A score of 1 on Standards 2, 4, or 5 (multiple probe) and not receiving a 0 on any of the Design Standards

0: Receiving a score of 0 on one or more of the Design Standards



Reviewer Initials

Article Number

Author, Year

Design Standard 1: Manipulation of the Independent Variable (p. 16)

1: Reports the manipulation of the independent variable

0: Does not report the manipulation of the independent variable

Design Standard 2: Reporting Inter-Assessor Agreement (IAA) 9pg. 11-13)

Reporting IAA (Standard 2A)

1: Reports IAA

0: Does not report IAA

Reporting IAA (Standard 2A)

1: Reports IAA

0: Does not report IAA

IAA Frequency (Standard 2B) (pg 13)

2: A minimum of 20% the sessions within each condition

1: A minimum of 20% the sessions without disaggregating by treatment or phase

0: No reporting of IAA for a minimum of 20% the sessions

IAA Quality (Standard 2C) (pg. 14)

1: Meets the minimum agreement of 80% for percent agreement or 60% for Kappa

0: Does not meet the minimum agreement of 80% for percent agreement or 60% for Kappa

Is the study a multiple baseline or alternating treatment design?

Yes

No

### Design Standard 3: Demonstration of Treatment Effects (p. 17)

1: Intervention effect shown by having three attempts over three points of time. OR Intervention effect shown by having three attempts over three points of time with a minimum of two condition (alternating treatment design)

0: Intervention effect not shown in three attempts over three points of time OR Intervention effect not shown in three attempts over three points of time with a minimum of two condition (alternating treatment design)

### Design Standard 4: Number of Data Points Per Phase

2: A minimum of five data points in the baseline and treatment phases OR A minimum five data points in the baseline and treatment phases (alternating treatment design)

1: A minimum of three data points in the baseline and treatment phases OR A minimum four data points in the baseline and treatment phases (alternating treatment design)

0: Less than three data points in the baseline and treatment phases OR Less than four data points in the baseline and treatment phases (alternating treatment design)

Is the study a multiple probe design?

Yes

No

## Design Standard 5: Multiple Probe Designs (pg. 23)

---

### Initial baseline (Standard 5A)

2: A minimum of three consecutive data points within the first three sessions of baseline for each level

1: A minimum of one data point within the first sessions of baseline for each level

0: Does not include a minimum of one data point within the first sessions of baseline for each level

---

### Probe Points Before the Intervention (Standard 5B)

2: A minimum of three consecutive data points within the first three sessions before introducing the intervention for each level.

1: A minimum of one data point within the first sessions before introducing the intervention for each level.

0: Does not include a minimum of one data point within the first sessions before introducing the intervention for each level

---

## Design Standard 5: Multiple Probe Designs (pg. 23)

---

### Initial baseline (Standard 5A)

2: A minimum of three consecutive data points within the first three sessions of baseline for each level

1: A minimum of one data point within the first sessions of baseline for each level

0: Does not include a minimum of one data point within the first sessions of baseline for each level

---

### Probe Points Before the Intervention (Standard 5B)

2: A minimum of three consecutive data points within the first three sessions before introducing the intervention for each level.

1: A minimum of one data point within the first sessions before introducing the intervention for each level.

0: Does not include a minimum of one data point within the first sessions before introducing the intervention for each level

---

## Consideration of Additional Probe Points (Standard 5C)

1: Each unit of analysis (e.g., participant, behavior, etc.) that was still in baseline when intervention is introduced for the previous unit of analysis (e.g., participant, behavior, etc.) had a data point when the previous unit(s) first received the intervention or when the previous unit(s) reached the prespecified intervention criterion (i.e., 3 out of 5 correct before entering intervention) AND this data point is consistent in level and trend with the previous baseline data points in that unit.

0: Each unit of analysis (e.g., participant, behavior, etc.) that was still in baseline when intervention is introduced for the previous unit of analysis (e.g., participant, behavior, etc.) did not have a data point when the previous unit(s) first received the intervention or when the previous unit(s) reached the prespecified intervention criterion (i.e., 3 out of 5 correct before entering intervention) AND this data point is not consistent in level and trend with the previous baseline data points in that unit.

## *Overall Design Quality*

2: The highest score possible across all standards

1: A score of 1 on Standards 2, 4, or 5 (multiple probe) and not receiving a 0 on any of the Design Standards

0: Receiving a score of 0 on one or more of the Design Standards

## REFERENCES CITED

- \*Alexander, M., Williams, N. A., & Nelson, K. L. (2012). When you can't get there: Using video self-monitoring as a tool for changing the behaviors of pre-service teachers. *Rural Special Education Quarterly*, 31, 18-24. <https://doi.org/10.1177/875687051203100404>
- Barton, E. E., Murray, R., O'Flaherty, C., Sweeney, E. M., & Gossett, S. (2020). Teaching object play to young children with disabilities: A systematic review of methods and rigor. *American Journal on Intellectual and Developmental Disabilities*, 125, 14-36. <https://doi.org/10.1352/1944-7558-125.1.14>
- Barton, E. E., Pustejovsky, J. E., Maggin, D. M., & Reichow, B. (2017). Technology-aided instruction and intervention for students with ASD: A meta-analysis using novel methods of estimating effect sizes for single-case research. *Remedial and Special Education*, 38(6), 371-386. <https://doi.org/10.1177/0741932517729508>
- Beauchamp, C. (2015). Reflection in teacher education: issues emerging from a review of current literature. *Reflective Practice*, 16, 123-141. <https://doi.org/10.1080/14623943.2014.982525>
- Benedict, A., Holdheide, L., Brownell, M., & Foley, A. M. (2016). Learning to teach: Practice-based preparation in teacher education [Special Issues Brief]. American Institutes for Research. [https://cedar.education.ufl.edu/wp-content/uploads/2016/07/Learning\\_To\\_Teach.pdf](https://cedar.education.ufl.edu/wp-content/uploads/2016/07/Learning_To_Teach.pdf)
- \*Bishop, C. D., Snyder, P. A., & Crow, R. E. (2015). Impact of video self-monitoring with graduated training on implementation of embedded instructional learning trials. *Topics in Early Childhood Special Education*, 35, 170-182. <https://doi.org/10.1177/0271121415594797>
- Borenstein, M. (1994). The case for confidence intervals in controlled clinical trials. *Controlled Clinical Trials*, 15, 411-428. [https://doi.org/10.1016/0197-2456\(94\)90036-1](https://doi.org/10.1016/0197-2456(94)90036-1)
- Borenstein, M. (2009). Effect sizes for continuous data. In Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.), *The handbook of research synthesis and meta-analysis*. (2nd ed., pp. 221-235). Russel Sage Foundation.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis: Statistics in practice*. Wiley.
- Borko, H., Jacobs, J., Eiteljorg, E., & Pittman, M. E. (2008). Video as a tool for fostering productive discussions in mathematics professional development. *Teaching and teacher education*, 24, 417-436. <https://doi.org/10.1016/j.tate.2006.11.012>

- \*Capizzi, A. M., Wehby, J. H., & Sandmel, K. N. (2010). Enhancing mentoring of teacher candidates through consultative feedback and self-evaluation of instructional delivery. *Teacher Education and Special Education, 33*, 191-212. <https://doi.org/10.1177/0888406409360012>
- Calandra, B., Gurvitch, R., & Lund, J. (2008). An exploratory study of digital video editing as a tool for teacher preparation. *Journal of Technology and Teacher Education, 16*(2), 137-153.
- Cohen J, (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
- Collin, S., Karsenti, T., & Komis, V. (2013). Reflective practice in initial teacher training: Critiques and perspectives. *Reflective practice, 14*, 104-117. <https://doi.org/10.1080/14623943.2012.732935>
- \*Coogle, C. G., Nagro, S., Regan, K., O'Brien, K. M., & Ottley, J. R. (2019). The impact of real-time feedback and video analysis on early childhood teachers' practice. *Topics in Early Childhood Special Education*. Advance online publication. <https://doi.org/10.1177/0271121419857142>
- Cook, B. G. (2014). A call for examining replication and bias in special education research. *Remedial and Special Education, 35*, 233-246. <https://doi.org/10.1177/0741932514528995>
- Cook, B. G., Collins, L. W., Cook, S. C., & Cook, L. (2016). A replication by any other name: A systematic review of replicative intervention studies. *Remedial and Special Education, 37*, 223-234. <https://doi.org/10.1177/0741932516637198>
- Cooper, H., Hedges, L., & Valentine, J. (Eds.). (2019). *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation. <http://www.jstor.org/stable/10.7758/9781610448864>
- Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. Available at [www.covidence.org](http://www.covidence.org)
- \*D'Agostino, S., Douglas, S. N., & Horton, E. (2020). Inclusive preschool practitioners' implementation of naturalistic developmental behavioral intervention using telehealth training. *Journal of Autism and Developmental Disorders, 50*, 864-880. <https://doi.org/10.1007/s10803-019-04319-z>
- Danielson, C. (2011). *The framework for teaching evaluation instrument*. The Danielson Group.



- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017, June). *Effective teacher professional development*. Learning Policy Institute.  
<https://learningpolicyinstitute.org/product/effective-teacher-professional-development-report>
- DeBettencourt, L. U., & Nagro, S. A. (2018). Tracking Special Education Teacher Candidates' Reflective Practices Over Time. *Remedial and Special Education, 40*, 277-288. <https://doi.org/10.1177/0741932518762573>
- Dewey, J. (1933). *How we think*. Prometheus Books.
- Digennaro-Reed, F. D., Coddling, R., Cantania, C. N., & Maguire, H. (2010). Effects of video modeling on treatment integrity of behavioral interventions. *Journal of Applied Behavior Analysis, 43*, 291-295. <https://doi.org/10.1901/jaba.2010.43-291>
- \*Englund, L. W. (2010). *Evaluating and improving the quality of teacher's language modeling in early childhood classrooms* [Doctoral dissertation, University of Nevada, Las Vegas]. Digital Scholarship@UNLV.  
<https://digitalscholarship.unlv.edu/thesesdissertations/722/>
- Etscheidt, S., Curran, C. M., & Sawyer, C. M. (2012). Promoting reflection in teacher preparation programs: A multilevel model. *Teacher education and Special education, 35*, 7-26. <https://doi.org/10.1177/0888406411420887>
- Every Student Succeeds Act of 2015, Pub. L. No. 114-95 § 114 Stat. 1177 (2015).  
<https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf>
- 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. <https://doi.org/10.1177/0014402914551738>
- \*Fedders, A. M. (2011). The effect of video self-monitoring on novice special educators' implementation of advanced direction instruction reading techniques [Doctoral dissertation, University of California, Santa Barbara]. ProQuest.  
<https://www.proquest.com/docview/923804213>
- Fernandez, C. (2002). Learning from Japanese approaches to professional development: The case of lesson study. *Journal of teacher education, 53*, 393-405.  
<https://doi.org/10.1177/002248702237394>
- Fisher, Z., & Tipton, E. (2015). *Robumeta: An R-package for robust variance estimation in meta-analysis* [Statistical package]. <https://cran.r-project.org/web/packages/robumeta/vignettes/robumetaVignette.pdf>

- Gaudin, C., & Chaliès, S. (2015). Video viewing in teacher education and professional development: A literature review. *Educational Research Review*, *16*, 41-67. <https://doi.org/10.1016/j.edurev.2015.06.001>
- Gearing, R. E., El-Bassel, N., Ghesquiere, A., Baldwin, S. Gillies, J., & Ngeow, E.. (2011). Major ingredients of fidelity: A review and scientific guide to improving quality of intervention research implementation. *Clinical Psychology Review*, *31*(1), 79–88. <https://doi.org/10.1016/j.cpr.2010.09.007>
- \*Hager, K. D. (2012). Self-monitoring as a strategy to increase student teachers' use of effective teaching practices. *Rural Special Education Quarterly*, *31*, 9-17. <https://doi.org/10.1177/875687051203100403>
- Harn, B. (2017) Making RTI effective by coordinating the system of instructional supports. *Perspectives on Language and Literacy* , *43*, 15-18.
- Harn, B. A., Forbes-Spear, C., Fritz, R., & Berg, T. (2011). *Quality of Intervention Delivery and Receipt (QIDR) observation tool*. Eugene, OR.
- Harn, B., & Meline, M. (2019). Developing Critical Thinking and Reflection in Teachers Within Teacher Preparation. In Mariano, G. J. & Figiliano, F. J. (Eds.), *Handbook of Research on Critical Thinking Strategies in Pre-Service Learning Environments* (pp. 126-145). IGI Global.
- \*Hawkins, S. M., & Heflin, L. J. (2011). Increasing secondary teachers' behavior-specific praise using a video self-modeling and visual performance feedback intervention. *Journal of Positive Behavior Interventions*, *13*, 97-108. <https://doi.org/10.1177/1098300709358110>
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107–128. <https://doi.org/10.3102/10769986006002107>
- Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods*, *3*, 224-239. <https://doi.org/10.1002/jrsm.1052>
- 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. <https://doi.org/10.1002/jrsm.1086>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, *1*, 39-65. <https://doi.org/10.1002/jrsm.5>

- Hong, C. E., & Van Riper, I. (2016). Enhancing teacher learning from guided video analysis of literacy instruction: An interdisciplinary and collaborative approach. *Journal of Inquiry and Action in Education*, 7(2), 94-110. <https://eric.ed.gov/?id=EJ1133602>
- 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. <https://doi.org/10.1177/001440290507100203>
- Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004). <https://sites.ed.gov/idea/>
- Ioannidis, J. (2005). Why most published research findings are false. *PLoS Medicine*, 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- Ioannidis, J. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, 7, 645– 654. <https://doi.org/10.1177/1745691612464056>.
- Johnson, E. S., Carter, D. R., & Pool, J. L. (2013). Introduction to the special issue: The critical role of a strong tier 2 system. *Intervention in School and Clinic*, 48, 195-197. <https://doi.org/10.1177/1053451212462877>
- Kagan, D. M. (1993). Contexts for the use of classroom cases. *American Educational Research Journal*, 30, 703–723. <https://doi.org/10.3102/00028312030004703>
- Knight, J., Bradley, B. A., Hock, M., Skrtic, T. M., Knight, D., Brasseur-Hock, I., Clark, J, Ruggles, M. & Hatton, C. (2012). Record, replay, reflect. *Journal of Staff Development*, 33(2), 18-23. <https://www.proquest.com/docview/1015816235>
- \*Knight, D., Hock, M., Skrtic, T. M., Bradley, B. A., & Knight, J. (2018). Evaluation of video-based instructional coaching for middle school teachers: Evidence from a multiple baseline study. *The Educational Forum*, 82, 425-442. <https://doi.org/10.1080/00131725.2018.1474985>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 33, 159-174. <https://doi.org/10.2307/2529310>
- Lee, G. C., & Wu, C. C. (2006). Enhancing the teaching experience of pre-service teachers through the use of videos in web-based computer-mediated communication (CMC). *Innovations in Education and Teaching International*, 43, 369-380. <http://doi.org/10.1080/14703290600973836>
- Leins Dvorchak, J. (2015). *Increasing Secondary Teachers' use Of Praise With Video Performance Feedback* [Doctoral dissertation, University of Pittsburgh]. <http://d-scholarship.pitt.edu/25179/>

- Levy, Y., & Ellis, T. J. (2006). A systems approach to conduct an effective literature review in support of information systems research. *Informing Science: The International Journal of Emerging Transdiscipline*, 9, 181-212.  
<https://doi.org/10.28945/479>
- Lindsey, R. (2014). *Increasing the use of prompting strategies: A multiple baseline study across pairs of paraeducators of students with disabilities* [Doctoral dissertation, John Hopkins University]. ProQuest. [www.proquest.com/docview/1507591541](http://www.proquest.com/docview/1507591541)
- \*Lynes, M. J. (2012). *The effects of self-evaluation with video on the use of oral language development strategies by preschool teachers*. [Doctoral dissertation, The University of Utah]. <https://www.proquest.com/docview/1012121473>
- \*MacVittie, N. S. (2018). *Guided self-reflection with video and changes in teacher instructional behaviors* [Doctoral dissertation, George Mason University]. ProQuest. [www.proquest.com/docview/2070496459](http://www.proquest.com/docview/2070496459)
- Maggin, D. M., Pustejovsky, J. E., & Johnson, A. H. (2017). A meta-analysis of school-based group contingency interventions for students with challenging behavior: An update. *Remedial and Special Education*, 38, 353-370.  
<https://doi.org/10.1177/0741932517716900>
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, 7, 537-542. <https://doi.org/10.1177/1745691612460688>
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43, 304–316.  
<https://doi.org/10.3102/0013189X14545513>
- McLeod, R. H., Kim, S., & Resua, K. A. (2019). The effects of coaching with video and email feedback on preservice teachers' use of recommended practices. *Topics in Early Childhood Special Education*, 38, 192-203.  
<https://doi.org/10.1177/0271121418763531>
- Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: A comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research*, 14, 579. <https://doi.org/10.1186/s12913-014-0579-0>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Medicine* 6(7), e1000097.  
<https://doi.org/10.1371/journal.pmed1000097>

- Moher, D., Tetzlaff, J., Tricco, A. C., Sampson, M., & Altman, D. G. (2007). Epidemiology and reporting characteristics of systematic reviews. *PLoS Medicine*, 4(3). <https://doi.org/10.1371/journal.pmed.0040078>
- \*Morin, K. (2017). *The use of video analysis to change special educators' instructional practices: A single-case study and meta-analysis* [Doctoral dissertation, Texas A&M University]. Texas A&M University Libraries: OAKTrust. <http://hdl.handle.net/1969.1/165113>
- Morin, K. L., Ganz, J. B., Vannest, K. J., Haas, A. N., Nagro, S. A., Peltier, C. J., Fuller, M. C., & Ura, S. K. (2019). A systematic review of single-case research on video analysis as professional development for special educators. *The Journal of Special Education*, 53, 3-14. <https://doi.org/10.1177/0022466918798361>
- Morin, K. L., Nagro, S., Artis, J., Haas, A., Ganz, J. B., & Vannest, K. J. (2019). Differential effects of video analysis for special educators related to intervention characteristics, dependent variables, and student outcomes: A meta-analysis of single-case research. *Journal of Special Education Technology*. Advance online publication. <https://doi.org/10.1177/0162643419890250>
- Mosley Wetzel, M., Maloch, B., & Hoffman, J. V. (2017). Retrospective video analysis: a reflective tool for teachers and teacher educators. *The Reading Teacher*, 70, 533-542. <https://doi.org/10.1002/trtr.1550>
- \*Murphy, A., Robinson, S. E., Cote, D. L., Karge, B. D., & Lee, T. (2015). A teacher's use of video to train paraprofessionals in pivotal response techniques. *The Journal of Special Education Apprenticeship*, 4(2), 1-19. <https://eric.ed.gov/?id=EJ1127774>
- Nagro, S. A. (2020). Reflecting on others before reflecting on self: Using video evidence to guide teacher candidates' reflective practices. *Journal of Teacher Education*, 71, 420-433. <https://doi.org/10.1177/0022487119872700>
- Nagro, S. A., & Cornelius, K. E. (2013). Evaluating the evidence base of video analysis: A special education teacher development tool. *Teacher Education and Special Education*, 36, 312-329. <https://doi.org/10.1177/0888406413501090>
- Nagro, S. A., & deBettencourt, L. U. (2019). Reflection activities within clinical experiences: An important component of field-based teacher education. In *Handbook of Research on Field-Based Teacher Education* (pp. 565-586). IGI Global.
- Nagro, S. A., deBettencourt, L. U., Rosenberg, M. S., Carran, D. T., & Weiss, M. P. (2017). The effects of guided video analysis on teacher candidates' reflective ability and instructional skills. *Teacher Education and Special Education*, 40, 7-25. <https://doi.org/10.1177/0888406416680469>

- Nagro, S. A., Hirsch, S. E., & Kennedy, M. J. (2020). A self-led approach to improving classroom management practices using video analysis. *TEACHING Exceptional Children*, 53, 24-32. <https://doi.org/10.1177/0040059920914329>
- Odom, S. L. (2009). The tie that binds: Evidence-based practice, implementation science, and outcomes for children. *Topics in Early Childhood Special Education*, 29, 53-61. <https://doi.org/10.1177/0271121408329171>
- Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children*, 71, 137-148. <https://doi.org/10.1177/001440290507100201>
- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K-12 curriculum intervention research. *Review of Educational Research*, 78, 33-84. <https://doi.org/10.3102/0034654307313793>
- Olson, J. K., Bruxvoort, C. N., & Vande Haar, A. J. (2016). The impact of video case content on preservice elementary teachers' decision-making and conceptions of effective science teaching. *Journal of Research in Science Teaching*, 53(10), 1500-1523.
- Osipova, A., Prichard, B., Boardman, A. G., Kiely, M. T., & Carroll, P. E. (2011). Refocusing the lens: Enhancing elementary special education reading instruction through video self-reflection. *Learning Disabilities Research & Practice*, 26, 158-171. <https://doi.org/10.1111/j.1540-5826.2011.00335.x>
- Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science*, 7, 528-530. <https://doi.org/10.1177/1745691612465253>
- 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. <https://doi.org/10.1016/j.beth.2010.08.006>
- Partin, T. C. M., Robertson, R. E, Maggin, D, M, Oliver, R. M, & Wehby, J. H. (2009). Using teacher praise and opportunities to respond to promote appropriate student behavior. *Preventing School Failure*, 54, 172-178. <https://doi.org/10.1080/10459880903493179>
- \*Pelletier, K., McNamara, B., Braga-Kenyon, P., & Ahearn, W. H. (2010). Effect of video self-monitoring on procedural integrity. *Behavioral Interventions*, 25, 261-274. <https://doi.org/10.1002/bin.316>



- Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System™ : Manual K-3*. Paul H Brookes Publishing.
- \*Pinter, E. B., East, A., & Thrush, N. (2015). Effects of a video-feedback intervention on teachers' use of praise. *Education and Treatment of Children, 38*, 451-472. <https://doi.org/10.1353/etc.2015.0028>
- Pustejovsky, J. E., Hedges, L. V., & Shadish, W. R. (2014). Design-comparable effect sizes in multiple baseline designs: A general modeling framework. *Journal of Educational and Behavioral Statistics, 39*, 368-393. <https://doi.org/10.3102/1076998614547577>
- Pustejovsky, J. E. (2020). *scdhlms: Estimating hierarchical linear models for single-case designs* [Statistical package]. <https://cran.r-project.org/web/packages/scdhlms/scdhlms.pdf>
- Pustejovsky, J. E. (2018, February 1). *Effect size measures for single-case research: General considerations*. Advanced Training Institute on Single-Case Research Methods. <https://singlecaseinstitute.uoregon.edu/2018/02/01/effect-sizes-and-single-case-research/>
- Reinke, W. M., Lewis-Palmer, T., & Martin, E. (2007). The effect of visual performance feedback on teacher use of behavior-specific praise. *Behavior Modification, 31*, 247-263. <https://doi.org/10.1177/0145445506288967>
- Rich, P. J., & Hannafin, M. (2009). Video annotation tools: Technologies to scaffold, structure, and transform teacher reflection. *Journal of Teacher Education, 60*, 52-67. <https://doi.org/10.1177/0022487108328486>
- Richards, K. A. R., Templin, T. J., & Gaudreault, K. L. (2013). Understanding the realities of school life: Recommendations for the preparation of physical education teachers. *Quest, 65*, 442-457. <https://doi.org/10.1080/00336297.2013.804850>
- Roberts, C. A., Benedict, A. E., Kim, S. Y., & Tandy, J. (2018). Using lesson study to prepare preservice special educators. *Intervention in School and Clinic, 53*(4), 237-244. <https://doi.org/10.1177/1053451217712974>
- \*Robinson, S. E. (2011). Teaching paraprofessionals of students with autism to implement pivotal response treatment in inclusive school settings using a brief video feedback training package. *Focus on Autism and Other Developmental Disabilities, 26*, 105-118. <https://doi.org/10.1177/1088357611407063>
- Robinson, L., & Kelley, B. (2007). Developing reflective thought in preservice educators: Utilizing role-plays and digital video. *Journal of Special Education Technology, 22*(2), 31-43. <https://doi.org/10.1177/016264340702200203>

- Scheeler, M. C., Ruhl, K. L., & McAfee, J. K. (2004). Providing performance feedback to teachers: A review. *Teacher Education and Special Education*, 27, 396-407.  
<https://doi.org/10.1177/088840640402700407>
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13, 90–100.  
<https://doi.org/10.1037/a0015108>
- Shadish, W. R., Hedges, L. V., Horner, R. H., & Odom, S. L. (2015). *The role of between-case effect size in conducting, interpreting, and summarizing single-case research*. National Center for Education Research.  
<https://files.eric.ed.gov/fulltext/ED562991.pdf>
- Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention*, 2, 188-196.  
<https://doi.org/10.1080/17489530802581603>
- Sherin, M., & van Es, E. (2005). Using video to support teachers' ability to notice classroom interactions. *Journal of technology and teacher education*, 13(3), 475-491.
- Sims, L., & Walsh, D. (2009). Lesson study with preservice teachers: Lessons from lessons. *Teaching and Teacher Education*, 25(5), 724-733.
- Sinclair, A. C., Gesel, S. A., LeJeune, L. M., & Lemons, C. J. (2019). A review of the evidence for real-time performance feedback to improve instructional practice. *The Journal of Special Education*, 54, 90-100.  
<https://doi.org/10.1177/0022466919878470>
- \*Smith, C. L. (2015). *Effects of video feedback and self-assessment on the performance of evidence-based teaching strategies* [Doctoral dissertation, University of Georgia]. Athenaeum@UGA. <https://athenaeum.libs.uga.edu/handle/10724/34655>
- \*Snyder, C. K. (2013). *Effects of Training on Early Childhood Special Education Paraeducators' Use of Early Literacy Strategies During Book Reading* [Doctoral dissertation, University of Kansas]. KU ScholarWorks.  
<https://kuscholarworks.ku.edu/handle/1808/15109>
- Spalding, E., & Wilson, A., (2002). Demystifying reflection: A study of pedagogical strategies that encourage reflective journal writing. *Teachers College Record*, 104(7), 1393-1421.



- Stains, Marilyne, & Vickrey, Trisha. (2017). Fidelity of Implementation: An Overlooked Yet Critical Construct to Establish Effectiveness of Evidence-Based Instructional Practices. *CBE Life Sciences Education*, 16(1), rm1.  
<https://doi.org/10.1187/cbe.16-03-0113>
- \*Starling, N. R. (2015). *The Effectiveness of Video Self-Modeling on Increasing and Sustaining Teacher Use of Behavior-Specific Praise in the Alternative Classroom* [Doctoral dissertation, University of Connecticut]. OpenCommons@UConn.  
<https://opencommons.uconn.edu/cgi/viewcontent.cgi?article=6991&context=dissertations>
- Stephenson, J., Carter, M., & Arthur-Kelly, M. (2011). Professional learning for teachers without special education qualifications working with students with severe disabilities. *Teacher Education and Special Education*, 34, 7-20.  
<https://doi.org/10.1177/0888406410384407>
- Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, 2, 85-112.  
<https://doi.org/10.1007/s40865-016-0026-5>
- \*Thompson, M. T., Marchant, M., Anderson, D., Prater, M. A., & Gibb, G. (2012). Effects of tiered training on general educators' use of specific praise. *Education and Treatment of Children*, 35, 521-546. <https://doi.org/10.1353/etc.2012.0032>
- Therrien, W. J., Mathews, H. M., Hirsch, S. E., & Solis, M. (2016). Progeny review: An alternative approach for examining the replication of intervention studies in special education. *Remedial and Special Education*, 37, 235-243.  
<https://doi.org/10.1177/0741932516646081>
- Tracz, S. M., Daughtry, J., Henderson-Sparks, J., Newman, C., & Sienty, S. (2005). The impact of NBPTS participation on teacher practice: learning from teacher perspectives. *Educational Research Quarterly*, 28(3), 35-50.  
<https://files.eric.ed.gov/fulltext/EJ718123.pdf>
- Tripp, T. R., & Rich, P. J. (2012). The influence of video analysis on the process of teacher change. *Teaching and teacher education*, 28, 728-739.  
<https://doi.org/10.1016/j.tate.2012.01.011>
- Van Es, E. A., & Sherin, M. G. (2002). Learning to notice: Scaffolding new teachers' interpretations of classroom interactions. *Journal of Technology and Teacher Education*, 10(4), 571-596.

- Valentine, J. C., Tanner-Smith, E. E., Pustejovsky, J. E., & Lau, T. S. (2016). Between-case standardized mean difference effect sizes for single-case designs: a primer and tutorial using the scdhlms web application. *Campbell Systematic Reviews*, *12*, 1-31. <https://doi.org/10.4073/cmdp.2016.1>
- Watkins, M. W., & Pacheco, M. (2000). Interobserver agreement in behavioral research: Importance and calculation. *Journal of Behavioral Education*, *10*(4), 205-212. <https://doi.org/10.1023/A:1012295615144>
- Weber, K. E., Gold, B., Prilop, C. N., & Kleinknecht, M. (2018). Promoting pre-service teachers' professional vision of classroom management during practical school training: Effects of a structured online-and video-based self-reflection and feedback intervention. *Teaching and Teacher Education*, *76*, 39-49. <https://doi.org/10.1016/j.tate.2018.08.008>
- \*Westover, J. M. (2010). *Increasing the Literacy Skills of Students Who Require AAC through Modified Direct Instruction and Specific Instructional Feedback* [Doctoral dissertation, University of Oregon]. ProQuest. <https://www.proquest.com/docview/749881596>
- Westover, J. M., & Martin, E. J. (2014). Performance feedback, paraeducators, and literacy instruction for students with significant disabilities. *Journal of Intellectual Disabilities*, *18*, 364-381. <https://doi.org/10.1177/1744629514552305>
- What Works Clearinghouse. (2020). *What Works Clearinghouse Standards Handbook* (Version 4.1). Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <https://ies.ed.gov/ncee/wwc/Docs/referenceresources/WWC-Standards-Handbook-v4-1-508.pdf>
- Yoon, K. S., Duncan, T., Lee, S. W.-Y., Scarloss, B., & Shapley, K. L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement*. (REL 2007-No. 033). Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <https://files.eric.ed.gov/fulltext/ED498548.pdf>
- Zhang, M., Lundeberg, M., Koehler, M. J., & Eberhardt, J. (2011). Understanding affordances and challenges of three types of video for teacher professional development. *Teaching and teacher education*, *27*, 454-462. <https://doi.org/10.1016/j.tate.2010.09.015>
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Making replication mainstream. *Behavioral and Brain Sciences*, *41*. <https://doi.org/10.1017/S0140525X17001972>