THE USE OF VIDEO ANALYSIS TO CHANGE SPECIAL EDUCATORS' INSTRUCTIONAL PRACTICES: A SINGLE-CASE STUDY AND META-ANALYSIS

A Dissertation

by

KRISTI MORIN

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Jennifer B. Ganz
Co-Chair of Committee,	Kimberly J. Vannest
Committee Members,	Patricia Lynch
	Lei-Shih Chen
	Victor Willson
Head of Department,	Shanna Hagan-Burke

August 2017

Major Subject: Educational Psychology

ABSTRACT

The traditional, lecture-based model of professional development is generally not effective for changing the instructional practices of educators. While previous research has demonstrated that video analysis, a method of professional development that involves watching videos of oneself teaching, may be a viable alternative, a lack of high-quality design studies and statistical analyses of the literature base limits the conclusions that can be drawn. The purpose of this research is to remediate these gaps in the literature by conducting two studies: a well-designed single-case experimental design study, including a multiple-baseline across participants and two changing criterion designs, and a meta-analytic review of the research base on video analysis.

The results of video analysis were positive for both studies. Statistical and visual analyses indicated that video analysis was effective for changing the instructional practices of educators in the first study. Generalization, maintenance, and social validity data were also positive and indicated that video analysis (a) generalized to a second behavior, (b) maintained for all but one participant, and (c) was viewed favorably by all participants. Results of effect size analyses conducted in the second study showed moderate effects for video analysis when used to change the instructional practices of educators. Both methodological quality and publication type were investigated as potential moderators and neither were statistically significant, indicating they did not impact the results. Potential moderators related to participant and instructional

characteristics were also analyzed and all subgroups showed moderate to strong effects, with only role being statistically significant.

The results of this research have implications for providing professional development opportunities to educators. Both studies demonstrated moderate to strong effects, indicating that overall video analysis is a viable alternative to the traditional, lecture-based method. Several limitations are noted in both studies, including a short maintenance period and the omission of student outcome data in the first study and the inclusion of only single-case research data in the second study. Implications for future research are also addressed.

DEDICATION

First and foremost, to my Lord and Savior, Jesus, for giving me the passion to work with children with autism and for leading me to where I am today. Without You, I am nothing. I pray that I can glorify you in everything I do.

To my mother and grandparents. You taught me from an early age that I could achieve anything I set my mind to do. Thank you for always believing in me and for lending an ear when I needed to talk or saying a prayer for me when I needed one most.

To my father. You taught me so much during your time on this earth—honesty, integrity, and hard work, to name a few. While you won't be here to see me receive my degree, I can only pray that you are looking down from heaven and that you are proud.

To my in-laws. I cannot even begin to thank you enough for all you've done for our family. Your constant support and willingness to care for our children anytime it was needed helped make graduation a reality. And you did it all without complaint. I love you both and am forever grateful for having you in our lives.

To my children. My sweet boys. I love you both. While having a mother in school may not have been easy, I hope I inspired you to always pursue your dreams.

Lastly, and most importantly, to my amazing husband. To say this journey would not have been possible without you is an understatement. You have contributed so much over the years, from being a voice of reason when I needed it to taking on both of our roles as parents when I completed my internship. I cannot begin to express my gratitude and my love for you.

ACKNOWLEDGEMENTS

I would like to thank my chair and mentor, Dr. Jennifer Ganz, for taking me under her wing and teaching me what it means to be a researcher. The countless hours you have spent providing me with feedback and guidance have helped shaped me into the person I am today. I am incredibly grateful for the numerous opportunities you have provided me with over the years and the support you have shown when I wanted to pursue my own research interests. It is largely because of you that I will be taking the next step in my career following graduation. I am grateful for your mentorship and I'm fortunate to call you not only my advisor and mentor, but also my friend.

Thank you also to my co-chair, Dr. Kimberly Vannest. You have contributed much over the years to both my education and future career. Thank you for believing in me and for going above and beyond to help make what started out as ideas being thrown back and forth in your office to the completed work contained within these pages. I couldn't have asked for a better co-chair and future colleague. You always made me feel included, whether sitting around a table with other faculty members or having dinner at a conference with prominent researchers in the field. You cheered me on when I needed it, and were always available to offer advice when asked. Thank you for all this and more.

I am also appreciative for the guidance and advice of my committee members, Drs. Victor Willson, Patricia Lynch, and Lei-Shih Chen. Your suggestions and feedback over the years have helped shaped this work. I am grateful for your support and service on my committee.

I'd also like to thank my friends and colleagues who helped with the seemingly endless hours of data coding. Marcus Fuller, Sarah Ura, and Courtney Lavadia, I thank you for the time and effort you put into collecting inter-observer agreement data and meeting with me to discuss the data. I know it was grueling work at times, and yet you still agreed to help, even with your own commitments and deadlines to meet. Thank you. And a special thank you to April Haas who has helped with this dissertation from day one. You were there when I was implementing the single-case study and you continued to help all the way through data analysis for the meta-analysis. I am incredibly grateful for your help, and I look forward to many more future collaborations with you.

I'm also very grateful to Stephanie Gerow, Lisa Sanchez, Leslie Neely, Jennifer Ninci, Ee Rea Hong, and other special education doctoral students whom I am undoubtedly forgetting to mention. You taught me all the tips and tricks for surviving a doctoral program, helped me pass the dreaded stats classes, helped make conference experiences enjoyable ones, and provided me with opportunities to collaborate with you on research projects. In short, you ensured my time at Texas A&M was not only fun, but also memorable. Thank you all. I look forward to seeing you at conferences and continuing to collaborate on future writing projects.

Thank you also to all the educators in the single-case study, without whom this dissertation would not be complete. You invested a tremendous amount of time and energy into the single-case projects, and as a result, your instructional practices improved greatly. I hope you enjoyed the project as much as I did, and that you found it valuable. Watching your videos and seeing the tireless effort you put into improving the

lives of your students with special needs was inspiring. Thank you for your hard work and dedication. The world needs more educators like you.

Finally, a special thank you to the staff at the Office of Special Education Programs. My internship at the Department of Education was, without question, the best experience of my doctoral program. Everyone at OSEP went above and beyond to ensure my time there was exceptional, and that I felt included. You introduced me to eminent scholars and top-level officials, included me on federal writing projects, invited me to social gatherings, and made me feel like part of the team. I learned about educational policy, furthered my knowledge of research, and participated in all stages of the grant process. Thank you for your kindness and the time you invested in my education and career. I appreciate each of you, and I look forward to seeing you again at trainings and conferences.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Drs. Jennifer Ganz [chair], Kimberly Vannest [co-chair], Patricia Lynch, and Victor Wilson of the Department of Educational Psychology and Dr. Lei-Shih Chen of the Department of Health and Kinesiology.

All work for this dissertation was completed by the student, under the advisement of Drs. Jennifer Ganz and Kimberly Vannest of the Department of Educational Psychology.

Graduate study was supported by the *Preparation of Leaders in Autism Across the Lifespan* grant awarded by the U.S. Department of Education, Office of Special Education Programs (Grant No: H325D110046). This work was also supported in part by a Research Scholar Award awarded by the College of Education and Human Development at Texas A&M University.

The contents of this manuscript are solely the responsibility of the author and do not necessarily represent the official views of the Office of Special Education Programs or the College of Education and Human Development.

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

The U.S. National Center for Education Statistics (Gray & Taie, 2015) report 1 in 5 teachers leave the classroom within the first five years of teaching, contributing to a national crisis of teacher attrition (Shockley, Guglielmino, & Watlington, 2006) and creating a barrier to quality instruction (Alliance for Excellent Education, 2014). However, teacher attrition is not only a concern for beginning teachers; Borman and Dowling (2008) found that attrition is also a concern for experienced teachers. Given the exorbitant cost of teacher-attrition, any reasonable effort to reduce the number of teachers leaving the teaching profession each year could potentially save the United States billions of dollars annually (Alliance for Excellent Education, 2014).

While the causes of teacher attrition are varied and complex, one predictor that is strongly related to teacher attrition is pedagogical knowledge, or the knowledge of how to teach (Ingersoll, Merrill, & May, 2014). According to Ingersoll and colleagues (2014), teachers who received instruction in effective teaching strategies were significantly less likely to leave the classroom. In addition to decreasing rates of attrition, professional development opportunities provide additional benefits, such as increased self-efficacy (Ross & Bruce, 2007), a greater commitment to the profession (Billingsley, 2004), and an increase in student achievement (Corcoran, 2007; Wei, Darling-Hammond, & Adamson, 2010).

When planning professional development opportunities for educators, administrators have different models to choose from. The traditional model of professional development, which includes the one-stop workshop approach, is the most commonly used method of providing professional development in the United States (Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009). Unfortunately, it is also the most ineffective (Boudah, Blair, & Mitchell, 2003; Garet, Porter, Desimone, Birman, & Yoon, 2001; Harwell, 2003; Darling-Hammond et al., 2009; Wei, Darling-Hammond, & Adamson, 2010). There are many disadvantages to using this model of professional development, including low rates of implementation of learned skills (Boudah et al., 2003), the provision of information that is not directly translatable to practice (Penuel, Fishman, Yamaguchi, & Gallagher, 2007), minimal impact on student achievement (Darling-Hammond et al., 2009), and a sense of dissatisfaction among educators over the quality of professional development (Nir & Bogler, 2003; Quick, Holtzman, & Chaney, 2009).

The authentic model of professional development, on the other hand, provides educators with authentic learning opportunities based on their expressed needs and thus remediates many of the concerns that are inherent in the traditional model. For example, the authentic model of professional development (also termed the reform-based approach) provides educators with opportunities to practice what they have learned (Corcoran, 2007), which leads to higher rates of implementation fidelity (Boudah et al., 2003). In addition, the learned practices are integrated into educators' daily lives (Garet et al., 2001), which contributes to information that is more directly translatable to

practice. Also, because the authentic model of professional development is more effective (Boudah et al., 2003), it potentially has a higher impact on student achievement (Wei et al., 2010; Darling-Hammond et al., 2009). Lastly, the authentic model of professional development takes educators' needs into account when planning and implementing professional development opportunities (Boudah et al., 2003), which leads to a greater satisfaction among educators (Nir & Bogler, 2003).

Video analysis, a method of professional development aimed at improving one's teaching by analyzing self-recorded videos, includes many of the characteristics of the authentic model. For example, implementing video analysis requires that educators be videotaped while teaching, to reflect upon or analyze the events in the video, and to make changes based on their reflection or analysis (Nagro & Cornelius, 2013). Because educators are reflecting upon and analyzing self-identified areas for improvement in their own teaching, their needs are being considered, the practices are integrated into their daily lives, and the information learned is directly translatable to their practice (Tripp & Rich, 2012). In addition, video analysis is typically implemented over several sessions to document changes in educators' instructional practices, which provides educators with the opportunity to practice what they have learned (Hager, 2012). Lastly, because video analysis includes many of the characteristics of authentic professional development, it often leads to a high level of satisfaction among educators (Alexander, Williams, & Nelson, 2012) and has the potential to lead to an increase in student achievement.

While video analysis seems to be a viable method for providing professional development to educators, many questions still remain. For example, while many studies have documented the effectiveness of video analysis, most of them use qualitative methods or designs that are not experimental (Tripp & Rich, 2012a). In addition, while a systematic review has been completed on the topic (Nagro & Cornelius, 2012), single-case studies and dissertations were excluded from the review and effect sizes were not calculated. Thus, the purpose of this dissertation is to remediate these gaps in the literature by conducting a series of single-case studies designed to experimentally investigate the effects of video analysis on the instructional practices of educators (Study 1) and a meta-analysis designed to investigate the magnitudes of effect of video analysis and the impact of moderator variables (Study 2). The following research questions will be addressed in each of the two studies in this dissertation:

Study 1:

1. What are the effects of video analysis on the self-identified instructional practices of educators?
2. Will the results of video analysis on one instructional practice generalize to another instructional practice?
3. Will educators maintain improved rates of behavior over time?
4. Do educators find video analysis feasible, and do their views change over the course of implementation?

Study 2:

1. What is the status of the literature base on video analysis regarding study characteristics (i.e., publication type, design quality), participant characteristics (i.e., role, education level, experience level, and age), student characteristics (i.e., disability type and collection of student outcomes), and setting characteristics (i.e., grade level, group size, type of instruction, and setting)?
2. What effects do publication type and methodological quality have on the effectiveness of video analysis?
3. What is the magnitude of effect of video analysis on the instructional practices of educators?
4. What effects do participant characteristics (e.g., role, education level, experience level, age) have on the effectiveness of video analysis?
5. What effects do instructional characteristics (i.e., group size, type of instruction, grade level, setting) have on the effectiveness of video analysis?

CHAPTER II

THE USE OF VIDEO ANALYSIS TO IMPROVE SPECIAL EDUCATORS' INSTRUCTIONAL SKILLS

Teacher quality is a topic of concerned discussion, dating back nearly half a century (Hanushek, 1970). These discussions have influenced federal legislation (No Child Left Behind [NCLB], 2001; Every Student Succeeds Act [ESSA], 2015) and have led to an increased focus on teacher training and preparation (U.S. Department of Education, 2011). Despite a strong correlation between the quality of teacher professional development and the academic achievement of their students (Wei, Darling-Hammond, & Adamson, 2010), most professional development opportunities provided to teachers in the United States are inadequate (Borko, 2004; Darling-Hammond, Wei, Andree, Richardson, & Orphanos, 2009) and do not incorporate effective practices (Wei et al., 2010). For example, while the number of professional development opportunities has increased over the years, most of these opportunities are still provided in short-term workshops, which has little effect on teachers' practice (Wei et al., 2010). Delivering professional development via an ineffective, lecture-based model not only wastes valuable resources, such as time and money, but can also lead to dissatisfaction among teachers with the professional development opportunities they receive (Quick, Holtzman, & Chaney, 2009).

In contrast to the traditional, workshop-based approach to professional development, the reform-based approach (also termed authentic professional

development) includes practices that positively impact teacher behavior (Boudah, Blair, & Mitchell, 2003). Some of the salient features of the reformed-based approach include giving teachers an opportunity to practice what they have learned (i.e., active learning; Corcoran, 2007), integrating the practices being taught into the everyday lives of teachers (i.e., coherence; Garet, Porter, Desimone, Birman, & Yoon, 2001), and sustaining the professional development over longer periods of time (as opposed to one-time workshops; Harwell, 2003). In addition to these features, it is also important to take teachers' perspectives into account and plan the content and delivery of professional development with their needs in mind. Teachers desire professional development opportunities that align with their own goals (Penuel, Fishman, Yamaguchi, & Gallagher, 2007), are differentiated based on the needs of their students (Quick et al., 2009), involve them in the decision-making process (Nir & Bogler, 2008), and are delivered at their school site (Nir & Bogler, 2008). When professional development is planned and delivered in accordance with the salient features of the reform-based approach, and with the teachers' needs and goals in mind, it is more likely to be effective and have a positive impact on teacher behavior.

Video analysis, a method of evaluating one's own teaching (Nagro & Cornelius, 2014), has the potential to maximize the effectiveness of professional development efforts. For example, when implementing video analysis, educators evaluate their own teaching (active learning) within the context of the everyday activities in their classroom (coherence) and continue the process until they have mastered their goal (sustained over longer periods of time; Hager, 2012). In addition, teachers have the ability to select

target behaviors (i.e., teaching skills that they want to improve) that are aligned with their own goals and with the needs of their students (Alexander et al., 2012). Because teachers are self-selecting target behaviors, they are inherently involved in the decision-making process. Lastly, video analysis is implemented at the teachers' school sites, which teachers have identified as their preferred location for professional development (Nir & Bogler, 2008).

While video analysis has a fairly large literature base (Tripp & Rich, 2012), many prior studies used either qualitative or quasi-experimental designs and thus were unable to establish a functional relationship between video analysis and changes in teacher behavior. For example, both Alexander and colleagues (2012) and Hager (2012) conducted studies that investigated the use of video analysis with pre-service teachers and found positive effects; however, the designs used in both studies were quasi-experimental, single-case A-B designs (e.g., baseline and intervention) which are no longer considered rigorous enough to establish a functional relationship (Horner et al., 2005; Kazdin, 2011). While other studies (Aartman-Meeker & Hemmeter 2012; Vuran & Gul, 2012) did use an experimental design (i.e., multiple-baseline design; Kazdin, 2011), they involved a high level of researcher involvement which may lower the social validity and generalizability of the results. In addition, generalization and maintenance were only investigated in one study (Hager, 2012), and social validity was only investigated in two studies (Alexander et al., 2012; Vuran & Gul, 2012).

The purpose of this study was to experimentally investigate the effects of video analysis on the instructional practices of educators in schools and to extend prior

research by using an experimental design, calculating effect sizes, including social validity measures, and programming for generalization and maintenance. The specific research questions investigated in this study include the following:

1. What are the effects of video analysis on the self-identified instructional practices of educators?
2. Will the results of video analysis on one instructional practice generalize to another instructional practice?
3. Will educators maintain improved rates of behavior over time?
4. Do educators find video analysis feasible and do their views change over the course of implementation?

Method

Participant Selection

The participants in this study are sampled from 34 students enrolled in a 3-credit hour Master's level course in the Special Education Program of a university in the southern United States. The participants were part of a cohort model and this was the final semester prior to graduation. The Special Education Master's degree course sequence includes courses pre-approved by the Behavior Analyst Certification Board as meeting the coursework requirements for the Board Certified Behavior Analyst (BCBA) Certification and many of the students in this class were seeking BCBA certification. To promote clarity, the term "participant" will be used throughout the article to refer to the students in the Master's course who consented to be included in this study and the word

"student" will be used to refer to the students in the participants' classrooms to whom they were responsible for delivering instruction.

Participants were selected based on their full-time employment as a special education teacher (n = 4) or a special education paraprofessional (n = 1) at the time the study was conducted, their participation in either a multiple-baseline or changing criterion design, and subsequent consent after course completion. A total of 24 of the 36 participants consented; however, only 19 of the 24 had usable data (e.g., were either part of an intra-participant design or a multiple-baseline design where all participants consented). Of the 19 who had usable data, five met criteria for actively teaching in a special education setting and participation in a multiple-baseline or changing criterion design. IRB approval occurred prior to beginning the study; participants were blind to the study, with consent provided post-hoc through a third party after the submission of grades to avoid undue influence or the Hawthorne effect. Pseudonyms are used for all participants to protect confidentiality.

Course Description

The course was delivered completely online with weekly synchronous classes conducted on Blackboard Collaborate. The purpose of the course was to teach participants how to conduct single-case research through direct application. A major requirement of the course was for participants to conduct their own single-case research project using video analysis to change self-selected target behaviors. All phases of the study were implemented independently by the participants, including the collection of

baseline, intervention, generalization, and maintenance data, as well as inter-observer agreement and social validity data.

Participant, Setting, and Materials Description

Stephanie. Stephanie was a White female, between the ages of 30-39 years, who worked as a special education paraprofessional at a school in Germany. She had worked as a paraprofessional for the past 7 months and had a Bachelor's degree in Psychology. Stephanie had a total of three elementary children with disabilities on her caseload, but she only worked with one of those children, a Kindergarten boy with a learning disability, for the purpose of this study. All sessions were conducted in a special education resource classroom in a 1:1 teaching setting while Stephanie was teaching letter/sound recognition and beginning reading skills. Stephanie used an iPhone to record all videos. The instructional materials she used during the teaching sessions included worksheets and flashcards with upper and lowercase letters printed on them.

Crystal. Crystal was a White female, between the ages of 30-39 years, who worked as social behavior skills teacher at a school in the southern United States. She had worked in this position for the past 5 years and had a Bachelor's degree in Multidisciplinary Studies. Crystal had a total of six secondary students with disabilities on her caseload, but she only worked with five of these students for the purpose of this study. All five students, one girl and four boys, were diagnosed with emotional/behavioral disorder and ranged in age from 12 to 15 years. All sessions were conducted in the special education classroom in either a small group or a 1:1 teaching arrangement. During these sessions, Crystal taught the student(s) various social skills,

such as anger management, problem solving, strategies for working cooperatively, and skills for reflecting on behavior. Crystal used a flip video recorder to record all sessions. The instructional materials she used during the teaching sessions included handouts for students to follow along and take notes.

Mary Anne. Mary Anne was a White female, between the ages of 30-39 years, who worked as a special education inclusion teacher at a school in the southern United States. She had worked as a special education teacher for the past 9 years and had a Bachelor's degree in General Education with an endorsement in Special Education. Mary Anne had a total of 14 elementary students on her caseload, but she only worked with 6 students for the purpose of this study. Five of the students were 6 years old and diagnosed with autism spectrum disorder, while the sixth student was 8-years old and diagnosed with Down Syndrome. All sessions were conducted at a teaching table in a resource classroom in a small group teaching arrangement. During these sessions, Mary Anne taught the students beginning reading skills. Mary Anne used either a digital camera or a laptop computer with a webcam to record all sessions. The instructional materials she used during the teaching sessions included a packaged reading curriculum, worksheets, and writing materials.

Pamela. Pamela was a White female, between the ages of 30-39 years old, who worked as a special education resource teacher at a school in the southern United States. This was Pamela's first year working as a special education teacher; she had a Bachelor's degree in Psychology. Pamela had a total of 10 elementary students with disabilities on her caseload, but she only worked with two of these students, both first-

graders, for the purpose of this study. All sessions were conducted in a resource classroom in a small group teaching arrangement. During these sessions, Pamela taught the students beginning reading or math fluency skills. Pamela used a laptop with a webcam to record all sessions. The instructional materials she used during the teaching sessions included an iPad, books, matching cards, and other manipulatives.

Angela. Angela was a White female, between the ages of 18-29 years, who worked as a Direct Support Professional Content Teacher in a post-secondary education setting for individuals with disabilities in the southern United States. This was Angela's first year working in this position; she had a Bachelor's degree in Special Education. Angela had 17 students, 10 males and 7 females, with disabilities in her classroom. The students had a variety of disabilities, including autism spectrum disorder (n=1), intellectual disability (n=3), learning disability (n=7), attention deficit disorder (n=1), physical disability (n=2), visual impairment (n=2), and Down Syndrome (n=1). All sessions were conducted in a university classroom in a large group teaching arrangement. Angela used either an iPad or a MacBook Pro with a webcam to record all sessions. The instructional materials she used during the teaching sessions included a computer and projector to display PowerPoint presentations.

Investigator

The study was conducted by a third-year doctoral student in special education with seven years of experience as a classroom teacher and behavior therapist. She also had three years of experience designing, implementing, and participating in single-case research. The instructor of record for the course was a Professor in the Special Education

Division, within the Department of Educational Psychology. She has authored or co-authored nearly 100 scholarly products and dozens of manuscripts and books in single-case research design, reviews single-case methods for the Institute of Education Sciences and the National Science Foundation, and has taught in public and higher education for three decades.

Research Design

The participants in this study were assigned to either a changing criterion or multiple-baseline design across participants (Kazdin, 2011; Kennedy, 2005) in order to investigate the effects of video analysis on self-selected target behaviors. The participants also collected generalization and maintenance data to investigate whether the effects of video analysis would maintain over time and whether they would generalize to a second target behavior.

With a changing criterion design, subphases with different criterion standards are set in intervention and experimental control is demonstrated when performance corresponds closely to the shifts in criterion (Kazdin, 2011). While some interventions deliver reinforcement to participants as they reach the different criteria (Kazdin, 2011), other times, particularly when used with a self-monitoring intervention as is the case in this study, reaching a criterion itself is reinforcing and extraneous reinforcement is not used (Klein, Houlihan, Vincent, & Panahon, 2017). Because changing criterion designs can maintain adequate experimental control with a variety of populations, target behaviors, and settings, they are ideal to use with practitioners (Klein et al., 2017). Additionally, the use of changing criteria to incrementally change behavior makes the

changing criterion design particularly well suited for changing behavior that may be otherwise resistant to large, immediate changes, such as habits (Klein et al., 2017). For these reasons, a changing criterion design was an ideal choice to answer the research questions presented in this study.

While there are procedural variations that can be implemented to increase the rigor of changing criterion designs, such as using multiple criterion changes and different phase lengths, both of which were implemented in this study, changing criterion designs are still not considered to be as experimentally robust as a multiple-baseline design (Klein et al., 2017). Therefore, we also included a multiple-baseline design in this study to further experimentally demonstrate the effects of video analysis on the instructional practices of educators. In contrast to a changing criterion design, multiple-baseline designs involve introducing the independent variable sequentially to different baselines (Kennedy, 2005). These baselines can be across different settings, people, or behaviors, and experimental control is demonstrated by showing that a change in the dependent variable occurs only when the intervention is introduced (Kazdin, 2011).

Dependent Variables

Participants selected primary target behaviors (i.e., dependent variables) for reflection and improvement after viewing their self-recorded baseline videos and engaging in an instructional discussion and reading about the evidence to support four educational practices (i.e., rates of praise, opportunities to respond, and higher order questions, and wait time). Participants selected target behaviors independently and were

not required to choose one of the educational practices that was introduced during the instructional discussion if it was not a behavior that the participants needed to improve. After the target behaviors were selected, participants submitted these to the investigator for approval. Initial videos of participants were viewed by the investigator to confirm the target behavior chosen had occurred at low rates.

Each participant chose to improve the use of specific praise as their primary target behavior. Secondary target behaviors, which were selected to assess generalization during the generalization phase, included opportunities to respond (Crystal & Pamela) follow-through (Stephanie), reinforcement (Mary Anne), and fidgeting (Angela). Secondary target behaviors were selected using the same procedures as the primary targets. Operational definitions for both the primary and secondary target behaviors originated with the participants and the investigator validated the definitions prior to the participants engaging in data collection.

Specific praise. While specific praise was operationally defined by each participant, it was generally defined as a verbal comment of approval immediately followed by a statement that describes the behavior that earned the praise (e.g., "thank you for raising your hand," "nice try saying the letter sound," etc.).

Opportunities to respond. Crystal defined opportunities to respond as asking a question that required a verbal or physical response, while Pamela defined it as questions asked to show understanding or ask for clarification (e.g., "Who was in the book?", "Can you tell me…").

Follow-through. Stephanie defined follow-through in the following way: after modeling a task or giving explicit directions, the teacher will ensure the student completes, or attempts to complete with effort, the desired task prior to giving the next direction (e.g., when asking the student a question, the teacher does not move on until the student answers the question).

Reinforcement. Mary Anne defined reinforcement as a high five, thumbs up, or sticker that is delivered when providing specific praise.

Fidgeting. Angela defined fidgeting as (a) touching, twisting, or taking her rings on and off (excluding times when she would touch her rings incidentally due to her hands being folded or clasped together); (b) touching, readjusting, or sliding her necklace back and forth; or (c) touching or twisting her Fitbit (excluding times when she would tap the face of the Fitbit to check the time).

Measurement and Dosage

All sessions were 8-minutes in length, but the dosage was different, depending on the design. Participants in the multiple-baseline design recorded no more than two 8-minute videos per day, though they were encouraged to record no more than one video per day whenever possible. Participants in the changing criterion design recorded 4-5 videos per day. These schedules were variable to allow for flexibility due to absences and availability. Each 8-minute video equated one session.

Data were collected as either frequency (i.e., specific praise, opportunities to respond, reinforcement, and fidgeting) or percent of opportunities (i.e., follow-through). For frequency, participants counted the number of instances the target behavior occurred

in an 8-minute period. For percent of opportunities, Stephanie collected data on the number of times directions were given in an 8-minute period and then placed a plus (+) or minus (-) sign next to the direction depending on whether or not she followed through according to her operational definition. She calculated the data by dividing the number of times she did follow-through by the total number of times she did and did not follow-through and multiplied that number by 100 to obtain a percentage.

Procedures

Baseline. The participants were blind to the purpose of the study during baseline. The investigator directed the participants to take video of themselves teaching a student or group of students in their classroom. The participants chose a relevant skill to teach their student(s), but the investigator asked that they keep the skill they chose to teach fairly consistent throughout the videos. No other directions were given.

Primary target behavior. After the participants watched their baseline videos and selected their primary target behaviors, they graphed and analyzed their baseline data to determine an appropriate individualized goal for intervention. For Stephanie, Crystal, and Mary Anne, their goals were 20, 8, and 8 instances of specific praise in an 8-minute video, respectively. Both Pamela and Angela were in changing criterion designs and had multiple goals to represent the different criteria in the designs. Pamela's goals were 4, 6, 7, and 8 and Angela's goals were 1, 3, 5, and 6 instances of specific praise in an 8-minute video for criterions 1, 2, 3, and 4, respectively.

Intervention. Within 24 hours of recording a video in intervention, the participants watched it, collected data on their target behaviors, graphed their data, and

completed a reflection sheet. All of these steps were completed prior to recording the next day's video (with the exception of instances where multiple videos were recorded in a day). The reflection questions asked participants to (a) state whether or not they met their goals, (b) describe something that went well, (c) describe any challenges or areas for change, and (d) explain what could be done differently next time to help meet their goals (if applicable).

Generalization and Maintenance. After all baseline and initial intervention phases were complete, participants viewed their baseline videos to identify a second target behavior, with guidance from the investigator. Once the target behavior was identified and approved, participants coded five baseline videos for the second target behavior. Participants then followed the same procedures as in intervention for an additional five intervention videos, with the exception that they now took data on two target behaviors rather than one. After a minimum of one week following the collection of the last generalization video, participants collected three maintenance videos. Procedures for maintenance were identical to baseline in that participants did not view or code the videos until after all three videos were collected.

Social Validity

A social validity survey was administered online at five points in time to determine if participants' views changed throughout the study with regard to video analysis. The survey was administered immediately after baseline videos were collected (but prior to participants watching them), at the beginning and end of intervention, during maintenance, and after generalization. The survey included 4 open-ended

questions and 10 multiple-choice statements that required the participants to answer on a 5-point Likert scale to indicate the degree to which they agreed with the statement (5 = strongly agree, 4 = agree, 3 = unsure/neutral, 2 = disagree, 1 = strongly disagree). The multiple-choice statements and the participants' responses to them can be found in Appendix C (see Table C1), while the open-ended questions were as follows: (a) please provide information about changes you would make to the procedures for implementing video analysis or practices you would recommend stay the same, (b) please comment on the feasibility of implementing video analysis in your setting, (c) please comment on the advantages and disadvantages of using video analysis, and (d) please provide any additional comments.

Data Analysis

The data were analyzed using both visual and statistical analyses. Visually, the data were analyzed by evaluating changes in trend, level, and variability to determine if a functional relation was present between the intervention and dependent variables (Horner et al., 2005); statistically, the data were analyzed by conducting Tau-U analyses (Parker, Vannest, Davis, & Sauber, 2011), using a free, online calculator (www.singlecaseresearch.org), to complement the visual analyses. Tau-U (Parker. Vannest, & Davis, 2011) is a non-parametric effect size that is used in single-case research and offers several advantages over other non-parametric effect sizes, including the use of all data points, the ability to control for trend, high sensitivity, and ease of calculation. Tau-U is the "percentage of nonoverlap versus overlap" and effect sizes

range from −1.0 to 1.0, with positive scores indicating improvement and negative scores indicating deterioration of the data (Parker et al., 2011).

Generally, Tau-U can be roughly interpreted as follows when comparing baseline to intervention: small effect = 0 to .62; medium effect = .63 to .92; large effect = .93 to 1.00 (Parker, Vannest, & Davis, 2011); however, these interpretive guidelines do not apply when comparing intervention and maintenance data. Because the goal of maintenance is for the target behavior to maintain at the same levels as in intervention, a Tau-U score of 0 would be considered a positive finding as it indicates there was 100% overlap of the data and the behavior maintained completely. Any score above 0 would indicate the behavior improved in maintenance, and a score below 0 would indicate there was a loss of skills during maintenance.

When entering the data into the calculator, data were entered in reverse when the goal was to decrease the behavior. In other words, the intervention data were entered as the "A" phase and baseline data were entered as the "B" phase. Additionally, all results are weighted and reflect corrected baseline data, with the exception of analyses that compared intervention to maintenance phases, as an increasing trend during the "A" phase (i.e., intervention) is expected and desired. To aid in the interpretation of the data, a forest plot is presented in Appendix A (see Figure A1) to visually display the results of the effect size analyses. These analyses were used to complement the visual analysis in interpreting the effects of video analysis on the instructional skills of educators.

Inter-Observer Agreement

All participants collected inter-observer agreement (IOA) data for a peer for a minimum of 20% of sessions in each phase and for each dependent variable. The participants calculated IOA as total count IOA (Cooper, Heron, & Heward, 2007), meaning that they compared the number of instances of behavior recorded by one observer to the number of instances of behavior recorded by the second observer and divided the smaller count by the larger count and multiplied the result by 100 to obtain a percentage.

For the primary target behavior, the mean IOA was 88% (range 0-100) in baseline, 82% (range 44-100) in intervention, and 95% (range 80-100) in maintenance. For the secondary target behavior, the mean IOA was 93% (range 85-100) in baseline, 81% (range 60-100) in intervention, and 80% (range 67-100) in maintenance. Because the participants calculated IOA as total count rather than dividing the 8-minute observation period into equal intervals and collecting interval-by-interval data (Cooper et al., 2007), their IOA was sometimes low, especially when the behavior they were collecting data on occurred infrequently, such as in baseline. For example, Mary Anne's IOA was 0% for one video in baseline. This is because she scored one instance of the target behavior occurring and her peer scored zero instances of the target behavior occurring, leading to an IOA score of 0%. Had the 8-minute observation period been divided into 16 30-second intervals, then only one of those intervals would have been a disagreement and the IOA for that observation period would have been 94% (15/16 x 100) rather than 0%.

Fidelity of Implementation

To determine if participants followed the study procedures, their fidelity of implementation was assessed during all phases of the study using a checklist. While the items on the checklist varied depending on the design and phase of the study, the following items were the same for all phases and designs: (a) the required number of videos were recorded, (b) videos were a minimum of 8 minutes in length, and (c) data were graphed and submitted by the due date. In baseline and maintenance phases, the following criterion was also included: data collection and reflection sheets were not completed until all videos had been recorded. For intervention and generalization phases, the following criteria were included: no more than two videos were uploaded within a 24-hour time period (for participants in the multiple baseline design), no more than five videos were uploaded within a 24-hour time period (for participants in the changing criterion design), and data collection and reflection sheets were completed within 24 hours of uploading the video(s).

The results were calculated as the number of steps implemented correctly divided by the total number of steps and multiplied by 100 to obtain a percentage. Fidelity of implementation was calculated on 100% of the data for each phase and each participant. The mean fidelity of implementation score was 92% (range 82%-100%) for Stephanie, 98% (range 95%-100%) for Crystal, 98% (range 95%-100%) for Mary Anne, 98% (range 91%-100%) for Pamela, and 98% (93%-100%) for Angela.

Results

This study used a multiple-baseline across participants design and two changing criterion designs to answer research questions related to the overall effects of video analysis, whether the results of video analysis on one instructional practice would generalize to another instructional practice, whether educators would maintain their improved rates of behavior over time, whether educators would find video analysis feasible, and whether educators' views of video analysis would change over the course of implementation. These questions were answered using a traditional visual analysis of the data in terms of trend, level, and variability (Horner et al., 2005), as well as a statistical analysis of the data using Tau-U, a non-parametric effect size (Parker et al., 2011).

Effects of Video Analysis

The first research question was, "What are the effects of video analysis on the self-identified instructional practices of educators?" A visual analysis of the primary target behavior demonstrates a functional relation and an increase in level over baseline for all participants. When aggregating the Tau-U scores for all participants for the primary target behavior, the omnibus effect size is 0.93, 90% CI [0.72, 1.00], which indicates strong effects (see Appendix A, Figure A1). This complements the visual analysis, which indicates positive effects for all participants, with strong effects for four of the five participants and moderate effects for the fifth participant.

Stephanie. A visual analysis of the data in Appendix B (see Figure B1) indicates an increasing trend in both baseline and intervention for Stephanie's primary target

behavior (i.e., specific praise). There is a significant change in level from baseline (*M* = 4.2) to intervention (*M* = 16.3), and while the level decreased slightly when the secondary target behavior was introduced (*M* = 11.2), it is still elevated over baseline. The range of scores increased from 1-10 in baseline to 4-27 in intervention, indicating a fair amount of variability in intervention. The Tau score for Stephanie's primary target behavior is 0.92, 90% CI [0.44, 1.00]. When the trend in baseline is corrected, the Tau-U score drops slightly to 0.86, 90% CI [0.38, 1.00]; however, both scores indicate a moderate effect.

Crystal. There is an immediate change in level from baseline (*M* = 0.6) to intervention (*M* = 12.7) for Crystal's primary target behavior (i.e., specific praise), with no overlap of data (see Appendix B, Figure B1), as indicated by a Tau-U score of 1.00, 90% CI [0.59, 1.00]. While there is a decreasing trend and slight change in level for the primary target behavior in the second intervention phase when the secondary target behavior is introduced (*M* = 10), the level is still significantly increased over baseline (*M* = 0.6). There is a change in variability from baseline (range 0-2) to intervention (range 6-17); however, much of this is due to low levels of responding in baseline and an increasing trend in intervention. Overall, both a visual and statistical analysis of Crystal's data for her primary target behavior indicate strong effects.

Mary Anne. Mary Anne's data for her primary target behavior (i.e., specific praise) also has an immediate change in level from baseline (*M* = 0.4) to intervention (*M* = 7.8), with no overlap of data (see Appendix B, Figure B1), as indicated by a Tau score of 1.00, 90% CI [0.61, 1.00]. When the slight increasing trend in baseline is corrected,

the Tau-U score drops negligibly to 0.98, 90% [0.58, 1.00]; however, the data still indicate strong effects. While there is a minor decreasing trend towards the end of the first intervention phase, the data increase in the second intervention phase when the secondary target behavior is introduced, with the level increasing above the level of the first intervention phase (M = 7.5 for the first intervention phase; M = 8.8 for the second intervention phase). The range of data increase somewhat from baseline (range 0-2) to intervention (range 5-12), but overall, the data are fairly stable.

Pamela. Pamela's average response for her primary target behavior (i.e., specific praise) was 2.2 in baseline. Her first criterion was four praise statements, which she consistently met for all four data points (see Appendix B, Figure B2). Next, she set a criterion of six praise statements, which she met for three of the four data points in this phase (M = 5.75). Her third criterion was seven praise statements, which she met or exceeded for three of the four data points in this phase (M = 7). Pamela's last criterion was eight praise statements, which she met or exceeded for all five data points in this phase (M = 9). When the secondary target behavior was introduced, Pamela's level of responding for specific praise dropped slightly (M = 7.0); however, it was still elevated above baseline (M = 2.2). The Tau-U score for Pamela's primary target behavior indicates strong effects (1.00, 90% CI [0.56, 1.00]), confirming the strong effects found in the visual analysis.

Angela. Angela had a low level of responding in baseline for her primary target behavior (i.e., specific praise; M = 0.6). Angela's first criterion was one praise statement (see Appendix B, Figure B3), which she met or exceeded for three out of the four data

points in this phase (M = 1.75). Next, Angela set a criterion of three praise statements, which she met or exceeded for all five data points in this phase (M = 4.6). Angela's third criterion was five praise statements, which she met or exceeded for only two out of the five data points in this phase (M = 4.6). Lastly, Angela set a criterion of six praise statements, which she met or exceeded only once during this phase (M = 4.2). When the secondary target behavior was introduced, Angela's level of responding decreased further and was only slightly above baseline levels (M = 0.8 in the second intervention; M = 0.6 in baseline). The Tau-U score for Angela indicates moderate effects (0.72, 90% CI [0.24, 1.00]).

Generalization

The second research question was, "Will the results of video analysis on one instructional practice generalize to another instructional practice?" A visual analysis of the data shows a change in level for all five participants, indicating that video analysis was effective for changing a secondary target behavior. The data became more stable in intervention for two of the five participants, while another two participants increased their responding from zero levels in baseline, thereby naturally increasing the variability of their data in intervention; the final participant had no change in variability from baseline to intervention, although there was an increase in level. When aggregating the Tau-U scores for all participants, the omnibus effect size is 0.79, 90% CI [0.51, 1.00], which indicates moderate effects. This complements the visual analysis which also demonstrated moderate to strong effects for all participants.

Stephanie. A visual analysis of the data indicates an increasing trend in both baseline and intervention for Stephanie's secondary target behavior (i.e., follow through). Her average responses increased from a mean of 63% in baseline to 81% in intervention, indicating a change in level. Additionally, the variability of responses narrowed from baseline (range 50-85) to intervention (range 70-90). The uncorrected Tau score for Stephanie's secondary target behavior is 0.68, 90% CI [0.05, 1.00]; when correcting for trend in baseline, the Tau-U score drops to 0.40, 90% CI [-0.23, 1.00], indicating small effects.

Crystal. Crystal also had a slight increasing trend in baseline for her secondary target behavior (i.e., opportunities to respond). There is an immediate change in level from baseline (M = 7) to intervention (M = 13) with no overlap of data, as indicated by a Tau score of 1.00, 90% CI [0.37, 1.00]. When correcting for trend in baseline, the Tau-U score decreases somewhat to 0.80, 90% CI [0.17, 1.00], but still demonstrates moderate effects. The variability of data are similar in both baseline (range 5-10) and intervention (range 11-16).

Mary Anne. There was an immediate increase in level for Mary Anne's secondary target behavior (i.e., reinforcement), from 0 responses in baseline to an average of 6.0 responses in intervention, with no overlap of data, as confirmed by a Tau-U score of 1.00, 90% CI [0.37, 1.00]. In intervention, Mary Anne's responses ranged from 4-10 responses, indicating some variability.

Pamela. Pamela's responses for her secondary target behavior (i.e., opportunities to respond) increased from a 0 level of responding in baseline to an average of 5.8

responses in intervention. There is a high level of variability, as indicated by a range of 0-10 responses in intervention, and an increasing trend present in the intervention data. The Tau-U score for Pamela's secondary target behavior demonstrates moderate effects at 0.80, 90% CI [0.17, 1.00].

Angela. The goal of Angela's secondary target behavior was to decrease the frequency of fidgeting. While there is a slight decreasing trend in baseline, there is a significant change in level from baseline (M = 8.2) to intervention (M = 0.6), as well as a much narrower range of data (range 2-16 in baseline; range 0-2 in intervention), indicating that the data stabilized in intervention. The strong effects that are demonstrated by a visual analysis of the data are confirmed by a statistical analysis (Tau-U 0.96, 90% CI [0.33, 1.00]).

Maintenance

The third research question was, "Will educators maintain improved rates of behavior over time?". A visual analysis of the data shows the primary target behavior was maintained for four of the five participants (i.e., Stephanie, Crystal, Mary Anne, & Pamela), while the secondary target behavior maintained for all five participants. For the primary target behavior, the overall level in maintenance decreased slightly for two participants, increased for two participants, and returned to a zero level for the fifth participant. For the secondary target behavior, the overall level in maintenance improved over intervention levels for three participants and decreased slightly for another participant, although the level was still elevated over baseline levels. For the fifth participant, the level of responding for her secondary target behavior decreased to zero

during maintenance; however, this was a positive finding given that her goal was to decrease this behavior.

An overall statistical analysis of the data revealed similar results. When aggregating the Tau scores for all participants, the omnibus effect size for intervention compared to maintenance for the primary target behavior was -0.09, 90% CI [-0.36, 0.18], indicating a very slight loss of skills during maintenance; overall, however, it indicates the skills maintained across participants. This finding is confirmed by a Tau score of 0.70, 90% CI [0.39, 1.00] when comparing baseline to maintenance, indicating moderate effects. For the secondary target behavior, the omnibus effect size for intervention compared to maintenance is 0.33, 90% CI [0.00, 0.66], indicating an increase in skills from intervention to maintenance. This finding is complemented by an omnibus Tau score of 1.00, 90% CI [0.67, 1.00] when comparing baseline to maintenance for the secondary target behavior.

Stephanie. A visual analysis of the data for the primary target behavior for Stephanie (i.e., specific praise) indicated that, while the behavior slightly decreased in maintenance, it remained at a comparable level to intervention. When considering level, the mean dropped slightly from 15.2 in intervention to 11.7 in maintenance; however, the mean is still considerably higher than in baseline (M = 4.2, range 1-10). Consistent with the slight decrease in level, the Tau score also showed a slight decrease from intervention to maintenance (Tau-U = -0.47, 90% CI [-1.00, 0.13]); again, an analysis of the data in baseline compared to maintenance indicate an improvement over baseline (Tau-U = 0.47, 90% CI [-0.27, 1.00]). While the range of data is narrower in

maintenance (range 10-15) than in intervention (range 4-27), indicating more stable data, this conclusion must be interpreted with caution considering the short data series in maintenance. For Stephanie's secondary target behavior (i.e., follow-through), there was an increasing trend in intervention and an immediate increase in level in maintenance that was maintained at 100% for all three data points. When considering level, the mean increased from 81.4% (range 70%-90%) in intervention to 100% in maintenance. This is an increase from 63.4% (range 50%-85%) in baseline. A statistical analysis of the data from intervention to maintenance for the secondary target behavior results in a Tau-U score of 1.00, 90% CI [0.26, 1.00].

Crystal. A visual analysis of the data for the primary target behavior for Crystal (i.e., specific praise) indicate that the behavior maintained after the conclusion of intervention. There is complete overlap in the data from intervention to maintenance, as confirmed by a Tau-U score of -0.13%, 90% CI [-0.73, 0.47]. When considering level, average responses changed slightly from 12.0 in intervention to 11.0 in maintenance. The range is similar in both intervention (range 6-17) and maintenance (range 6-16), indicating no change in variability. For the secondary target behavior (i.e., opportunities to respond), an analysis of the data also indicate that the behavior maintained. When considering level, the average response increased slightly from 13.0 in intervention to 13.7 in maintenance. The range narrowed from intervention (range 11-16) to maintenance (range 13-14), indicating the data stabilized. Comparing intervention to maintenance, the Tau-U score for the secondary target behavior is 0.47, 90% CI [-0.27, 1.00], indicating the behavior improved in maintenance.

Mary Anne. When comparing Mary Anne's data from intervention to maintenance, a visual analysis indicates the primary target behavior (i.e., specific praise) maintained at the conclusion of intervention. There is complete overlap of data from intervention to maintenance, with a slight increase in level (M = 7.8 in intervention; M = 8.0 in maintenance). The range of scores indicate comparable variability when comparing intervention to maintenance (range 5-10 in intervention; range 7-9 in maintenance). The Tau-U score for the primary target behavior when comparing intervention to maintenance is 0.15, 90% CI [-0.46, 0.75]. For the secondary target behavior (i.e., reinforcement), the data also indicate it maintained after intervention; however, there was a slight decrease in level (M = 6.0 in intervention; M = 4.3 in maintenance), which is consistent with the Tau-U score of -0.40, 90% CI [-1.00, 0.34] . Despite the decrease in level, the data are still elevated above baseline, which was consistently zero. The range of data indicate that the data were slightly less variable in maintenance (range 4-10 in intervention; range 2-7 in maintenance); however, the short data series in maintenance makes it difficult to analyze variability.

Pamela. When visually analyzing the primary target behavior for Pamela (i.e., specific praise), the data indicate that not only did the behavior maintain after the conclusion of intervention, it increased over intervention levels. These findings are confirmed by analyzing the average response in intervention (M = 6.7) and maintenance (M = 10.7), as well as the Tau score (Tau = 0.91, 90% CI [0.31, 1.00]). Pamela's secondary target behavior (i.e., opportunities to respond) also maintained after intervention. There is complete overlap in data, as indicated by a Tau score of 0.02, 90%

CI [-0.54, 0.94]. When analyzing level, Pamela's response increased from an average of 5.8 in intervention to 7.0 in maintenance. The variability also stabilized in maintenance, as indicated by a tighter range of responses (range 0-10 in intervention; range 6-8 in maintenance).

Angela. Angela's primary target behavior (i.e., specific praise) did not maintain after intervention, as her level of responding decreased from an average of 3.3 responses in intervention to 0 responses in maintenance. The lack of maintenance for the primary target behavior is confirmed by a Tau score of -0.88, 90% CI [-1.00, 0.28]. Angela's secondary target behavior (i.e., fidgeting), however, not only maintained but further decreased to a zero level of responding. Because the goal was to decrease fidgeting, a decrease from a mean of 0.6 responses in intervention to 0 responses in maintenance was a positive result. The Tau score for Angela's secondary target behavior, which was affected by a floor effect, was 0.40, 90% CI [-0.34, 1.00].

Social Validity

The fourth research question was, "Do educators find video analysis feasible and do their views change over the course of implementation?" The results are in Appendix C (see Table C1). To answer the first part of the question, "Do teachers find video analysis feasible?" the overall results of the survey were analyzed. The average response was a 4.0 or higher on six out of the 10 questions, indicating the participants were satisfied with the intervention. The six statements that participants agreed or strongly agreed with relate to the usefulness, cost-effectiveness, ease of implementation, and beneficial aspects of video analysis. Statements for which the participants felt unsure or

neutral related to the feasibility of video analysis, the time and effort required, the degree to which participants minded watching themselves on video, and whether or not they intended to continue using video analysis after the project was complete.

To answer the second part of the research question, "Do educators' views change over the course of implementation?", the teachers' responses on the survey were compared across all five points in time. The results indicate that the participants viewed video analysis more favorably as the study progressed. For every question, the mean response increased at the last administration of the survey from the first administration. In some instances, the response increased an entire point or more (i.e., Video analysis is a cost-effective way to improve my teaching skills; I was able to implement video analysis without much assistance from others; and Video analysis is worth the time invested). According to the results of a paired t-test, the differences between the scores from the first administration of the social validity survey (i.e., Time 1) to the last administration (i.e., Time 5) were statistically significant for two statements—Video analysis is a cost-effective way to improve my teaching skills and Video analysis is worth the time invested, both of which produced the following results: $t(4) = 3.162$, $p = 0.034$.

Discussion

The first research question investigated the effects of video analysis on the self-identified instructional practices of educators. While all participants increased their levels of responding during intervention, some participants showed stronger effects than others. Although Angela did increase her levels of responding for specific praise during

most of the intervention, she did not meet her goals during criterion three or four and her level of responding returned to near baseline levels during the last intervention phase when the secondary target behavior was introduced. One rival explanation for the lack of effect is that her class was a discussion-based class with a large amount of teacher-lecture, precluding frequent opportunities for specific praise. For example, during one synchronous class, Angela reported the reason she did not meet her goal for the week was because one student dominated the class discussion by talking for long periods of time, which limited the number of times she could deliver specific praise. Additionally, during the last five data points in intervention (i.e., the phase when the secondary target behavior was introduced), Angela noted on her reflection sheet that the reason she had such low levels of specific praise was because this class period was devoted to discussing the details of a final class project, thereby necessitating a large amount of teacher instructions and fewer opportunities for the students to participate and earn specific praise.

Because Angela's class only met once per week, she had to take all of her videos for the week during one class period; thus, if she had limited opportunities to provide specific praise, it affected multiple data points, a point which Angela commented on several times during the different administrations of the social validity survey. Despite Angela's unique situation, the positive effects of video analysis were demonstrated through the complete reduction of her secondary target behavior (i.e., fidgeting) which was more under her control and less susceptible to the changing demands of her classroom. In addition, the other four participants, who were able to record video daily,

had an immediate increase in their skills when intervention was introduced, thereby demonstrating that video analysis is effective in many cases.

The second research question investigated whether the results of video analysis would generalize to another instructional practice. A visual analysis of the data indicated the secondary target behavior improved over baseline for all participants, although the full effects of the intervention were not seen for Stephanie until maintenance. Stephanie had an increasing trend and fairly high levels of responding for her secondary target behavior (i.e., follow through) in baseline, which impacted the amount of improvement that could be shown in intervention. Despite the small improvement that was made in intervention, Stephanie was able to increase her follow through to levels well above baseline during the maintenance phase.

One unexpected finding during generalization was a decrease in level for the primary target behavior when the secondary target behavior was introduced for four of the five participants. While this could be due to difficulty trying to focus on improving multiple behaviors at once while teaching, as some participants mentioned during synchronous classes, for some participants it was because increasing one target behavior naturally decreased responding for the secondary target behavior. For example, Stephanie noted that in order to improve follow-through, she had to allow her students more wait time to respond, which decreased the amount of opportunities she had to provide specific praise. In Pamela's case, her target behaviors were complementary and the frequency with which both target behaviors occurred was almost identical for most of the data points in generalization. This is because every time she provided her students

with an opportunity to respond, she also delivered specific praise contingent upon their response. Similarly, Mary Anne's secondary target behavior was tangible reinforcement, which was delivered at the same time as specific praise; as a result, her level of delivering specific praise did not decrease during generalization.

The third research question investigated whether participants' behavior would maintain over time. A visual analysis of the data indicates that, for the majority of the participants, their skills maintained in the absence of intervention. An analysis of the omnibus Tau scores confirm that the primary target behavior decreased only slightly during maintenance and that, overall, the secondary target behavior actually increased over intervention levels. One exception is Angela's primary target behavior, specific praise, which did not maintain. The fact that the three data points in intervention reflect one teaching session for Angela likely impacted her results. In response to her low levels of responding for her primary target behavior during maintenance, Angela wrote the following on her reflection survey: "…I found myself struggling to come up with a specific praise statement quickly that would make sense and actually sound like praise. In this video, I missed two opportunities to give specific praise. I also had a few students give long winded answers, so my opportunities to give specific praise were even more limited." While Angela struggled to provide specific praise statements during this class period, it is encouraging that she noticed missed opportunities. Had Angela been able to record multiple videos a week rather than taking all of her maintenance videos during one class session, she likely would have been able to improve delivery of specific praise in future class periods, a point that she made on every administration of the social

validity survey.

The last research question investigated the feasibility and acceptability of video analysis and whether participants' views changed over the course of implementation. Participants' responses to the Likert-type statements indicated that, overall, they found video analysis to be beneficial and a useful tool for improving their instructional behaviors. While the responses the participants gave to the open-ended questions were largely positive, they also addressed several areas of concern for their particular situations. Angela mentioned several times that she wished she could record videos several times a week rather than taking all of her videos on one day as she felt that taking multiple videos in one day negatively impacted her data if she did not have many opportunities to provide specific praise that day. Additionally, while some participants felt that video analysis was fairly easy to implement in their settings, other participants did not have the same experience. This was particularly true for Crystal, as she taught students with emotional and behavioral disorders, and she often commented that she worried her students would try to break her camera when they became agitated during class. While she eventually decided to hold the camera to protect it from her students, she mentioned this was not easy and recording would have been easier in her situation if she had a paraprofessional who could record for her. Another topic that came up frequently in participants' narrative responses was the length of the videos. While three of the five participants felt that eight minutes was an appropriate amount of time to record their teaching, Pamela felt that eight minutes was too long as she mentioned it was hard to record for eight minutes straight without interruption.

When asked to discuss the advantages and disadvantages of using video analysis, every participant mentioned the ability to see yourself and your teaching on video was an advantage. As Mary Anne stated, "[It] gives me a true visual of myself. [I] can't forget things that was said/done by myself or student (everything is concrete)." Additionally, several participants commented on the cost-effectiveness of the intervention, the ability to improve teaching as a result of watching yourself, and the immediacy of feedback. As Crystal mentioned, "Advantages are getting to see yourself and give yourself feedback instead of having to wait for an observation [from a supervisor]. It is immediate and you can make changes for the next day's lesson." In addition to the advantages, participants also mentioned several disadvantages, including the risk that your camera will malfunction and not capture sound and/or video, the amount of memory that the videos take up on recording devices, and the time it takes to upload videos. Despite these disadvantages, the participants all agreed that video analysis was both effective and worth the time invested (see Appendix C, Table C1).

The topic of watching oneself on video came up often, both in class and on the social validity surveys. While initially aversive, the participants' narrative responses indicated it did get easier to watch themselves on video as the study progressed. For example, the first time the social validity survey was administered, Stephanie mentioned, "It is hard to watch myself and not be so critical about things that don't matter and see things that do."; however, by the fourth administration of the survey, she wrote, "I like using videos to see what I can improve. It is easier to watch afterwards and find changes that you wouldn't notice at the time." This trend towards becoming more accustomed to

watching oneself on video is also reflected in the participants' responses to the statement, "I don't mind watching myself on video" (see Appendix C, Table C1), as they rated this statement more favorably after the first administration of the survey.

Limitations and Future Research

There are several limitations to this study that should be noted. For one, Angela could only record video once per week. As a result, she received a lower dosage of the intervention. Additionally, while the purpose of improving educators' behavior is to ultimately impact student outcomes, the time constraints of the semester and the added layer of consent that comes with taking student data did not allow for the collection of data on student outcomes; thus, it is unknown if the positive effects experienced by the participants had any impact on their students. Future research should investigate the impact that improving educators' skills through the use of video analysis has on student outcomes. A third limitation is that the participants took inter-observer agreement data using total count recording rather than interval by interval recording, which may have negatively impacted the inter-observer agreement results for several participants.

An additional limitation of this study is that the maintenance period was only one week following intervention. Due to the time constraints of the semester, it was not possible to take extended maintenance data; therefore, it is unknown if the improvement in skills experienced by the participants maintained for extended periods of time. Future research should examine if the effects of video analysis maintain for weeks or months following the conclusion of intervention. A final limitation is that the need to know who completed the social validity surveys for grading purposes necessitated that the

participants type their name on their survey. While there is a chance that the lack of anonymity on the survey may have artificially inflated the participants' responses, the participants knew they were not being graded on their responses, but rather the completion of the survey; thus, it is likely that their responses accurately reflected their views on video analysis.

Implications for Practice

Video analysis has many advantages that make it an appealing choice for educators wishing to improve their instructional skills. For example, because educators self-select the instructional skills they want to improve, the intervention is meaningful to them and individualized to the unique needs of their students and classroom. Additionally, having educators take and graph data on their own behavior increases the sustainability of the intervention by teaching them valuable skills that they can use with other behaviors and in other contexts when they no longer have the guidance and supervision of the teacher educator.

Supervisors who wish to take advantage of these benefits and use video analysis as a form of professional development with the educators they supervise should consider several points before implementing it. First, supervisors need to consider the amount of space that videos take up. Participants in this study had access to 1 terabyte of storage on a cloud server as part of their university fees that was shared with the investigator. However, if educators do not have access to free cloud storage, supervisors will need to consider other options to avoid using all of the hard drive space on educators' recording devices. Another point to consider is the amount of time it takes to upload videos. For

some participants, particularly those in rural areas or those with slower internet speeds, it would sometimes take a long time to upload the videos. In these cases, supervisors might suggest educators connect their device directly to the internet via an Ethernet cable rather than trying to upload files via a wireless connection. Another option is for educators to "zip" their files prior to uploading to reduce the size of the upload. If supervisors are local, they can provide educators with an external hard drive to transfer the videos for later viewing. Lastly, supervisors must expect technical issues to occur and have a plan in place when they do. For example, providing educators with a list of common technical issues and ways to overcome them may reduce the amount of questions they receive. Supervisors might also suggest educators record a few "test sessions" to find the best camera angle and position to capture quality sound and video. Additionally, having the camera running for a few sessions may help the students become accustomed to having a video camera in the room, thereby reducing the reactivity that is sometimes associated with the novelty of a camera.

By having a plan in place to reduce the amount of space that videos take up and the time it takes to upload videos, as well as to address technical issues when they occur, supervisors can increase the likelihood that educators will receive the maximum benefit from video analysis. Prior research has shown that professional development that is meaningful to educators, sustained over longer periods of time, and integrated into educators' everyday lives are more impactful and have lasting benefits on educators' practices (Garet et al., 2001; Harwell, 2003). Video analysis includes all of these features and was found to be an effective method for improving the skills of educators.

CHAPTER III

EFFECTS OF VIDEO ANALYSIS ON THE INSTRUCTIONAL SKILLS OF SPECIAL EDUCATORS: A META-ANALYSIS OF SINGLE-CASE STUDIES

Professional development opportunities for educators in the U.S. are often lacking or inadequate (Wei, Darling-Hammond, Andree, Richardson, & Orphanos, 2009). Despite overwhelming evidence to discredit its use, many professional development opportunities provided to educators follow the traditional model, which includes the one-stop workshop (Wei, Darling-Hammond, & Adamson, 2010). Delivering ineffective professional development has many negative consequences, such as leaving teachers frustrated and without the requisite skills needed to effectively teach students (Nir & Bogler, 2008; Wei et al., 2010). Because there is a strong correlation between the quality of teaching students receive and their academic success (Wei et al., 2010), it is vital that teachers receive professional development that is based on research and is aimed at improving their instructional practices (Wei et al., 2009).

Fortunately, the negative consequences associated with ineffective professional development can be avoided when a more effective model is used. Authentic professional development (also termed the reform-based approach) is an effective model of professional development (Boudah, Blair, & Mitchell, 2003) and includes the following characteristics: focusing on educators' needs when planning and implementing professional development opportunities (Boudah et al., 2003), giving educators the opportunity to practice learned skills (Corcoran, 2007), incorporating learned skills into

educators' daily lives (Garet, Porter, Desimone, Birman, & Yoon, 2001), and sustaining professional development over longer time periods (Harwell, 2003).

Video analysis, a method of evaluating one's teaching by watching previously recorded video, is one intervention that meets the characteristics of effective professional development. While there are variations to how video analysis is implemented, the core features include recording a video of an educator teaching, watching and analyzing the video, targeting an instructional behavior for improvement, and using the information learned to improve instructional practices (Nagro & Cornelius, 2013). Because educators can choose the behavior to be targeted for change (Alexander, Williams, & Nelson, 2012), video analysis focuses on their needs. In addition, because the videos are recorded in the educators' classrooms during typical instruction, educators have the opportunity to practice and improve upon the target behavior they identified within their daily lives (Tripp & Rich, 2012). Lastly, because multiple videos are typically recorded, video analysis is sustained over longer periods of time, as opposed to a one-time workshop (Hager, 2012).

While a number of studies have demonstrated that video analysis may be an effective means of increasing the instructional practices of educators (e.g., Milburn, Girolametto, Weitzman, & Greenberg, 2014; Wright, Ellis, & Baxter, 2012; Zan & Donegan-Ritter, 2014), many questions remain about whether or not variables, such as participant characteristics, instructional practices, and setting, differentially impact the effectiveness of video analysis. While there are two previous reviews on the topic (Nagro & Cornelius, 2013; Tripp & Rich, 2012), neither evaluated the differential effects

these variables have on video analysis; therefore, it is still unknown whether they have an impact. Meta-analysis, a method of aggregating and evaluating the results of a body of research on a topic, is one way to answer these questions (Lipsey & Wilson, 2001).

Video analysis has been implemented with a range of populations and experience levels, including pre-service educators (e.g., Alexander et al., 2012; Hager, 2012; Morgan, Menlove, Salzberg, & Hudson, 1994; Saudargas, 1973), novice in-service educators (i.e., three or less years of experience; e.g., Fedders, 2012; Lindsey, 2014; Reamer, 1996), and experienced in-service educators (i.e., more than three years of experience; e.g., Englund, 2011; Hawkins & Heflin, 2011; Lynes, 2013). The experience level of educators has been shown to impact the quality of their instruction, particularly during the first few years of teaching (Rice, 2010); therefore, it is important to evaluate whether or not this variable impacts the effectiveness of video analysis. In addition, video analysis has been implemented with participants from different age categories and with different levels of education. Because video analysis involves technology, which younger populations may be more apt to use (Black, 2010), it is important to evaluate the effect that age has on the effectiveness of video analysis. Evaluating the education level of participants is also important to determine if educators with higher levels of education benefit more or less from video analysis. Knowing how different educator characteristics differentially impact the effectiveness of video analysis can assist administrators and supervisors when they consider with whom to use this intervention.

In addition to participant characteristics, it is also important to evaluate the impact that setting has on the effectiveness of video analysis. Video analysis has been

implemented in general education (e.g., Ahuja, 2000; Fullerton, Conroy, & Correa, 2009), self-contained (e.g., Bingham, Spooner, & Browder; 2007; Hawkins & Heflin, 2011; Lindsey, 2014; Westover, 2011), resource (e.g., Alexander et al., 2012; Capizzi, Wehby, & Sandmel, 2010; Hager, 2012), and inclusion classrooms (e.g., Bose-Deakins, 2006; Carnine & Fink, 1978; Snyder, 2013). Video analysis has also been implemented in a variety of grade levels, including preschool (i.e., kindergarten and below; e.g., Bishop, Snyder, & Crow, 2015; Englund, 2011; Lynes, 2013), elementary (i.e., first through fifth grade; e.g., Ahuja, 2000; Saudargas, 1973; Westover, 2011), and middle/secondary (i.e., sixth through twelfth grade; e.g., Capizzi et al., 2010; Hawkins & Heflin, 2011; Pelletier, McNamara, Braga-Kenyon, & Ahearn, 2010), as well as with different student groupings, including whole group (e.g., Englund, 2011; Hawkins & Heflin, 2011; Pinter, East, & Thrush, 2015), small group (Ahuja, 2000; Carnine & Fink, 1978; Saudargas, 1973), and one-on-one (e.g., Lindsey, 2014; Reamer, 1996; Westover, 2011) . Lastly, video analysis has been implemented with various types of instruction, including academic (e.g., Fedders, 2012; Morgan et al., 1994; Pinter et al., 2015), communication (e.g., Bingham et al., 2007; Englund, 2011; Robinson, 2011), and daily living skills (e.g., Reamer, 1996; Peck, Killen, & Baumgart, 1989). Because video analysis has been implemented in such a wide range of settings, it may be difficult for educators and administrators to know in which setting video analysis will be most beneficial; therefore, it is important to investigate this variable to assist practitioners in choosing the best setting in which to implement this intervention.

When conducting a meta-analysis, researchers must decide whether or not to include studies based on basic design quality standard ratings (Bernard, Borokhovski, Schmid, & Tamin, 2014). To improve the rigor of educational research and provide readers with data on the methodological quality of studies, several sets of quality indicators have been developed to evaluate the quality of single-case research (Council for Exceptional Children [CEC], 2014; Horner et al., 2005; U.S. Department of Education, 2016). While What Works Clearinghouse (WWC) typically excludes studies from evidence-based reviews based on the overall design quality rating (U.S. Department of Education, 2016), others have advocated for a more inclusive approach, particularly in areas that have a limited amount of research (Dijkers, 2009). There are concerns, however, that including studies that do not meet basic design quality standards will invalidate the results of a meta-analysis or mislead readers with inaccurate results (Cook et al., 2015). While these concerns are understandable, there are reasons for including studies that do not meet design quality standards, such as the ability to provide a more accurate representation of the body of research on a topic (Lipsey & Wilson, 2001) and to identify areas in need of additional research. In addition, researchers often disagree on what qualifies as a "high quality" study, even when a set of quality indicators are used (Cooper, 2010). Instead of excluding studies based on design quality, a better approach might be to "let the data speak" by including all studies that meet initial inclusion criteria, coding the methodological quality of studies based on a set of quality indicators, and empirically examining the effects of the variations in methods (Cooper, 2010, p. 124). By doing so, researchers can avoid excluding studies

unnecessarily and include a larger body of research, with a potentially more diverse range of participants and dependent variables, from which to draw conclusions.

The purpose of this study is to use meta-analytic methods to investigate the effectiveness of video analysis on the instructional practices of educators. The research questions that will be investigated in this study are as follows:

1. What is the status of the literature base on video analysis regarding (a) study characteristics (i.e., publication type, design quality), (b) participant characteristics (i.e., role, education level, experience level, and age), (c) student characteristics (i.e., disability type and collection of student outcomes), and (d) setting characteristics (i.e., grade level, group size, type of instruction, and setting)?
2. What effects do publication type and methodological quality have on the effectiveness of video analysis?
3. What is the omnibus magnitude of effect of video analysis on the instructional practices of educators?
4. What effects do participant characteristics (e.g., role, education level, experience level, age) have on the effectiveness of video analysis?
5. What effects do instructional characteristics (i.e., group size, type of instruction, grade level, setting) have on the effectiveness of video analysis?

Method

Study Identification

Primary search method. Three primary searches were conducted: an ancestral search, a forward search, and a first author search. All searches were limited to peer-reviewed articles and dissertations and were conducted using the following electronic databases: *ERIC, PsycARTICLES, PsycINFO, Education Source, Teacher Reference Center, Academic Search Complete*, and *Education Full Text*. The first primary search, which was conducted on 6/25/15, was carried out by combing the term *teacher** with the following search terms: *video**, *analy**, *evaluat**, *reflect**, and *feedback**. The second primary search, which used the same search terms as the initial search, was conducted on 11/5/16 and covered the time period from 6/25/15 to 11/5/16. The final primary search was also conducted on 11/15/16 and combined the terms *paraeducator*, "teach* assistant*," paraprofessional*,* and *"instructional assistant*"* with *video**, *analy**, *evaluat**, *reflect**, and *feedback**. The date published was not limited for the third search.

Title/abstract review. Once documents were identified, they were exported into folders in Refworks, a web-based bibliography and database manager, and duplicates were removed from each folder. After duplicates were removed, the searches resulted in 7,583 documents, 1,132 documents, and 215 documents for the first, second, and third primary searches, respectively. Titles and abstracts of all identified documents were evaluated to determine if they met inclusion criteria. If it could not be determined whether the article met inclusion criteria from the title and abstract alone, the full text of

the document was searched. As a result of the title/abstract screening, 7,816 records were excluded via the primary searches. Appendix D (see Figure D1) summarizes the number of documents that remained for each search once duplicates were removed.

Full-text review. A total of 1,114 documents from the primary searches were screened by full-text. Application of the inclusion/exclusion criteria resulted in the exclusion of 979 documents. The primary reasons that documents were excluded from the full-text search were as follows: the studies did not include (a) quantitative data (*n* = 394; 40%), (b) at least one dependent variable related improving observed teacher behavior (*n* = 187; 19%), or (c) at least one participant who was the focus of the video (*n* = 124; 12%). Additional reasons for exclusion were as follows: the study did not (a) include at least one participant who was a teacher or paraprofessional in early childhood through 12th grade (*n* = 69; 7%), (b) analyze recorded videos of teachers or paraprofessionals (*n* = 69; 7%), (c) include comparative data (*n* = 55; 5%), or (d) include an evaluation/feedback component (*n* = 49; 5%). Articles were also excluded if they were not in English (*n* = 47; 5%).

Design review. In an effort to be comprehensive, documents were not initially screened for type of design; however, after the inclusion/exclusion criteria were applied to documents identified from the primary searches, design type was coded and studies that did not include a single-case experimental design were excluded. This resulted in the exclusion of an additional 103 documents. Additionally, five single-case design documents were excluded because they were either not comparable to the included studies or, upon closer inspection, did not meet the inclusion criteria. Specifically,

Herbert (1993) was excluded because only one participant's data were presented, despite being described as a multiple-baseline design; Sharpe, Spies, Newman, and Spickelmier-Vallin (1996) and Sloat, Tharp, & Gallimore (1977) were excluded because the phases that used video were likely affected by sequencing effects; Kirk-Martinez (2011) was excluded because there is not a true baseline phase due to the provision of training in baseline; and Venn & Wolery (1992) was excluded because the setting was in a daycare, and the dependent variables were not related to teaching behaviors.

Ancestral/forward/first author search method. In addition to the primary searches, an ancestral, a forward, and a first author search of the documents that met inclusion criteria were conducted using the same procedures as the primary search. Each included document was entered into Scopus, an abstract and citation database, and a several lists of documents were generated and exported into a folder in Refworks. These lists included (a) documents included in the reference list of the included document (i.e., ancestral search), (b) documents that have cited the included document (i.e., forward search), and (c) other documents that were authored by the first author of the included document (i.e., first author search). A hand search of the reference list was conducted for any documents that were not included in Scopus. Appendix D (see Figure D1) provides information on the number of documents remaining after this stage of the search process. In addition to the documents that were excluded due to not meeting inclusion criteria, an additional five articles were excluded because they did not include a single-case line graph and another article was excluded because it was already included from the primary searches. The reasons documents were excluded from this search included the following:

they did not include (a) an analysis of recorded videos of teachers or paraprofessionals (*n* = 77; 64%); (b) quantitative data (*n* = 18; 15%); (c) at least one participant who was a teacher or paraprofessional in early childhood through 12th grade (*n* = 16; 13%); (d) at least one participant who was the focus of the video (*n* = 7; 6%); (e) an evaluation or feedback component (*n* = 1; 1%); or (f) at least one dependent variable on improving observed teacher behavior (*n* = 1; 1%).

Inclusion/exclusion criteria. The following inclusion criteria were used for the first primary search: (a) the study included quantitative data, (b) at least one participant was a teacher in an early childhood through 12th grade classroom or was in a teacher preparation program for early childhood through 12th grade, (c) recorded videos of the in-service or pre-service teacher were analyzed, (d) the intervention included an evaluation or feedback component, (e) the participant is the focus of the video, (f) the study included comparative data (e.g., pre-/post-, treatment/control, single case graph, and/or data at different points in time), (g) at least one dependent variable is related to improving observed teacher behavior, and (h) the study is in English. The second and third searches used the same inclusion criteria with exception that the word teacher was replaced with teacher or paraprofessional. A paraprofessional was defined as an assistant teacher who worked under the direct supervision of a teacher, and a teacher was defined as a lead teacher in an early childhood through 12th grade setting. Early childhood was defined as at least 36 months of age.

Exclusion criteria included the following: (a) qualitative studies, reviews, and discussion articles, (b) direct care staff at residential facilities who did not work under

the direct supervision of a certified teacher, (c) day care workers who did not provide any academic or behavioral instruction, (d) home and clinic settings, (e) studies that included only videos of others (e.g., videos depicting exemplary practice by someone other than the participants), and (f) unobserved or non-behavioral dependent variables, such as answers to a content knowledge test or survey or data on the participants' reflections or ability to reflect while watching the video.

Application of the What Works Clearinghouse Design Quality Standards

Once all studies were identified, they were evaluated for design quality using the What Works Clearinghouse (WWC) Basic Design Quality Standards (Kratochwill et al., 2010; U.S. Department of Education, 2016). Because one of the purposes of this study was to empirically investigate the effects of the methodological quality of studies, studies were not excluded based on design quality.

Basic Design Quality Standards (U.S. Department of Education, 2016). Each study was evaluated at the experiment level, defined as one single-case experimental design, using the WWC Basic Design Quality Standards (U.S. Department of Education, 2016). For example, if a document included three multiple-baseline design experiments, then each experiment was evaluated separately. Experiments were evaluated on the presence of the following WWC Basic Design Quality Standards (U.S. Department of Education, 2016): (a) manipulation of the independent variable (Design Standard 1), (b) whether or not inter-observer agreement was reported (Design Standard 2A), (c) the percentage of data for which inter-observer agreement was collected (Design Standard 2B), (d) whether or not inter-observer agreement scores met minimum quality thresholds

(Design Standard 2C), (e) whether or not the experiment included a minimum of three attempts to demonstrate treatment effects at three different points in time (Design Standard 3), and (f) the number of data points per phase (Design Standard 4). Multiple-probe designs (MPB) were also evaluated on the presence of the following additional Basic Design Quality Standards (U.S. Department of Education, 2016): (a) the number of data points within the initial baseline sessions (Design Standard 5A), (b) the number of consecutive probe points prior to intervention (Design Standard 5B), and (c) the collection of data points in subsequent levels when the previous level first received intervention (Design Standard 5C). Once each experiment was coded on the basic design standards, an overall design quality rating was assigned. Experiments with an overall design quality rating of "2" were considered to meet the standards without reservations, experiments with an overall rating of "1" were considered to meet the standards with reservations, and experiments with an overall rating of "0" were considered to not meet the standards (U.S. Department of Education, 2016). Each standard, and the criteria used to evaluate each experiment, are described in more detail in Appendix E (see Table E1).

Variable Coding

A portion of the included articles were coded using descriptive data in order to develop a coding menu. Once enough information had been extracted to allow patterns in the data to develop, a coding menu was created, and each study was coded for the following variables: (a) role, (b) education level, (c) experience level, (d) age, (e) group size, (f) type of instruction, (g) grade level, (h) setting, (i) design type (i.e., multiple, baseline, multiple probe, reversal, etc.), and (j) publication form (i.e., peer-reviewed

article or dissertation). Additionally, studies were coded for the length and number of videos/sessions, role of the primary and secondary evaluators, type and timing of feedback, the collection of generalization and maintenance data, and dependent variable type; however, these variables will be investigated in a future study and are not reported here.

Appendix F (see Table F1) provides operational definitions and subgroup categories for role, group size, type of instruction, grade level, and setting. Education, experience, and age variables were coded as follows: education (high school/GED only, some college or specialized training [e.g., Associate's degree, pre-service teachers enrolled in an undergraduate teacher preparation program, early childhood certificate, etc.], Bachelor's degree, Master's degree), experience (none or first year, in 2nd or 3rd year of teaching [1 to 2 full years of teaching experience], or in 4th year or more [3 full years or more of teaching experience]), age (18-29, 30-39, 40-49, 50 and over). Some subgroups were initially coded separately (i.e., experience: none, first year; age: 50-59, 60 and over; grade level: middle school, high school; type of instruction: reading/ELA, math) but were later combined prior to moderator analyses due to a low number of contrasts in each category. In addition, while setting originally included a code for general education, there were not enough contrasts to analyze for this subgroup and it was later dropped. For all variables, if a participant, characteristic, or intervention did not fit into any of the categories created for each variable, or if they fit into multiple categories, the study was coded as "0" for that particular variable and was excluded from

further analysis. Studies that did not include information on the variable being coded were also coded as a "0" and excluded from further analysis.

While the initial inclusion criteria did not include a stipulation that the teacher or paraprofessional had to view his or her videos—only that there had to be a feedback or evaluation component—after coding for potential moderators, any studies or contrasts that did not involve the teacher or paraprofessional watching his or her videos were excluded. The decision to exclude these documents was made because having the teacher or paraprofessional view his or her videos is more aligned with the purpose of video analysis (Nagro & Cornelius, 2013). The following studies were excluded as a result of this decision: Artman-Meeker and Hemmeter (2012), Cardinal (2012), and Courtemanche et al. (2014). Additionally, the following studies were excluded because, after closer inspection, they were found to either not meet the original inclusion criteria or to differ significantly from other studies in terms of setting, participants, or procedures: Rule (1973) was excluded because the setting was an alternative school staffed only by volunteers and it wasn't clear whether it was comparable to a typical early childhood through 12th grade setting; Ford (1984) was excluded because the setting was a "mental retardation" facility rather than an early childhood through 12th grade setting; Westover and Martin (2014) was excluded because it is the published form of Westover (2011) with the exception that Westover and Martin (2014) only included a subset of the data that were presented in Westover (2011); and Duker, Hensgens, and Venderbosch (1995) was excluded because the setting is a training facility rather than an early childhood through 12th grade education setting.

Data Extraction and Effect Size Calculation

Data extraction. Data were extracted from each graph using the free, online software, GetData Graph Digitizer. Similar graph digitizers have been shown to have high reliability (Shadish et al., 2009) and have been used in several previous systematic reviews of single-case experimental research (Gage & Lewis, 2014; Lequia, Wilkerson, Kim, & Lyons, 2015; Losinski, Cuenca-Carlino, Zablocki, & Teagarden, 2014). A JPEG image of each graph was scanned into the program and the coordinates and data points were plotted. The resulting digitized results of the baseline and intervention data for each AB contrast were exported to an Excel file.

Data analysis. An effect size was calculated for each study and for potential moderators using Tau-U (Parker, Vannest, Davis, & Sauber, 2011), which can be interpreted as the "percentage of nonoverlap versus overlap." Tau-U (Parker et al., 2011) was selected as the effect size measure for this meta-analysis because it has several advantages over other non-parametric effect sizes (Parker et al., 2011), including (a) the use of all data points, making it less susceptible to outliers, (b) greater statistical power and precision than other nonoverlap effect sizes, (b) the ability to control for undesirable baseline trend, (d) the ability to calculate confidence intervals, (d) high sensitivity, and (e) simple calculation. Additionally, Tau-U has been found to be consistent with visual analysis of data (Brossart et al., 2014; Parker et al., 2011) and has been used increasingly in single-case research studies (Hong, Ganz, Gilliland, & Ninci, 2014; Hutchins & Prelock, 2013) and single-case meta-analyses (Bowman-Perrott, Burke, Zaini, Zhang, & Vannest, 2016; Hong et al., 2016, Neely et al., 2016; Ninci et al., 2015). Interpretive

guidelines for Tau-U effects are as follows: small effect = 0 to .62; medium effect = .63 to .92; large effect = .93 to 1.00 (Parker, Vannest, & Davis, 2011).

Effect sizes were calculated by first entering baseline and intervention data into the free, online Tau-U calculator (Vannest, Parker, Goen, & Adiguzel, 2016) to obtain a Tau-U value for each AB contrast (i.e., baseline versus intervention). In cases where the goal of the intervention was to decrease behavior, the data were entered in reverse. In other words, the intervention data were entered as the A-phase and the baseline data were entered as the B-phase. Next, the effect sizes were combined in the Tau-U calculator to produce one effect size per study. Because trend was present in 81% of the baseline data and, of that, 31% was undesired trend, all results reflect corrected baseline data, with the exception of studies whose goal was to decrease behavior. When data are entered into the Tau calculator in reverse, trend is expected in the A-phase because the A-phase consists of intervention data. Therefore, uncorrected data were used for these contrasts. Lastly, an omnibus effect size was generated by entering the Tau-U value and its standard error (SD_{Tau}) into the Comprehensive Meta-Analysis software program (Version 3; Borenstein, Hedges, Higgins, & Rothstein, 2005). Moderator analyses were also conducted by entering the Tau-U value and its standard error (SD_{Tau}) for each AB contrast into the Comprehensive Meta-Analysis software program (Version 3; Borenstein, et al., 2005) and generating an effect size for each potential moderator and its associated subgroups.

The Comprehensive Meta-Analysis software program (Version 3; Borenstein, et al., 2005) generates omnibus effect sizes for both a random effects and fixed effects

model. While neither a fixed-effects nor random-effects model is an "exact fit" for single-case data (Bowman-Perrott et al., 2016), a random effects model was preferred in this case because the studies included in this meta-analysis vary in terms of the participants, outcome measures, procedures, and settings, and it was hypothesized that the variance between studies was due to systematic differences rather than sampling error alone (Bornenstein et al., 2009; Lipsey & Wilson, 2001). Statistical significance for potential moderators was determined using the between *Q*-statistic. The *Q*-statistic is a test of homogeneity, and the null hypothesis assumes that all studies share a common effect size and any variance between subgroups is due to chance or random error (Borenstein, Hedges, Higgins, & Rothstein, 2009). When the associated *p*-value is less than 0.05, there is evidence that the differences in the dispersion (i.e., range) of effect sizes between the subgroups is due to real differences and not random error (Borenstein et al., 2009).

Inter-rater reliability. A minimum of 20% of data for all phases of the study were independently coded for inter-rater reliability by one of four doctoral students—two first-year students in special education, one second year student in school psychology, and one third year student in school psychology. All raters were trained to criterion for each stage of the process using a subset of data. If there was a disagreement on whether or not an article should be included, a third evaluator independently rated the studies or the first two evaluators discussed the disagreement until they came to a consensus. In the cases where three evaluators were needed, the final decision was based on the agreement of two evaluators. Inter-rater reliability was calculated by dividing the

number of agreements by the number of agreements plus disagreements and multiplying that number by 100 to obtain a percentage.

For study identification, a total of 43% of the documents from the primary searches and 42% of the documents from the ancestral/forward search were coded by a second independent rater for reliability, resulting in 98% agreement for the primary searches and 99% agreement for the ancestral search. For the WWC Basic Design Quality Standards (U.S. Department of Education, 2016), a second rater coded 47% of experiments for Design Standards 1-4 and the overall evaluation and 78% of experiments for Design Standards 5A-5C, resulting in an overall inter-rater reliability score of 95% (range 86%-100%). More specifically, the reliability scores were 100%, 100%, 86%, 94%, 100%, 91%, 100%, 100%, 86%, and 89% for Design Standards 1, 2A, 2B, 2C, 3, 4, 5A, 5B, 5C, and the overall evaluation, respectively. Disagreements among raters for Design Standards 2B and 5C were slightly higher than disagreements for other Design Standards because variations in wording used by authors to report the frequency in which they collected inter-observer agreement data (Design Standard 2B) and the manner in which graphs were formatted (Design Standard 5C) resulted in a higher level of subjectivity when coding for these standards.

When coding for variables, a second rater coded 51% of the contrasts for each variable, resulting in a mean agreement of 95% (range 92%-99%). More specifically, the inter-rater reliability scores for role, education level, experience level, age, group size, type of instruction, grade level, and setting were 93%, 99%, 97%, 92%, 95%, 96%, 94%, and 92%, respectively. For data extraction, a second rater extracted the data for 23% of

the data points in baseline and 21% of the data points in intervention, resulting in 99.4% agreement for baseline data points and 99.8% agreement for intervention data points. When comparing the data between raters for data extraction, if scores fell within 0.1 point for scores rounded to the nearest tenth (e.g., rate) and 1.0 point for scores rounded to the nearest one (e.g., frequency or percent), the difference was considered rounding error and not counted as a disagreement. Because graph digitizers extract data with extreme precision (13 decimal points), some rounding error is to be expected. For scores that were counted as a disagreement, the GetData workspace files were reviewed to determine which score was accurate. For data analysis, a second rater entered 25% of the extracted AB contrast data into the Tau-U calculator (Vannest et al., 2016) and effect sizes obtained for the first and second raters were compared for reliability purposes, resulting in 98% agreement. There was one disagreement which was resolved by re-calculating the data to determine where the disagreement lay.

Results

Status of the Literature Base

The first research question was, "What is the status of the literature base on video analysis regarding study characteristics (i.e., publication type, design quality), participant characteristics (i.e., role, education level, experience level, and age), student characteristics (i.e., disability type and collection of student outcomes), and setting characteristics (i.e., grade level, group size, type of instruction, and setting)?" To answer this research question, narrative data on each of these variables were analyzed to determine the percentage of studies or participants that were coded for each variable.

Study characteristics. A total of 61% (*n* = 17) of included documents were peer-reviewed articles and 39% (*n* = 11) were dissertations (see Appendix G, Table G1). There were 58 single-case experiments across the 28 included documents. Of the 58 experiments, 62% (*n* = 36) were multiple-baseline designs, 21% (*n* = 12) were a variation of an AB design, 8.5% (*n* = 5) were multiple-probe designs, and 8.5% (*n* = 5) were reversal designs. Across the 28 included documents, half (50%; *n* = 14) met the WWC Design Quality Standards (U.S. Department of Education, 2016) with reservations, 39% (*n* = 11) did not meet standards, and 11% (*n* = 3) had some experiments that met the standards with reservations and some that did not meet standards. None of the included documents or experiments met standards without reservations. Of the included experiments, participants ranged from one to eight per study, AB contrasts ranged from two to 24, and the number of pairs ranged from 60 to 4,144.

Participant characteristics. There was a total of 105 participants included across the 28 articles and dissertations. While Appendix G (see Table G2) provides a general overview of each study, specific information regarding the participants' role, education level, experience level, and age is provided here. In some instances, narrative data are provided in the table, but the variable was coded as "other" and excluded from analyses due to the authors describing the participants in general terms rather than giving specific demographic information on each participant.

Regarding role, there were a total of 105 participants included across the 28 articles and dissertations. Of these, more than half (52%; *n* = 55) were in-service

teachers, 24% (*n* = 25) were paraprofessionals, 13% (*n* = 14) were pre-service teachers, and 11% (*n* = 11) were coded as "other.". Regarding education level, most participants were coded as "other" (37%; *n* = 39), with another 24% (*n* =25) having a Bachelor's degree, 18% (*n* = 19) having some college or specialized training, 12% (*n* = 13) having a Master's degree, and 9% (*n* = 9) having a high school diploma or GED only. Regarding age, the majority of participants were coded as "other" (60%; *n* = 63), with an additional 21% (*n* = 22) between the ages of 18-29, 8.5% (*n* = 9) between the ages of 30-39, 5.5% (*n* = 6) between the ages of 40-49, and 5% (*n* =5) ages 50 and over. For experience level, most of the participants were in their fourth year or more of teaching (41%; *n* = 43), with an additional 22% (*n* = 23) of the participants in their second or third year of teaching, 19% (*n* = 20) coded as "other," 10.5% (*n* = 11) in their first year of teaching, and 7.5% (*n* = 8) having never taught.

Student characteristics. Studies were also coded for the type of disability that students in the classroom exhibited and the type of student outcomes that were collected. These results can be found in Appendix G (see Table G3). Fewer than half of the studies collected data on student outcomes (43%; *n* = 12). Across the 12 studies that did collect data on student outcomes, a total of 21 outcomes were reported. Out of these 21 outcomes, 43% (*n* = 9) were academic outcomes (i.e., correct responses [*n* = 4], rate of responses [*n* = 1], no response [*n* = 1], engagement [*n* = 2], literacy skills [*n* = 1]), 14% (*n* = 3) were behavioral outcomes (i.e., challenging behavior, compliance, following directions), 10% (*n* = 2) were social outcomes (i.e., social communication goals, student effect), and 33% (*n* = 7) were communication outcomes (i.e., use of AAC device,

prompted and unprompted use of communication targets, total communicative responses, one word utterances, number of words spoken, labelling actions, answering yes/no questions).

A total of 42 disabilities were reported in the classrooms in which the teachers and paraprofessionals taught. Of these, 38% (*n* = 15) were developmental disabilities (i.e., autism spectrum disorder [*n* = 7], intellectual disability [*n* = 5], Down syndrome [*n* = 2], and developmental disorder [*n* = 1]), 14% (*n* = 6) were physical disabilities (i.e., physical disability [*n* = 4], cerebral palsy [*n* = 2]), 14% (*n* = 6) were mental disabilities (i.e., anxiety disorder, conduct disorder, bipolar disorder, depression, schizophrenia, and attention deficit hyperactivity disorder), 12% (*n* = 5) were emotional or behavioral disorders, 9.5% (*n* = 4) were learning disabilities, 9.5% (*n* = 4) were other disabilities (i.e., multiple disabilities [*n* = 3], other health impairment [*n* = 1]), and 5% (*n* = 2) were cognitive disabilities (i.e., cognitive impairment, brain injury). In addition to these reported disabilities, seven studies reported having students with developmental delays, three studies reported having other types of delay (i.e., fine motor, literacy, language, and cognitive), seven studies did not report whether or not the students had a disability, and one study reported that the students had challenging behavior (see Appendix G, Table G3).

Setting characteristics. Narrative data were also collected on variables related to setting or instructional characteristics. Specifically, grade level, group size, type of instruction, and setting in which the educators taught (see Appendix G, Table G3) were coded for each study. In some instances, narrative data are provided in the table, but the

variable was coded as “other” and was excluded from the analyses, primarily because the authors provided a general description only or the type of instruction fell in multiple categories.

Regarding grade level, 34% (*n* = 36) of educators taught in a preschool setting, 33% (*n* = 35) taught in an elementary setting, 6% (*n* = 6) taught in a middle school setting, 11% (*n* = 11) taught in a high school setting, and 16% (*n* = 17) were coded as “other.” Regarding group size, most participants (39%; *n* = 41) provided instruction in a one-to-one instructional setting, while another 31.5% (*n* = 33) taught in a small group setting, 18% (*n* = 19) in a large group setting, and 11.5% (*n* = 12) were coded as “other.” Regarding the type of instruction the educator was delivering when implementing video analysis, most of the participants delivered instruction in the “other” category (43%; *n* = 45). A large percentage of participants also delivered academic instruction (39%; *n* = 41), and the remaining participants delivered communication or language instruction (13%; *n* = 14) or daily living skills instruction (5%; *n* = 5). Of the academic instruction that was delivered, all but one participant delivered instruction in reading or language arts (*n* = 40); the remaining participant delivered instruction in math skills. Lastly, regarding setting, most participants taught in a self-contained (39%; *n* = 41) or inclusion (31%; *n* = 33) classroom. An additional 12% (*n* = 13) taught in a resource classroom, 9% (*n* = 9) taught in a general education classroom, and 9% (*n* = 9) were coded as “other.”

Methodological Quality and Publication Type

To answer the second research question, "What effect does type and methodological quality have on the effectiveness of video analysis?," the effects of video analysis were analyzed separately for AB contrasts that were peer-reviewed (n = 70) and that were dissertations (n = 108) and for AB contrasts that met WWC Design Quality Standards (U.S. Department of Education, 2016) with reservations (n = 93) and those that did not meet standards (n = 85). These results can be found in Appendix H (see Figure H1). The effects of video analysis were slightly larger for peer-reviewed articles (ES = 0.90) than for dissertations (ES = 0.83); however, the difference in the two effect sizes were not statistically significant (Q_b = 1.95; p = 0.16), indicating that publication bias was not present in this data set (Ferguson & Brannick, 2012). When considering the methodological quality of the included studies, the effects of video analysis were stronger for studies that met WWC Design Quality Standards (U.S. Department of Education, 2016) with reservations (ES = 0.90) than they were for studies that did not meet WWC Design Quality Standards (U.S. Department of Education, 2016; ES = 0.82); however, these differences were not statistically significant (Q_b = 2.95; p = 0.09). Because there were no statistically significant differences between studies that met WWC Design Quality Standards (U.S. Department of Education, 2016) and those that did not meet standards, the studies that did not meet WWC Design Quality Standards (U.S. Department of Education, 2016) were included in all subsequent analyses.

Overall Effects of Video Analysis

To answer the third research question, "What is the magnitude of effect of video analysis on the instructional practices of educators?," effect sizes for each study were first calculated separately and then aggregated to obtain an omnibus effect size (see Appendix H, Figure H2). Overall, the use of video analysis to change special educators' behavior produced moderate effects (ES = 0.85) across the 28 included studies. This effect size was calculated from 178 AB contrasts across 105 participants (see Table 3).

The overall *Q*-value was 44.39 with a *p* value of 0.02, indicating that at least some of the dispersion in the effect sizes of the studies is due to real differences in study effects as opposed to random error. Of the dispersion that is seen between the studies' effect sizes, the I^2 value indicates that approximately 39% of that dispersion is probably due to real differences ($I^2 = 39.18$). Therefore, potential moderators related to participant and instructional characteristics were empirically investigated to determine where these differences lay.

Participant Characteristics

To answer the fourth research question, "What effects do participant characteristics have on the effectiveness of video analysis?" AB contrasts were coded according to the role, education level, experience level, and age of the participants and differences in the homogeneity of the effect sizes between subgroups were investigated using the *Q* statistic (see Appendix H, Figures H1 & H3). Overall, moderate to large effects were found for all subgroups, with role being the only potential moderator found to have statistically significant differences between the subgroups.

Role. To determine whether role had a statistically significant effect on the effectiveness of video analysis, AB phase contrasts were coded as paraprofessional (n = 66), pre-service teacher (n = 21), or in-service teacher (n = 77). Video analysis showed larger effects for paraprofessionals (ES = 0.91) than for both pre-service teachers (ES = 0.80) and in-service teachers (ES = 0.78), and this difference was statistically significant (Q_b = 6.78, p = 0.03).

Education level. To determine whether the participants' level of education had a statistically significant effect on the effectiveness of video analysis, phase contrasts were coded as high school/GED (n = 21), some college (n = 29), Bachelor's degree (n = 59), or Master's degree (n = 18). Participants with a high school diploma or GED only had a larger effect size (ES = 0.96) than those with some college or specialized training (ES = 0.85), a Bachelor's degree (ES = 0.83), or a Master's degree (ES = 0.78); however, these differences were not statistically significant (Q_b = 4.08, p = 0.25).

Experience level. To determine whether the experience level of educators had a statistically significant effect on the effectiveness of video analysis, AB phase contrasts were coded as none or first year (n = 41), second or third year (n = 45), or fourth year or more (n = 63). Video analysis showed larger effects for educators who had no experience or who were in their first year of teaching (ES = 0.93) than for those who were in their second or third year of teaching (ES = 0.78) or for those in their fourth year or more of teaching (ES = 0.85), but these differences were not statistically significant (Q_b = 4.72, p = 0.10).

Age. To determine whether the age of educators had a statistically significant effect on the effectiveness of video analysis, AB phase contrasts were coded as 18-29 (n = 49), 30-39 (n = 18), 40-49 (n = 19), and 50 and over (n = 10). Video analysis showed equally large effects for educators who were 50 and over (ES = 0.94) and for those who were 40-49 (ES = 0.94) and these effects were larger than for educators who were 18-29 (ES = 0.84) and 30-39 (ES = 0.78); however, these differences were not statistically significant (Q_b = 3.51, p = 0.32).

Instructional Characteristics

To answer the fifth research question, "What effects do instructional characteristics (i.e., group size, type of instruction, grade level, setting) have on the effectiveness of video analysis?" AB contrasts were coded according to the group size, type of instruction, grade level, and setting in which the participants taught. Differences in the homogeneity of the effect sizes between subgroups were then investigated using the Q statistic (see Appendix H, Figure H3). Moderate to large effects were found for all subgroups, with no statistically significant differences found for any of the potential moderators.

Group size. To determine whether group size had a statistically significant effect on the effectiveness of video analysis, AB phase contrasts were coded as one-to-one (n = 82), small group (n = 53), or large/whole group (n = 23). Video analysis showed larger effects for educators who taught in a one-to-one grouping arrangement (ES = 0.93) than for those who taught in a small group (ES = 0.82) or large/whole group (ES = 0.80); however, these differences were not statistically significant (Q_b = 5.57, p = 0.06).

Type of instruction. To determine whether the type of instruction delivered by the educators had a statistically significant effect on the effectiveness of video analysis, AB phase contrasts were coded as communication (n = 18), academic (n = 78), or daily living skills (n = 18). Video analysis showed larger effects when educators taught communication skills (ES = 0.97) than when they taught academic (ES = 0.89) or daily living skills (ES = 0.84); however, these differences were not statistically significant (Q_b = 1.50, p = 0.47).

Grade level. To determine whether the grade level in which the educators taught had a statistically significant effect on the effectiveness of video analysis, AB phase contrasts were coded as preschool (n = 61), elementary (n = 60), or middle/high school (n = 30). Video analysis showed larger effects for educators who taught in an elementary classroom (ES = 0.89) than for those who taught in a middle/high school (ES = 0.84) or preschool classroom (ES = 0.83); however, these differences were not statistically significant (Q_b = 1.54, p = 0.46).

Setting. To determine whether setting had a statistically significant effect on the effectiveness of video analysis, AB phase contrasts were coded as self-contained (n = 74), resource (n = 25), or inclusion (n = 57). Video analysis showed larger effects for educators who taught in a self-contained setting (ES = 0.91) than for those who taught in a resource (ES = 0.85) or inclusion setting (ES = 0.81); however, these differences were not statistically significant (Q_b = 3.63, p = 0.16).

Discussion

This meta-analysis investigated the effects of video analysis on the instructional practices of educators by analyzing the results of 28 single-case experimental design studies, which were implemented by a diverse pool of educators in a range of settings. The current study appears to be the first to use meta-analytic techniques to review and analyze the body of literature on video analysis. When considering individual study effects, the majority of studies had either strong or moderate effects, with only a few studies demonstrating low effects. Overall, the results demonstrated moderate effects and support the use of video analysis to change educators' instructional practices. The results of moderator analyses provide further information regarding for whom and under what circumstances video analysis is more or less effective.

The first research question addressed in this study focused on the status of the literature base on video analysis regarding study, participant, student, and setting characteristics. Regarding study characteristics, while most of the included documents were peer-reviewed articles, a large percentage were also dissertations. Of the included experiments, the majority were multiple baseline designs and met the WWC Design Quality Standards (U.S. Department of Education, 2016) with reservations. While it is encouraging that most experiments met WWC Design Quality Standards (U.S. Department of Education, 2016) with reservations, the finding that none of the experiments met the standards without reservations and a large percentage of experiments did not meet the standards indicates that more high-quality research is warranted in this area. Additionally, while the total number of documents included in

this meta-analysis is large enough to generate conclusions about the overall status of the literature base on video-analysis, only 11 of these documents were peer-reviewed articles, suggesting that more peer-reviewed research is needed. Adding more high-quality, peer-reviewed research to the literature base will potentially strengthen any conclusions drawn from future meta-analyses on the topic.

The literature base was also analyzed in regard to participant characteristics. While most participants were in-service teachers, a large percentage of paraprofessionals were also represented in this meta-analysis. Pre-service teachers were the minority across all studies, suggesting more research is needed with this population, particularly since statistically significant differences were found among the subgroups in this variable category. Of the participants whose education level was described in the study, most held a Bachelor's degree. This is consistent with the finding that most participants were in-service teachers, as a Bachelor's degree is typically required for this position. While a large percentage of studies did not provide the age of the participants, of those that did report age, most participants were between the ages of 18-29 with an equally small number of participants between the ages of 40-49 and 50 and over. More research is warranted with older educators, because, although this age group had the smallest number of participants, it also had the largest effects. More research is needed to determine if the strong effects found for educators ages 40 and over will maintain when video analysis is investigated with more participants. In terms of experience, most of the participants were experienced educators. Less than 20% of the participants had either never taught or had less than one year of experience. None/first year educators had the

strongest effects, despite having the least amount of participants and AB contrasts in this category; however, caution should be used when interpreting these results as this finding may be an artifact of the low number of contrasts in this variable category. Considering the differences in instructional quality of novice and experienced educators (Cortina, Miller, McKenzie, & Epstein, 2015), it is encouraging that the results demonstrated video analysis is a viable way to develop the instructional skills of novice educators; however, more research is warranted to determine if these strong effects will maintain when used with a larger number of participants.

Student and setting characteristics were also investigated to determine the status of the literature base regarding these variables. While the inclusion criteria were not initially limited to studies that contained students with disabilities, after coding studies for setting and disability type, it was apparent that all studies contained one of the following characteristics: (a) was conducted in a classroom with students with disabilities, (b) used an instructional curriculum that is often used with students who have disabilities or who are at risk (i.e., Ahuja, 2000; Direct Instruction), (c) was part of a federally funded project for students at risk (i.e., Saudargas, 1973; Project Follow Through: Behavior Analysis Model), or (d) included students who displayed challenging behavior such that it interfered with their learning and the learning of others (Fullerton et al., 2009). Of the studies that did include students with disabilities, there were a variety of disability types represented across the literature base. The most common disability category was developmental disabilities, with the largest number of studies reporting the inclusion of students with autism. Intellectual disabilities and emotional or behavioral

disorders were also fairly well represented across the studies. The large number of studies that were conducted in settings that included students with disabilities, particularly those with more severe disabilities, is encouraging as relevant, high quality professional development can mitigate the stress and burnout that is often experienced by educators of students with more significant needs (Billingsley, 2004; Nichols & Sosnowsky, 2002).

While it is important to collect data on student outcomes to determine if the changes in educators' teaching practices have a resulting impact on the skills of the students they teach, fewer than half of the included studies reported data on student outcomes. This finding is concerning considering the goal of professional development designed to improve the instructional practices of educators is to impact student outcomes. Without data on student outcomes, it is unknown whether the moderate to strong effects found for video analysis positively affected the educators' students. Of those studies that did report data on student outcomes, most outcomes were related to academics, with a large percentage also related to communication or language. Social and behavioral outcomes were reported less frequently and are an area for additional research, particularly since they have been shown to affect academic achievement (Malecki & Elliott, 2002).

When evaluating setting characteristics, most participants taught in a preschool or elementary setting and delivered academic instruction, particularly reading or language arts instruction. The literature included in this study that was conducted in secondary settings is minimal, which is reflective of research in the field of special

education (Wong et al., 2013). Thus, more research in secondary settings involving video analysis is warranted. Regarding the instructional arrangement, most educators provided instruction in a one-on-one or small group instructional arrangement, which is not surprising considering that over half of the educators taught in a self-contained or resource classroom. Including studies with more diverse settings in future meta-analyses can broaden the conclusions that can be drawn regarding the effectiveness of video analysis, as well as assist administrators and practitioners in choosing appropriate settings in which to conduct this intervention.

Additional research questions investigated in this study related to whether or not differential effects existed based on participant and instructional characteristics. Of the four potential participant characteristic moderators analyzed, only role was found to have statistically significant differences among the subgroups, with paraprofessionals showing the largest effects. One possible reason for this finding is the high number paraprofessionals who held college degrees. Of the 66 AB contrasts with paraprofessionals as participants, 45 of these held a Bachelor's degree or higher. Additionally, the paraprofessionals in the included studies were experienced, with more than a third having three or more years of teaching experience. The finding that paraprofessionals had the largest effects is encouraging as paraprofessionals are typically not compensated for attending professional development events after school hours, and therefore, may not receive as many professional development opportunities as teachers (Brock & Carter, 2015). Because paraprofessionals can watch recorded videos of themselves during times in the school day when they are not providing instruction to

students, they can potentially improve their instructional skills without the need for professional development outside of school hours.

Considering instructional characteristics, video analysis had moderate to strong effects for all subgroups analyzed with no statistically significant differences found among group sizes, types of instruction, grade levels, or settings. This finding is positive as it suggests that video analysis is effective when implemented in a range of settings with different instructional characteristics; however, these results should be viewed with caution as a nonsignificant *p*-value does not necessarily mean the true effects do not vary (Borenstein et al., 2009). Regardless, the moderate to strong effects found among the subgroups supports the use of video analysis with educators across all grade levels, settings, and instructional characteristics.

Limitations

There are several limitations which should be noted. First, while publication bias was investigated by using publication type as a potential moderator, no other tests of publication bias were conducted. Therefore, the possibility that publication bias exists in this data set cannot be ruled out. Second, only single-case research was included in this meta-analysis; as such, the entire body of literature on video analysis is not represented in these results. Lastly, while there were no statistically significant differences between the studies that met WWC Design Quality Standards (U.S. Department of Education, 2016) with reservations and those that did not meet the standards, there is the possibility that the data from the studies that did not meet standards are not valid. For example, the reasons that studies did not meet standards included (a) not including the minimum

number of data points in baseline or intervention; (b) not collecting inter-observer agreement data, not taking inter-observer agreement data on a minimum of 20% of the data points, or not meeting minimum quality thresholds for inter-observer agreement results, and (c) not using a single-case design that is capable of demonstrating experimental control. Failing to meet each of these standards poses a threat to internal validity; however, it could be argued that there are still threats to validity even among studies that meet design quality standards, such as those proposed by the WWC (U.S. Department of Education, 2016), or meet them with reservations. For example, even if a researcher does collect inter-observer agreement data on the requisite 20% of data points and those results show high agreement, there is still the possibility that low agreement would be found on the rest of the data points, thus posing a threat to the validity of the results. Additionally, while not included in the WWC Design Quality Standards (U.S. Department of Education, 2016), the collection and reporting of implementation fidelity data results is considered by some to be an important component of high-quality designs (Council for Exceptional Children, 2014; Horner et al., 2005) and neglecting to report these data could pose a threat to internal validity. Thus, even studies that meet WWC Design Quality Standards (U.S. Department of Education, 2016) with reservations could still include design flaws that threatened the internal validity of the results.

Conclusions

Overall, video analysis appears to be effective for a variety of educators and under a variety of circumstances. This finding is promising as prior research has shown that the lecture style of professional development—the most commonly used form of

professional development for educators—is generally not effective. Considering the cost-effectiveness, ease of implementation, and ability to individualize the professional development to the needs of the educator, video analysis has the potential to replace the historically used lecture model with a more effective method of improving educators' instructional skills. Being able to implement the intervention within their own classroom, to select their own behaviors to improve, and to be in charge of the decision-making process are all aspects of video analysis that may make it attractive to educators as an authentic form of professional development.

CHAPTER IV

SUMMARY AND CONCLUSIONS

This dissertation aimed to investigate the effects of video analysis on the instructional practices of special educators through two studies. The first study used a series of single-case experimental designs to investigate the effects of video analysis on the self-selected instructional behaviors of special educators. The second study consisted of a systematic search of the literature and used meta-analytic procedures to investigate the overall effects of video analysis on the instructional skills of special educators, as well as potential moderators related to participant and instructional characteristics. The second study also summarized narrative data on the included studies and answered questions about the status of the literature base of video analysis regarding study characteristics, participant characteristics, student characteristics, and setting characteristics.

The first study used a multiple-baseline across participants and changing criterion designs to answer the research question, "What are the effects of video analysis on the self-identified instructional practices of educators?" Generalization, maintenance, and social validity data were also collected to determine if (a) the results of video analysis on one instructional practice would generalize to another instructional practice, (b) educators would maintain their improved rates of behavior over time, (c) educators would find video analysis feasible, and (d) educators' views of video analysis would change over the course of implementation. The results of this study were largely

positive. The results of visual and statistical analyses showed that the use of video analysis produced moderate to strong effects for all participants and that these results generalized to additional target behaviors. The positive effects demonstrated in intervention maintained for four of the five participants for the primary target behavior and for all five participants for the secondary target behavior. The results of the social validity survey were also positive and indicated the participants viewed video analysis more favorably as the study progressed. In particular, results of a paired *t*-test revealed statistically significant results between the participants' responses on the first and last administration of the survey for two questions—"Video analysis is a cost-effective way to improve my teaching skills" and "Video analysis is worth the time invested."

The second study systematically reviewed the single-case research base on video analysis to answer the question, "What is the status of the literature base on video analysis regarding study characteristics (i.e., publication type, design quality), participant characteristics (i.e., role, education level, experience level, and age), student characteristics (i.e., disability type and collection of student outcomes), and setting characteristics (i.e., grade level, group size, type of instruction, and setting)?" This study also calculated Tau-U (Parker et al., 2011) effect sizes to determine (a) the effect publication type and methodological quality have on the effectiveness of video analysis, (b) the magnitude of effect of video analysis on the instructional practices of educators, (c) the effects that participant characteristics (e.g., role, education level, experience level, age) have on the effectiveness of video analysis, and (d) the effects that instructional characteristics (i.e., group size, type of instruction, grade level, setting) have on the

effectiveness of video analysis. Results of Tau-U analyses (Parker et al., 2011) indicated that video analysis is effective for changing the instructional practices of educators. Moderator analyses also revealed that neither publication type nor methodological quality impacted the results. Additionally, Tau-U effect sizes (Parker et al., 2011) were moderate to large for every variable category and subgroup analyzed, although only role was found to have statistically significant differences among the subgroups.

Implications for Practice

The results of this work revealed several implications for practice. First, it is encouraging that video analysis was found to be effective for changing the instructional practices of educators. Given the negative aspects of the traditional, lecture-based model of professional development, namely that it generally does not produce positive results when used to change educators' instructional practices, video analysis may be a viable alternative. The positive aspects of video analysis, such as being cost-effective, easy to implement, and effective, makes it an appealing choice for educators. Additionally, the finding that there were no statistically significant differences found between any of the subgroups for the potential moderator categories tested, other than role, is positive. These results indicate that video analysis may be equally effective for a variety of educators with different backgrounds, as well as for educators who teach in a variety of instructional settings. Lastly, conducting a single-case study with educators revealed several considerations for supervisors wishing to replicate the experience with educators under their supervision, including (a) the use of a cloud storage server to minimize the use of disc space on educators' recording devices, (b) connecting directly to the internet,

"zipping" a file prior to uploading, or reducing the length of the video to minimize the time it takes to upload videos to a cloud server, and (c) having a plan in place for anticipated technical issues.

Limitations

There were several limitations in this research which should be noted. For the first study, the use of total count recording rather than interval by interval recording negatively impacted the inter-observer agreement results for several participants. Also, one participant had a lower dose of the intervention due to only being able to record once per week, which may have negatively impacted her results. Further, due to time constraints, student outcome and extended maintenance data were not collected. Finally, the administration of social validity surveys were not anonymous, which may have influenced the participants' responses.

Limitations were also noted in the meta-analysis. Because publication bias was only tested through a heterogeneity analysis of peer-reviewed studies and dissertations, it is possible that publication bias exists in this data set. Additionally, the entire body of literature on video analysis is not represented in these results as only single-case research was included. Finally, while there is the possibility that design flaws inherent in the studies that did not meet WWC Design Quality Standards (U.S. Department of Education, 2016) may have adversely impacted the internal validity of the results from those studies, this risk cannot be ruled out from any study and thus was not used as a justification to exclude these studies from analyses, particularly since moderator analyses showed no statistically significant results between the subgroups of studies that

did and did not meet the WWC Design Quality Standards (U.S. Department of Education, 2016).

Future Research

Analyzing the status of the literature base on video analysis revealed several areas in need of additional research, including well-designed, peer-reviewed studies. Many of the studies included in the meta-analysis were dissertations and/or did not meet minimum WWC Design Quality Standards (U.S. Department of Education, 2016), and while a large percentage of studies did meet the WWC Design Quality Standards (U.S. Department of Education, 2016) with reservations, none of the included studies met the standards without reservations. Additional areas in need of further research include studies that evaluate the effectiveness of video analysis with pre-service teachers, novice educators, and educators ages 40 years old and older. Student outcomes should also be evaluated in future research, particularly outcomes related to the social and behavioral skills of students. Finally, more research is needed in middle school and high school settings and in academic subjects other than reading or language arts.

REFERENCES

*Indicates documents that were included in the meta-analysis.

*Ahuja, R. S. (2000). *The relative effectiveness of two different peer coaching programs and a structured self-evaluation program with elementary teachers using direct instruction* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 9978924)

*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.

Alliance for Excellent Education (2014). *On the path to equity: Improving the effectiveness of beginning teachers.* Retrieved from http://all4ed.org/reports-factsheets/path-to-equity/

Artman-Meeker, K. M., & Hemmeter, M. L. (2012). Effects of training and feedback on teachers' use of classroom preventive practices. *Topics in Early Childhood Special Education, 33,* 112-123. doi:10.1177/0271121412447115

Bernard, R. M., Borokhovski, E., Schmid, R. F., & Tamin, R. M. (2014). An exploration of bias in meta-analysis: The case of technology integration research in higher education. *Journal of Computing in Higher Education, 26,* 183-209. doi:10.1007/s12528-014-9084-z

Billingsley, B. S. (2004). Special education teacher retention and attrition: A critical analysis of the research literature. *The Journal of Special Education, 38,* 39-55.

*Bingham, M. A., Spooner, F., & Browder, D. (2007). Training paraeducators to promote the use of augmentative and alternative communication by students with significant disabilities. *Education and Training in Developmental Disabilities, 42, 339-352.*

*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. doi:10.1177/0271121415594797

Black, A. (2010). Gen Y: Who they are and how they learn. *Educational Horizons, 88,* 92-101.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive Meta-Analysis (Version 3) [Computer software]. Retrieved from https://www.meta-analysis.com/

Borenstein, M. Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* West Sussex, UK: John Wiley & Sons, Ltd.

Borko, H. (2004). Professional development and teacher learning: Mapping the terrain. *Educational Researcher, 33*(8), 3-15. doi:10.3102/0013189X033008003

Borman, G. D., & Dowling, N. M. (2008). Teacher attrition and retention: A meta-analytic and narrative review of the research. *Review of Educational Research, 78,* 367-409. doi:10.3102/003465430832455

*Bose-Deakins, J. E. (2006). *Increasing early literacy instruction behavior of Head Start teachers using videotape consultation* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3199456)

Boudah, D. J., Blair, E., & Mitchell, V. J. (2003). Implementing and sustaining strategies instruction: Authentic and effective professional development or "business as usual"? *Exceptionality, 11,* 3-23. doi:10.1207/S15327035EX1101_2

Bowman-Perrott, L., Burke, M. D., Zaini, S., Zhang, N., & Vannest, K. (2016). Promoting positive behavior using the good behavior game: A meta-analysis of single-case research. *Journal of Positive Behavior Interventions, 18,* 180-190. doi:10.1177/1098300715592355

Brock, M. E., & Carter, E. W. (2015). Effects of a professional development package to prepare special education paraprofessionals to implement evidence-based practice. *The Journal of Special Education, 49,* 39-51. doi:10.1177/0022466913501882

Brossart, D. F., Vannest, K. J., Davis, J. L., & Patience, M. A. (2014). Incorporating nonoverlap indices with visual analysis for quantifying intervention effectiveness in single-case experimental designs. *Neuropsychological Rehabilitation, 24,* 464–491. doi:10.1080/09602011.2013.868361

*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. doi:10.1177/0888406409360012

Cardinal, J. R. (2012). *"Rethink Autism": Effectiveness of web-based applied behavior video modeling program on the performance of paraeducators and students with autism spectrum disorder* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3503058)

*Carnine, D. W., & Fink, W. T. (1978). Increasing the rate of presentation and use of signals in elementary classroom teachers. *Journal of Applied Behavior Analysis, 11,* 35-46.

Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., & Test, D. W. (2015). CEC's standards for classifying the evidence base of practices in special education. *Remedial and Special Education, 36,* 220-234. doi:10.1177/0741932514557271

Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Los Angeles, CA: Sage.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied Behavior Analysis* (2nd ed.). Upper Saddle River, NJ: Pearson.

Corcoran, T. B. (2007). Teaching Matters: How state and local policymakers can improve the quality of teachers and teaching. *CPRE Policy Briefs.* Retrieved from http://repository.upenn.edu/cpre_policybriefs/36

Council for Exceptional Children. (2014). *Council for Exceptional Children standards for evidence-based practices in special education*. Retrieved from http://www.cec.sped.org/Standards/Evidence-Based-Practice-Resources-Original

Courtemanche, A., Sheldon, J., Sherman, J., Schroeder, S., Bell, A., & House, R. (2014). Assessing the effects of a staff training package on the treatment integrity of an intervention for self-injurious behavior. *Journal of Developmental and Physical Disabilities, 26,* 371-389. doi:10.1007/s10882-014-9372-6

Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, A. (2015). Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with expert and novice teachers. *International Journal of Science and Math Education, 13,* 389-403. doi:10.1007/s10763-014-9610-5

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

*Digennaro-Reed, F. D., Codding, 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. doi:10.1901/jaba.2010.43-291

Dijkers, M. P.J.M. for the NCDDR Task Force on Systematic Review and Guidelines. (2009). *When the best is the enemy of the good: The nature of research evidence used in systematic reviews of guidelines.* Austin, TX: SEDL.

Duker, P. C., Hensgens, Y., & Venderbosch, S. (1995). Effectiveness of delayed feedback on the accuracy of teaching communicative gestures to individuals with severe mental retardation. *Research in Developmental Disabilities, 16,* 479-488.

*Englund, L. W. (2011). *Evaluating and improving the quality of teacher's language modeling in early childhood classrooms* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3439991)

*Erbas, D., Tekin-Iftar, E., & Yucesoy, S. (2006). Teaching special education teachers how to conduct functional analysis in natural settings. *Education and Training in Developmental Disabilities, 41,* 28-36.

Every Student Succeeds Act (ESSA) of 2015, Pub.L. 114–95

*Fedders, A. M. (2012). *The effect of video self-monitoring on novice special educators' implementation of advanced direction instruction reading techniques* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3495674)

Ferguson, C. J., & Brannick, M. T. (2012). Publication bias in psychological science: Prevalence, methods for identifying and controlling, and implications for the use of meta-analyses. *Psychological Methods, 17,* 120-128. doi:10.1037/a0024445

Ford, J. E. (1984). A comparison of three feedback procedures for improving teaching skills. *Journal of Organizational Behavior Management, 6,* 65-77.

*Fullerton, E. K., Conroy, M. A., & Correa, V. I. (2009). Early childhood teachers' use of specific praise statements with young children at risk for behavioral disorders. *Behavioral Disorders, 34,* 118-135.

Gage, N. A., & Lewis, T. J. (2014). Hierarchical linear modeling meta-analysis of single-subject design research. *The Journal of Special Education, 48,* 3-16. doi:10.1177/0022466912443894

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

GetData Graph Digitizer (2013; version 2.26). [Computer software]. Retrieved from http://www.getdata-graph-digitizer.com/

Gray, L., & Taie, S. (2015). Public school teacher attrition and mobility in the first five years: Results from the first through fifth waves of the 2007-2008 beginning teacher longitudinal study (NCES 2015-337). U.S. Department of Education. Washington, DC: National Center for Education Statistics. Retrieved from http://nces.ed.gov/pubsearch

*Hager, K. D. (2012). Self-monitoring as a strategy to increase student teachers' use of effective teaching practices. *Rural Special Education Quarterly, 31*(4), 9-17.

Hanushek, E. (1970, February). *The production of education, teacher quality, and efficiency.* Paper presented at the Bureau of Educational Personnel Development Conference, Washington, D. C.

Harwell, S. H. (2003). *Teacher professional development: It's not an event, it's a process*. Waco, TX: CORD.

*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. doi:10.1177/1098300709

Herbert, J. M. (1993). Intervening to influence teachers' actions and students' involvement. *Journal of Personnel Evaluation in Education, 6,* 249-261.

Hong, E. R., Ganz, J. B., Gilliland, W., & Ninci, J. (2014). Teaching caregivers to implement an augmentative and alternative communication intervention to an adult with ASD. *Research in Autism Spectrum Disorders, 8,* 570-580. doi:10.1016/j.rasd.2014.01.012

Hong, E. R., Ganz, J. B., Mason, R., Morin, K., Davis, J. D., Ninci, J.,…& Gilliland, W. D. (2016). The effects of video modeling in teaching functional living skills to persons with ASD: A meta-analysis of single-case studies. *Research in Developmental Disabilities, 57,* 158-169. doi:10.1016/j.ridd.2016.07.001

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.

Hutchins, T. L., & Prelock, P. A. (2013). The social validity of Social Stories™ for supporting the behavioral and communicative functioning of children with autism spectrum disorder. *International Journal of Speech-Language Pathology, 15,* 383–395. doi:10.3109/17549507.2012.743174

Ingersoll, R., Merrill, L., & May, H. (2014). *What are the effects of teacher education and preparation on beginning teacher attrition?*. Research Report (#RR-82). Philadelphia: Consortium for Policy Research in Education, University of Pennsylvania.

*Kaiser, A. P., Ostrosky, M. M., & Alpert, C. L. (1993). Training teachers to use environmental arrangement and Milieu teaching with nonvocal preschool children. *Journal of the Association for Persons with Severe Handicaps, 18,* 188-199.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). New York: Oxford University Press, Inc.

Kirk-Martinez, J. F. (2011). *Coaching in special education: Toward a model of differentiated professional development for elementary school paraeducators* (Doctoral dissertation). Retrieved from ProQuest. (UMI 3495682)

Klein, L. A., Houlihan, D., Vincent, J. L., & Panahon, C. J. (2017). Best practices in utilizing the changing criterion design. *Behavior Analysis in Practice, 10,* 52-61. doi:10.1007/s40617-014-0036-x

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from: http://files.eric.ed.gov/fulltext/ED510743.pdf

*Lambour, G. P. (1976). *Variations of videotape feedback as a mechanism of behavior change* (Doctoral dissertation). Retrieved from ProQuest. (Accession No. 1977-30246-001)

Lequia, J., Wilkerson, K. L., Kim, S., & Lyons, G. L. (2015). Improving transition behaviors in students with autism spectrum disorders: A comprehensive evaluation of interventions in educational settings. *Journal of Positive Behavior Interventions, 17,* 146-158. doi:10.1177/1098300714548799

*Lindsey, R. (2014). *Increasing the use of prompting strategies: A multiple baseline study across pairs of paraeducators of students with disabilities* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3579519)

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis.* Applied Social Research Methods Series, Volume 49. Bickman, L., & Rog, D. J. (Eds.). Thousand Oaks, CA: Sage Publications.

Losinski, M., Cuenca-Carlino, Y., Zablocki, M., & Teagarden, J. (2014). Examining the efficacy of self-regulated strategy development for students with emotional or behavioral disorders: A meta-analysis. *Behavioral Disorders, 40*(1), 52-67.

*Lynes, M. J. (2013). *The effects of self-evaluation with video on the use of oral language development strategies by preschool teachers* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 3505030)

Malecki, C. K., & Elliott, S. N. (2002). Children's social behaviors as predictors of academic achievement: A longitudinal analysis. *School Psychology Quarterly, 17,* 1-23.

Milburn, T. F., Girolametto, L., Weitzman, E., & Greenberg, J. (2014). Enhancing preschool educators' ability to facilitate conversations during shared book reading. *Journal of Early Childhood Literacy, 14,* 105-140. doi:10.1177/1468798413478261

*Morgan, R. L., Menlove, R., Salzberg, C. L., & Hudson, P. (1994). Effects of peer coaching on the acquisition of direct instruction skills by low-performing preservice teachers. *The Journal of Special Education, 28,* 59-76.

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. doi:10.1177/0888406413501090

Neely, L. C., Ganz, J. B., Davis, J. L., Boles, M. B., Hong, E. R., Ninci, J., & Gilliland, W. D. (2016). Generalization and maintenance of functional living skills for individuals with autism spectrum disorder: A review and meta-analysis. *Review Journal of Autism and Developmental Disorders, 3*, 37-47. doi:10.1007/s40489-015-0064-7

Nichols, A. S., & Sosnowsky, F. L. (2002). Burnout among special education teachers in self-contained cross-categorical classrooms. *Teacher Education and Special Education, 25*, 71-86.

Ninci, J., Neely, L. C., Hong, E. R., Boles, M. B., Gilliland, W. D., Ganz, J. B., ... & Vannest, K. J. (2015). Meta-analysis of single-case research on teaching functional living skills to individuals with ASD. *Review Journal of Autism and Developmental Disorders, 2*, 184-198. doi:10.1007/s40489-014-0046-1

Nir, A. E., & Bogler, R. (2008). The antecedents of teacher satisfaction with professional development programs. *Teaching and Teacher Education, 24*, 377-386. doi:10.1016/j.tate.2007.03.002

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107 110, § 115, Stat. 1425

Odden, A., Archibald, S., Fermanich, M., & Gallagher, H. A. (2002). A cost framework for professional development. *Journal of Education Finance, 28*, 51-74.

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single case research: a review of nine nonoverlap techniques. *Behavior Modification, 35,* 303–322. doi:10.1177/0145445511399147

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. doi:10.1016/j.beth.2010.08.006

*Peck, C. A., Killen, C. C., & Baumgart, D. (1989). Increasing implementation of special education instruction in mainstream preschools: Direct and generalized effects of nondirective consultation. *Journal of Applied Behavior Analysis, 22,* 197-210.

*Pelletier, K., McNamara, B., Braga-Kenyon, P., & Ahearn, W. H. (2010). Effect of video self-monitoring on procedural integrity. *Behavioral Interventions, 25,* 261-274. doi:10.1002/bin.316

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What makes professional development effective? Strategies that foster curriculum implementation. *American Educational Research Journal, 44,* 921-958.

*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.

Quick, H. E., Holtzman, D. J., & Chaney, K. R. (2009). Professional development and instructional practice: Conceptions and evidence of effectiveness. *Journal of Education for Students Placed at Risk, 14,* 45-71. doi:10.1080/10824660802715429

*Reamer, R. B. (1996). *The effects of video self-modeling and self-assessment training on the delivery of instructional strategies provided by paraeducators working with children with developmental delays* (Doctoral dissertation). Retrieved from ProQuest. (UMI No. 9542615)

Rice, J. K. (2010). The impact of teacher experience: Examining the evidence and policy implications. *Brief, 11*. Washington, D.C.: Center for Analysis of Longitudinal Data in Education Research. Retrieved from http://www.caldercenter.org/publications/impact-teacher-experience-examining-evidence-and-policy-implications

*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. doi:10.1177/1088357611407063

Ross, J. A. & Bruce, C. D. (2007a). Professional development effects on teacher efficacy: Results of a randomized field trial. *Journal of Educational Research, 101*(1), 50-60.

Rule, S. (1973). *A comparison of three different types of feedback on teachers' performance* (Doctoral dissertation). Retrieved from ProQuest. (Accession No. 1973-27843-001)

*Saudargas, R. A. (1973). *Setting criterion rates of teacher academic approvals: The effects of videotape feedback with behavior analysis follow through teachers*

(Doctoral dissertation). Retrieved from ProQuest. (Accession No. 1973-12067-001)

Shadish, W. R., Brasil, I. C. C., Illingworth, D. A., White, K. D., Galindo, R., Nagler, E. D., & Rindskopf, D. M. (2009). Using Ungraph to extract data from image files: Verification of reliability and validity. *Behavior Research Methods, 41,* 177-183.

Sharpe, T., Spies, R., Newman, D., & Spickelmier-Vallin, D. (1996). Assessing and improving the accuracy of in-service teachers' perceptions of daily practice. *Journal of Teaching in Physical Education, 15,* 297-318.

Shockley, R., Guglielmino, P., Watlington, E. (2006). *National crisis in teacher education: What are the costs?* Retrieved from http://images.pearsonassessments.com/images/NES_Publications/2006_11ShockleyGuglielminoWatlington_558_1.pdf

Sloat, K. C. M., Tharp, R. G., & Gallimore, R. (1977). The incremental effectiveness of classroom-based teacher-training techniques, *8,* 810-818.

*Snyder, C. K. C. (2013). *Effects of training on early childhood special education paraeducators' use of early literacy strategies during book reading* (Doctoral dissertation). Retrieved from ProQuest. (UMI 3553235)

*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. doi:10.1177/0888406410384407

Tripp, T., & Rich, P. (2012). Using video to analyze one's own teaching. *British Journal of Educational Technology, 43,* 678-704. doi:10.1111/j.1467-8535.2011.01234.x

U.S. Department of Education (2011). *Our Future, Our Teachers: The Obama Administration's Plan for Teacher Education Reform and Improvement.* Washington, D.C.: Authors. Retrieved from https://www.ed.gov/teaching/our-future-our-teachers

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2016, September). *What Works Clearinghouse: Procedures and Standards Handbook (Version 3.0).* Retrieved from http://whatworks.ed.gov

Vannest, K. J., Davis, J. L., Davis, C. R., Mason, B. A., & Burke, M. D. (2010). Effective intervention for behavior with a daily behavior report card: A meta-analysis. *School Psychology Review, 39,* 654-672.

Vannest, K. J., Parker, R. I., Gonen, O., & Adiguzel, T. (2016). Single case research: Web-based calculators for SCR analysis. (Version 2.0) [Web-based application]. College Station, TX: Texas A&M University. Available from www.singlecaseresearch.org

Venn, M. L., & Wolery, M. (1992). Increasing day care staff members' interactions during caregiving routines. *Journal of Early Intervention, 16,* 304-319.

Vuran, S., & Gul, S. O. (2012). On-the-job training of special education staff: Teaching the simultaneous prompting strategies. *Educational Sciences: Theory & Practice, 12,* 2101-2110.

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

*Westover, J. M. (2011). *Increasing the literacy skills of students who require AAC through modified direct instruction and specific instructional feedback* (Doctoral dissertation). Retrieved from ProQuest. (UMI 3420436)

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. doi:10.1177/1744629514552305

Wong, C., Odom, S. L., Hume, K., Cox, A. W., Fettig, A., Kucharczyk, S.,…Schultz, T. R. (2013). *Evidence-based practices for children, youth, and young adults with Autism Spectrum Disorder*. Chapel Hill: The University of North Carolina, Frank Porter Graham Child Development Institute, Autism Evidence-Based Practice Review Group. Retrieved from http://autismpdc.fpg.unc.edu/sites/autismpdc.fpg.unc.edu/files/2014-EBP-Report.pdf

Wright, M., Ellis, D., & Baxter, A. (2012). The effect of immediate or delayed video-based teacher self-evaluation on head start teachers' use of praise. *Journal of Research in Childhood Education, 26,* 187-198.

Zan, B., & Donegan-Ritter, M. (2014). Reflecting, coaching and mentoring to enhance teacher-coach interactions in Head Start classrooms. *Journal of Early Childhood Education, 42*, 93-104. doi:10.1007/s10643-013-0592-7

APPENDIX A

SINGLE-CASE STUDY FOREST PLOT

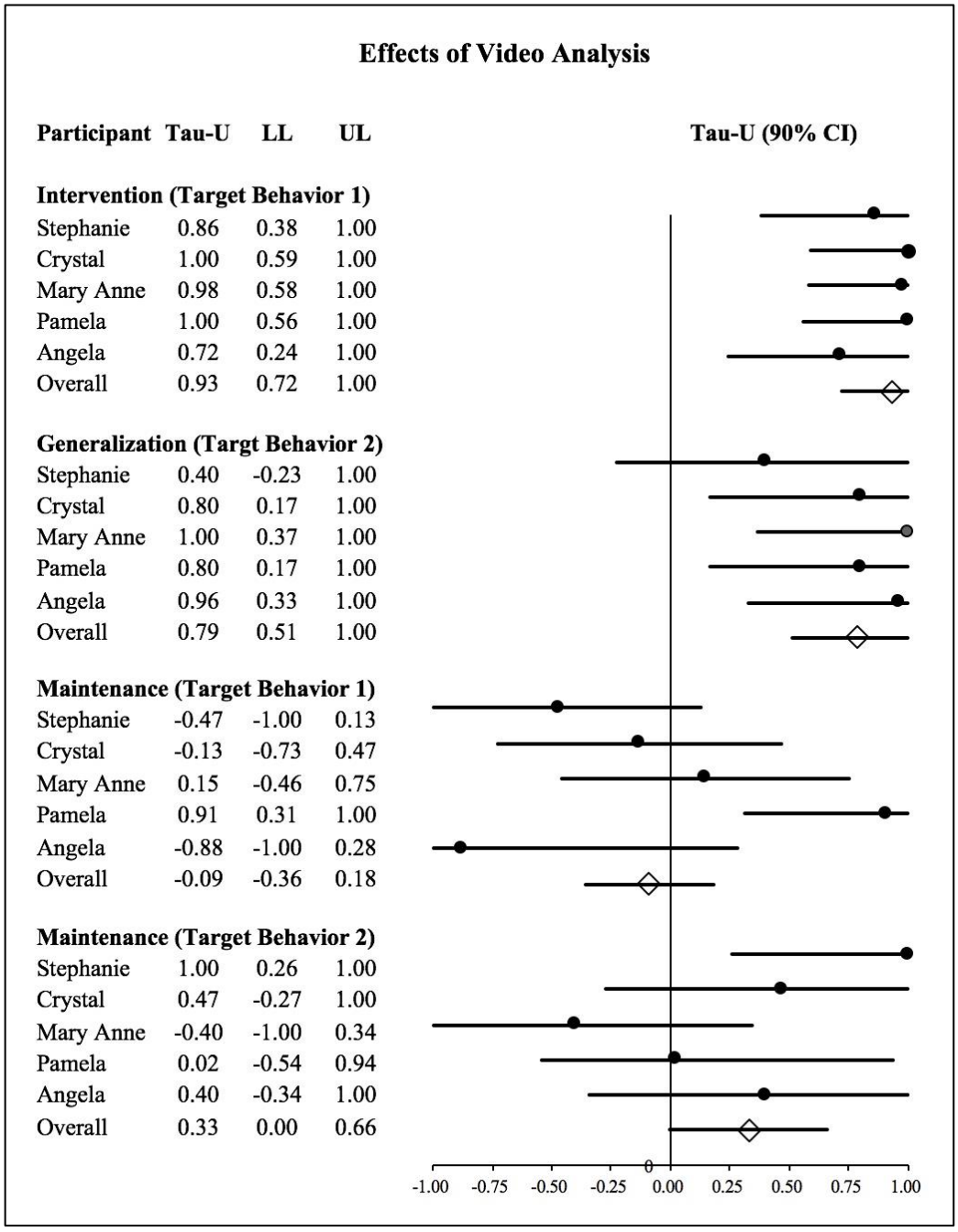

Figure A1. Forest plot and Tau-U effect size analyses for the effects of video analysis on selected target behaviors

APPENDIX B

SINGLE-CASE GRAPHS OF VIDEO ANALYSIS OUTCOMES

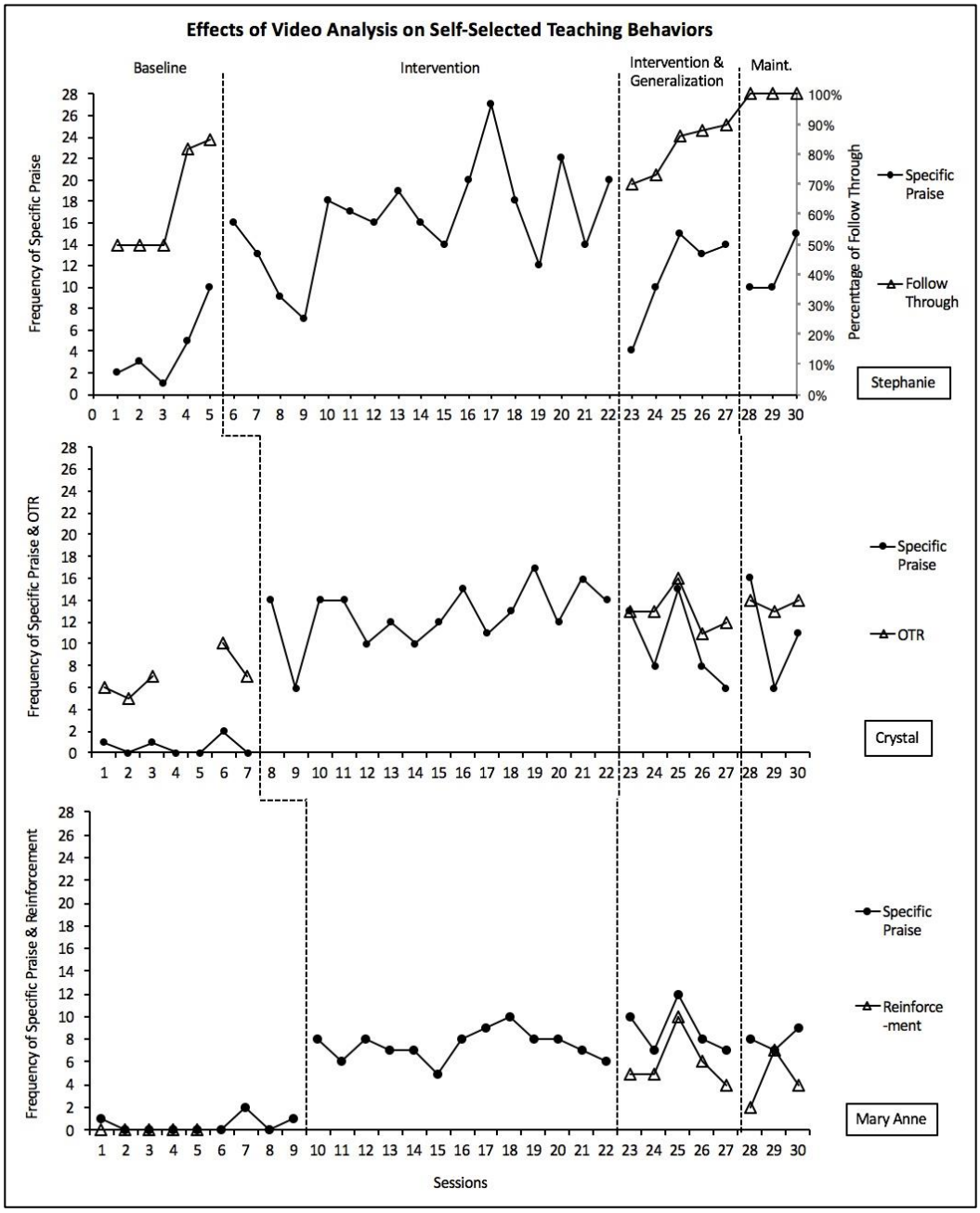

Figure B1. Effects of video analysis on the percentage of follow-through and the frequency of specific praise, reinforcement, and opportunities to respond (OTR) for Stephanie, Crystal, and Mary Anne.

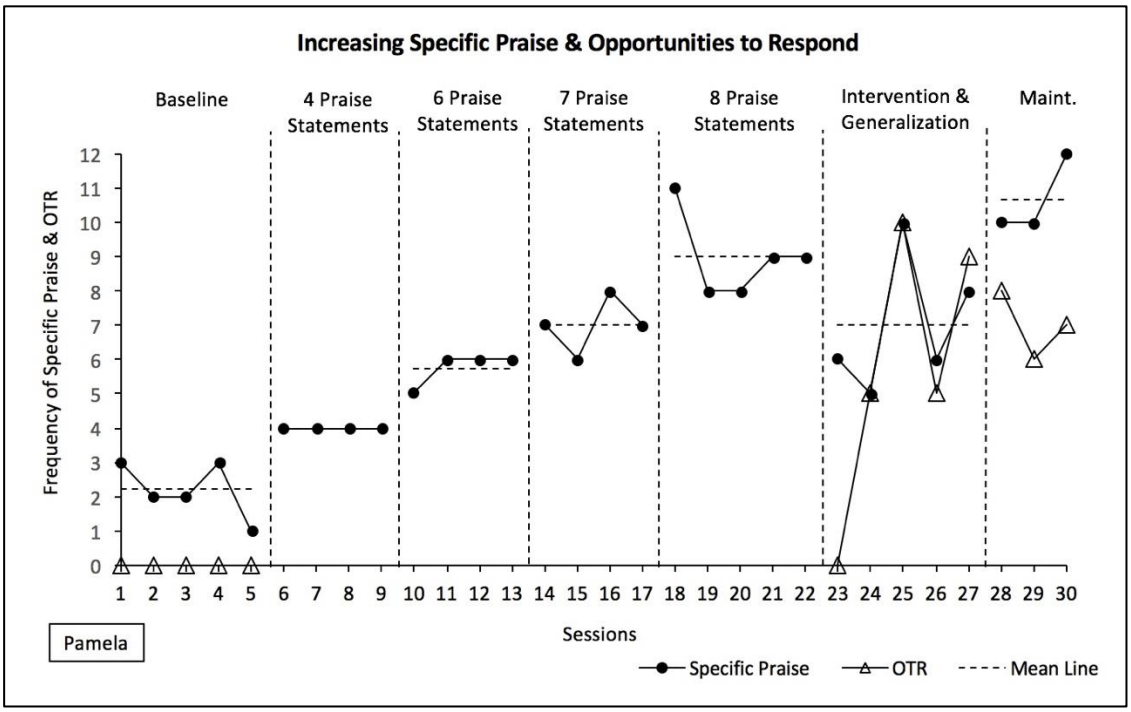

Figure B2. Effects of video analysis on the frequency of praise and opportunities to respond (OTR) for Pamela.

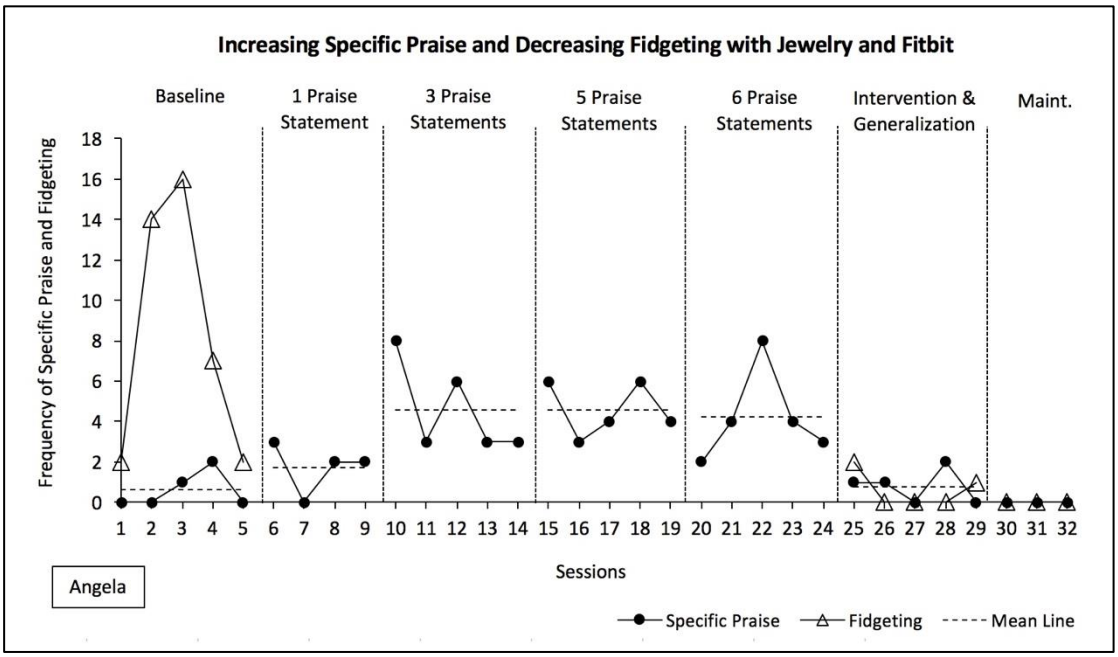

Figure B3. Effects of video analysis on the frequency of specific praise and fidgeting for Angela

APPENDIX C

SOCIAL VALIDITY SURVEY RESULTS

Table C1
Social Validity Responses

	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
Time 1										
Mean	4.4	3.4	3.6	3.8	3.0	2.8	3.8	3.4	3.0	3.8
Range	3-5	2-4	3-4	3-5	2-5	2-4	3-4	3-4	2-4	3-5
Time 2										
Mean	4.6	3.6	3.6	4.6	4.2	3.4	4.4	4.2	3.4	4.2
Range	4-5	2-5	2-5	4-5	4-5	1-5	3-5	3-5	2-5	3-5
Time 3										
Mean	4.4	4.0	3.8	4.6	4.2	3.2	4.2	4.2	3.2	4.2
Range	4-5	3-5	2-5	4-5	3-5	1-4	3-5	3-5	2-5	3-5
Time 4										
Mean	4.4	3.8	4.0	4.6	4.4	3.2	4.4	4.0	3.8	4.2
Range	3-5	2-5	3-5	4-5	4-5	1-4	3-5	2-5	2-5	3-5
Time 5										
Mean	4.6	3.8	4.0	4.8	4.4	3.4	4.4	4.4	3.6	4.4
Range	4-5	2-5	3-5	4-5	4-5	1-5	3-5	3-5	2-5	3-5
Overall	4.48	3.72	3.80	4.48	4.04	3.20	4.24	4.04	3.40	4.16

Note. S1 = video analysis can be used to improve many different teaching skills; S2 = video analysis is feasible to implement in my setting; S3 = the amount of time and effort it takes to implement video analysis is reasonable; S4 = video analysis is a cost effective way to improve my teaching skills; S5 = I was able to implement video analysis without much assistance from others; S6 = I don't mind watching myself on video; S7 = watching myself on video helped me to see things that I would not have noticed otherwise; S8 = video analysis is worth the time invested; S9 = I plan on continuing to use video analysis after this project is complete; S10 = Overall, I believe video analysis is an effective method for helping me improve my teaching skills; 5 = strongly agree; 4 = agree; 3 = unsure/neutral; 2 = disagree; 1 = strongly disagree

APPENDIX D

FLOWCHART

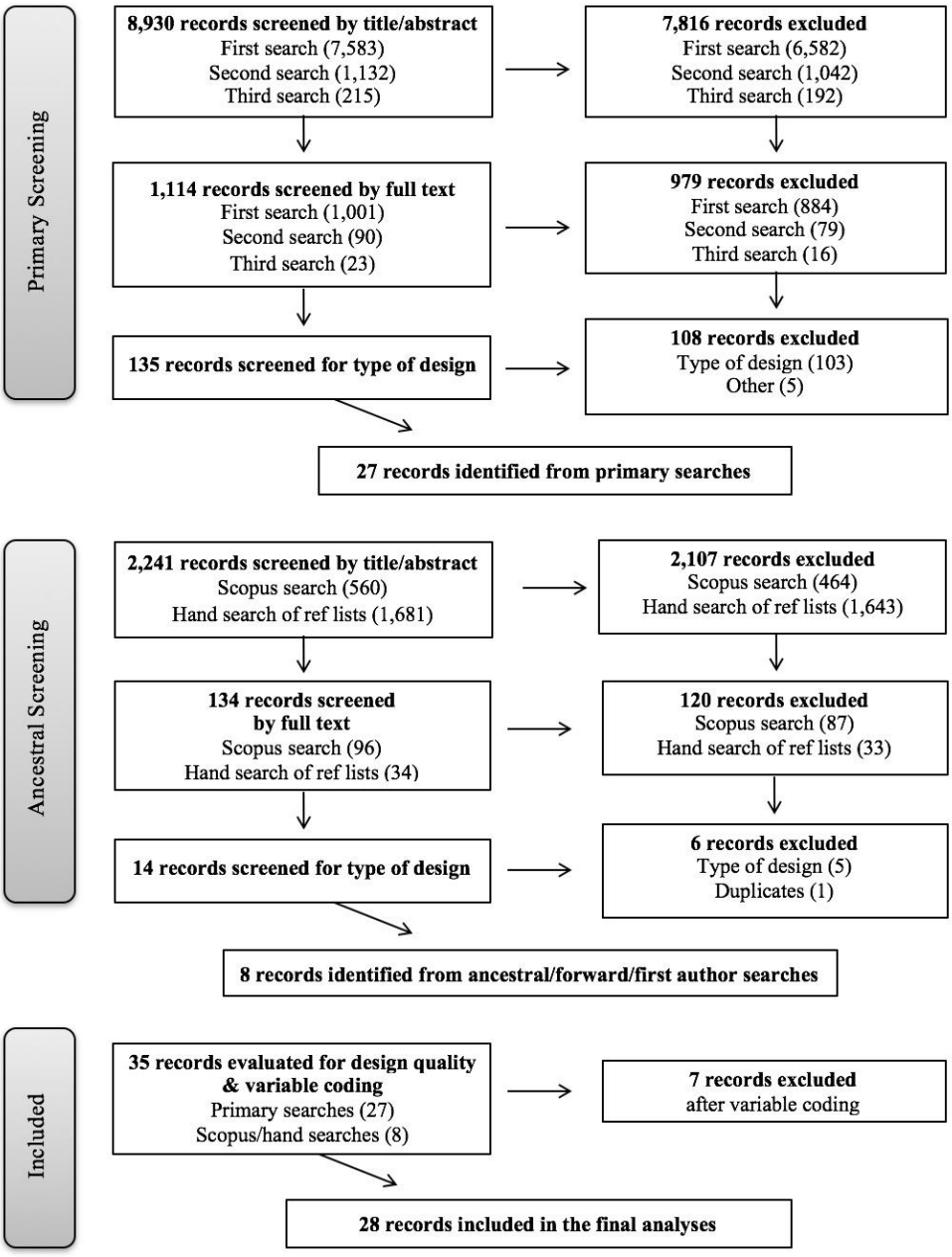

Figure D1. Flowchart indicating the number of articles excluded at each stage of the search process.

APPENDIX E

WWC BASIC DESIGN QUALTITY STANDARDS (U.S. DEPARTMENT OF EDUCATION, 2016) CODES AND OPERATIONAL DEFINITIONS

Table E1
Description of Single-Case Design Standards

Possible Score	Criteria for Score
Design Standard 1: Manipulation of the Independent Variable	
1	The study reported manipulation of the independent variable.
0	The study did not report manipulation of the independent variable
Design Standard 2A: IOA Reported	
1	The study reported an IOA score.
0	The study did not report an IOA score.
Design Standard 2B: IOA Frequency	
2	The study reported IOA for a minimum of 20% of the sessions within each condition.
1	The study reported IOA for a minimum of 20% of the sessions, but did not disaggregate the score by phase or condition.
0	The study reported IOA for less than 20% of the sessions.
Design Standard 2C: IOA Quality	
1	The IOA reported in the study met minimum quality thresholds (i.e., at least 80% for percentage agreement indices or 60% for *kappa* measures).
0	The IOA reported in the study did not meet minimum quality thresholds (i.e., less than 80% for percentage agreement indices or 60% for *kappa* measures).
Design Standard 3: Demonstration of Treatment Effects	
1	The study included a minimum of three attempts to demonstrate treatment effects at three different points in time. For alternating treatment designs, the study must include at least 2 conditions.
0	The study did not include a minimum of three attempts to demonstrate treatment effects at three different points in time or did not include at least two conditions (for alternating treatment designs).

Table E1
Continued

Possible Score	Criteria for Score
Design Standard 4: Number of Data Points Per Phase	
2	The study included at least 5 data points in baseline and intervention phases. For alternating treatment designs, the study included at least 5 data points per treatment for baseline and intervention phases.
1	The study included at least 3 data points in baseline and intervention phases. For alternating treatment designs, the study included at least 4 data points per treatment for baseline and intervention phases.
0	The study included less than 3 data points in baseline and intervention phases. For alternating treatment designs, the study included less than 4 data points per treatment for baseline and intervention phases.
Design Standard 5A (Multiple-Probe Designs Only): Initial Baseline Sessions	
2	The study included at least three consecutive data points within the first three sessions of baseline for each level.
1	The study included at least one data point within the first session of baseline for each level.
0	The study did not include at least one data point within the first session of baseline for each level.
Design Standard 5B (Multiple-Probe Designs Only): Probe Points Prior to Intervention	
2	The study included at least three consecutive data points immediately prior to introducing intervention for each level
1	The study included at least one data point immediately prior to introducing intervention for each level.
0	The study did not include at least one data point immediately prior to introducing intervention for each level.
Design Standard 5C (Multiple-Probe Designs Only): Additional Probe Point Considerations	
1	Each level that was still in baseline when intervention is introduced had a data point when the previous level(s) first received the intervention or when the previous level(s) reached the prespecified intervention criterion AND this data point is consistent in level and trend with the previous baseline data points in that level.

Table E1
Continued

Possible Score	Criteria for Score
Design Standard 5C (Multiple-Probe Designs Only): Additional Probe Point Considerations (Continued)	
0	Each level that was still in baseline when intervention was introduced did not have a data point when the previous level(s) first received the intervention or when the previous level(s) reached the prespecified intervention criterion OR this data point was not consistent in level and trend with the previous baseline data points in that level.
Overall Design Rating	
2	The study received the highest possible score for all Design Standards (e.g., a score of "2" for Design Standards 2B, and 4, and a score of "1" for Design Standards 1, 2A, 2C, and 3).
1	The study received a score of "1" for Design Standard 2B or 4 and received no scores of 0 for any of the Design Standards.
0	The study received a score of 0 on one or more Design Standards.

Note. These standards were adapted from What Works Clearinghouse Design Quality Standards (U.S. Department of Education, 2016); Design Standards 5A-5C were not used to determine the overall design rating; IOA = inter-observer agreement

APPENDIX F

VARIABLE CODES AND OPERATIONAL DEFINITIONS

Table F1
Operational Definitions for Variable Codes

Variable Code	Operational Definition
Role	
paraprofessional	consider aides, staff, etc. as paraprofessionals if they are supervised by an in-service teacher or lead teacher
pre-service teacher	individual enrolled in a teacher preparation program
in-service teacher	individual leads his/her own classroom or is the primary individual responsible for designing/delivering instruction
Group Size	
one-to-one	only one student received instruction
small group	a subset of students from the whole group received instruction; for example, small group activities, centers, etc.
large group	all students in the classroom received instruction; for example, circle time, calendar time, etc.
Type of Instruction	
academic	math, reading, language arts, literacy, writing; for example, letter identification, handwriting, reading comprehension, grammar, sentence structure, etc.
communication	for example, requesting, expanding communication, labeling, answering questions, using AAC, etc.
daily living skills	for example, eating, dressing, washing hands, toileting, cleaning, brushing teeth, cooking, preparing food, etc.
Grade Level	
preschool	Kindergarten and below; younger than 6 years old
elementary	grades 1-5; 6 years old to less than 12 years old
middle school	grades 6-8; 12 years old to less than 14 years old
high school	grades 9-12; 14 years old and older
Setting	
general education	none of the students in the class had a disability; assume general education if there is no mention of the students having a disability
self-contained	students with disabilities spent all their time in a special education classroom; includes separate/specialized schools for students with disabilities

Table F1
Continued

Variable Code	Operational Definition
resource	students with disabilities spent some time in a separate special education classroom, but also spent time in a general education classroom
inclusion	the classroom in which the educator taught included both students with and without disabilities

Note. Education, experience, and age were also coded, but are not included here because they are concrete and did not require operational definitions.

APPENDIX G

NARRATIVE TABLES

Table G1
Study Characteristics

Study	Pub Type	Design[a]	Design Quality	Participants	Contrasts	Pairs
Ahuja (2000)	DISS	MBD (3)	MWR, DNM	7	7	165
Alexander, Williams, & Nelson (2012)	PR	AB (2)	DNM	2	2	106
Bingham, Spooner, & Browder (2007)	PR	MPD (2)	DNM	3	6	318
Bishop, Snyder, & Crow (2015)	PR	MPD	MWR	2	2	432
Bose-Deakins (2006)	DISS	MPD	MWR	3	3	131
Capizzi, Wehby, & Sandmel (2010)	PR	MBD (3)	MWR	3	9	312
Carnine & Fink (1978)	PR	MBD (2)	MWR	3	6	1274
Digennaro-Reed, Codding, Cantania, & Maguire (2010)	PR	MBD	MWR	2	2	113
Englund (2011)	DISS	MBD (2)	MWR	6	6	180
Erbas, Tekin-Iftar, & Yucesoy (2006)	PR	MBD	MWR	6	6	4144
Fedders (2012)	DISS	MBD (2)	MWR	3	6	120
Fullerton, Conroy, & Correa (2009)	PR	MBD	DNM	3	3	110
Hager (2012)	PR	MBD	DNM	1	2	79
Hawkins & Heflin (2011)	PR	reversal (3)	MWR	3	6	235
Kaiser, Ostrosky, Alpert (1993)	PR	MBD (2)	MWR	1	2	264
Lambour (1976)	DISS	reversal (2), ABA, ABAC, MBD	MWR, DNM	7	9	665

Table G1
Continued

Study	Pub Type	Design[a]	Design Quality	Participants	Contrasts	Pairs
Lindsey (2014)	DISS	MBD (3)	DNM	8	24	2123
Lynes (2013)	DISS	MBD (2)	DNM	6	12	564
Morgan, Menlove, Salzberg, & Hudson (1994)	PR	MBD	MWR	5	5	420
Peck, Killen, & Baumgart (1989)	PR	MBD	MWR	3	6	614
Pelletier, McNamara, Braga-Kenyon, & Ahearn (2010)	PR	MBD	MWR	3	3	60
Pinter, East, & Thrush (2015)	PR	MBD	DNM	4	4	465
Reamer (1996)	DISS	MBD (3), AB	MWR, DNM	3	15	565
Robinson (2011)	PR	MPD	DNM	4	4	68
Saudargas (1973)	DISS	ABACBC (2), ABCACD, ABCDADED, ABABACD	DNM	5	9	1016
Snyder (2013)	DISS	MBD (2)	DNM	4	8	290
Stephenson, Carter, & Arthur-Kelly (2011)	PR	AB (2)	DNM	2	2	108
Westover (2011)	DISS	MBD (3)	MWR	3	9	2259

Note. PR = peer-reviewed; DISS = dissertation; MBD = multiple baseline design; MPD = multiple probe design; MWR = met with reservations; DNM = did not meet; for Ahuja (2000), figure 1 did not meet standards and figures 2 and 3 met with reservations; for Lambour (1975), Subjects 1 and 2 met with reservations, but all other participants did not meet standards; for Reamer (1995), figure 1 (rate of positive feedback) and figure 6 did not meet standards, but all other figures and dependent variables met with reservations; [a] indicates the number of experiments, when more than one, included in the study

Table G2
Participant Characteristics

Study	Participants	Age	Role	Education	Experience
Ahuja (2000)	SC1, SD1, SD2, SD3, SE1, SE2, SE3	NS	in-service	NS	1-4 years
Alexander, Williams, and Nelson (2012)	Susan, Rachel	NS	pre-service	Some college	none – 5 years
Bingham, Spooner, and Browder (2007)	Paras 1, 2, 3	20-52	para	High school diploma, GED	first year – 5 years
Bishop, Snyder, and Crow (2015)	Natalie, Brenda	24	in-service	Master's in Early Childhood, Bachelor's in Psychology	2 – 5 years
Bose-Deakins (2006)	Teachers A, B, C	31-37	in-service	Some college	2 – 10 years
Capizzi, Wehby, and Sandmel (2010)	Amy, Sarah, Scott	24-30	in-service, pre-service	Bachelor's in Early Childhood, Social Work, English, and Marketing	none – 2 years
Carnine and Fink (1978)	Teachers 1, 2, 3	NS	in-service, para	NS	NS
Digennaro-Reed, Codding, Cantania, & Maguire (2010)	Lauren, Shannon	28, 35	in-service	Master's in Education; Bachelor of Arts	first year – 4 years
Englund (2011)	Participants A/1, B/2, C/3, D/4, E/5, F/6	26-35	in-service	Some college, Bachelor of Education, Master of Education	2 – 15 years
Erbas, Tekin-Iftar, and Yucesoy (2006)	Ali, Ozge, Cicek, Gokhan, Ilknur, Oyku	NS	pre-service, in-service	Master of Arts, Bachelor's in SPED	3 – 14 years

Table G2
Continued

Study	Participants	Age	Role	Education	Experience
Fedders (2012)	Teachers 1, 2, 3	24-27	in-service	Education Specialist Credential Program	first year – 3rd year
Fullerton, Conroy, and Correa (2009)	Teachers 2, 3, 4	NS	in-service	Associate of Arts	6-13 years
Hager (2012)	Jennifer	23	pre-service	Some college	none
Hawkins and Heflin (2011)	Cantelli, Thomas, Williams	28-32	in-service	Master's in Special Education & Social Work	2 – 7 years
Kaiser, Ostrosky, and Alpert (1993)	Teacher A	NS	in-service	Bachelor's in Special Education	less than 3 years
Lambour (1976)	Experiment 1: Subjects 1, 2, 3, 4; Experiment 2: Subjects 1, 2, 3	23-31	in-service	NS	1 – 3 years
Lindsey (2014)	Paraeducators 1, 2, 3, 4, 5, 6, 7, 8	24-35	para	Bachelor of Arts, Master's in Special Education	first year – 4 years
Lynes (2013)	Teachers 1, 2, 3, 4, 5, 6	NS	in-service	Associate's Degree/Child Development Associate Credential, Bachelor's in Early Childhood & Psychology	2 – 28 years
Morgan, Menlove, Salzberg, and Hudson (1994)	Sharon, Cora, Linda, Candy, Nora	21-49	pre-service	Some college	none
Peck, Killen, and Baumgart (1989)	Ann, Alice, Carol	NS	in-service	Bachelor's in Early Childhood Education	NS

Table G2
Continued

Study	Participants	Age	Role	Education	Experience
Pelletier, McNamara, Braga-Kenyon, and Ahearn (2010)	Layla, Bob, Sam	NS	in-service	NS	NS
Pinter, East, and Thrush (2015)	Linda, Ava, Leeza, Mick	NS	in-service	Master's in Teaching and Leadership, Special Education, and Educational Leadership	2 – 13 years
Reamer (1996)	Jackie, Maria, Carol	20-42	para	High school diploma, Some college, Bachelor of Arts/Sociology	first year – 1 year
Robinson (2011)	Anna, Deborah, Sandra, Mary	18-60	para	High school diploma, Bachelor's in Psychology & Biology	first year – 17 years
Saudargas (1973)	Teachers A, B, C, D, E	NS	pre-service	NS	second year
Snyder (2013)	Amanda, Leah, Kristin, Tricia	24-49	para	High school diploma, Some college, Associate of Arts, Bachelor of Arts	1 – 13 years
Stephenson, Carter, and Arthur-Kelly (2010)	School A, Class 2; School B	NS	in-service, para	NS	2 – 11 years
Westover (2011)	Dyads A, B, C	42-53	para	High school diploma, Bachelor's degree	first year – 17 years

Note. NS = not specified; para = paraprofessional

Table G3
Student and Setting Characteristics

Study	Disability Type	Grade Level	Group Size	Instruction	Setting	Student Outcomes
Ahuja (2000)	NS	preschool, elementary	small group	reading, NS	SPED-NS, general education	correct responses
Alexander, Williams, and Nelson (2012)	NS	elementary, middle	NS	language arts	resource	N/A
Bingham, Spooner, and Browder (2007)	multiple disabilities	elementary, middle school, high school	individual	communication (AAC)	self-contained	use of AAC device, challenging behavior
Bishop, Snyder, and Crow (2015)	delays in fine motor, literacy, & receptive language skills	preschool	individual	literacy, functional living skills	inclusion	N/A
Bose-Deakins (2006)	NS	preschool	large group	circle time	inclusion	N/A
Capizzi, Wehby, and Sandmel (2010)	LD, BD, ID	elementary, middle school	NS	reading, math	resource	N/A
Carnine and Fink (1978)	DD, Down syndrome, language delay	preschool	small group	reading	inclusion	N/A

Table G3
Continued

Study	Disability Type	Grade Level	Group Size	Instruction	Setting	Student Outcomes
Digennaro-Reed, Codding, Cantania, and Maguire (2010)	autism, brain injury, developmental disorders	NS	individual	independent work tasks	self-contained	N/A
Englund (2011)	physical disabilities, DD	preschool	large group	language modeling	inclusion	N/A
Erbas, Tekin-Iftar, and Yucesoy (2006)	NS	NS	individual	NS	self-contained	N/A
Fedders (2012)	autism	elementary	individual	reading	resource, self-contained	correct responses
Fullerton, Conroy, and Correa (2009)	challenging behavior	preschool	individual	transitions	NS	engagement, compliance
Hager (2012)	NS	elementary	small group	reading	resource	N/A
Hawkins and Heflin (2011)	EBD, ADHD, anxiety disorder, CD, bipolar, depression, schizophrenia	high school	large group	NS	self-contained	N/A

Table G3
Continued

Study	Disability Type	Grade Level	Group Size	Instruction	Setting	Student Outcomes
Kaiser, Ostrosky, and Alpert (1993)	significant cognitive and language delays, severe physical disabilities	preschool	small group	communication	self-contained	prompted & unprompted use of communication targets, total communicative responses
Lambour (1976)	LD, BD	elementary, middle school	NS	NS	self-contained	N/A
Lindsey (2014)	DD, ID, autism	preschool, elementary, middle school	individual	reading	self-contained	correct responses
Lynes (2013)	NS	preschool	small group	circle time	inclusion	one word utterances, number of words spoken
Morgan, Menlove, Salzberg, and Hudson (1994)	mild ID, LD, BD	elementary	small group	spelling, reading	resource	rate of response

Table G3
Continued

Study	Disability Type	Grade Level	Group Size	Instruction	Setting	Student Outcomes
Peck, Killen, and Baumgart (1989)	DD	preschool	large group, small group, individual	circle time, lunch	inclusion	following directions, labelling actions, answering yes/no questions
Pelletier, McNamara, Braga-Kenyon, and Ahearn (2010)	autism	high school	individual	communication & other skill domains (NS)	self-contained	N/A
Pinter, East, and Thrush (2015)	cognitive impairment, multiple disabilities, OHI, autism, LD, EBD	middle school, high school	whole group	reading, language arts, science/social studies, math, vocational skills	self-contained, SPED-NS	N/A
Reamer (1996)	DD	preschool, elementary	individual	functional living skills	resource, inclusion	N/A
Robinson (2011)	autism	preschool, elementary	individual	social communication	inclusion	social communication goals, student affect
Saudargas (1973)	NS	elementary	small group	NS	NS	N/A

Table G3
Continued

Study	Disability Type	Grade Level	Group Size	Instruction	Setting	Student Outcomes
Snyder (2013)	DD	preschool	small group	reading	inclusion	engagement
Stephenson, Carter, and Arthur-Kelly (2010)	cerebral palsy, physical disability, ID	elementary, middle school, high school	large group	morning arrival routine, morning tea routine	self-contained	N/A
Westover (2011)	DD, ID, multiple disabilities, physical disabilities, autism, cerebral palsy, Down syndrome	elementary	individual	reading	self-contained	literacy skills, correct responses, no-responses

Note. ADHD = attention deficit hyperactivity disorder, BD = behavior disorder, DD = developmental delay, EBD = emotional/behavioral disorder, ID = intellectual disability, LD = learning disability, OHI = other health impairment, NS = not specified, SPED = special education; N/A = not applicable (i.e., student outcomes not reported)

APPENDIX H

META-ANALYSIS STUDY FOREST PLOTS

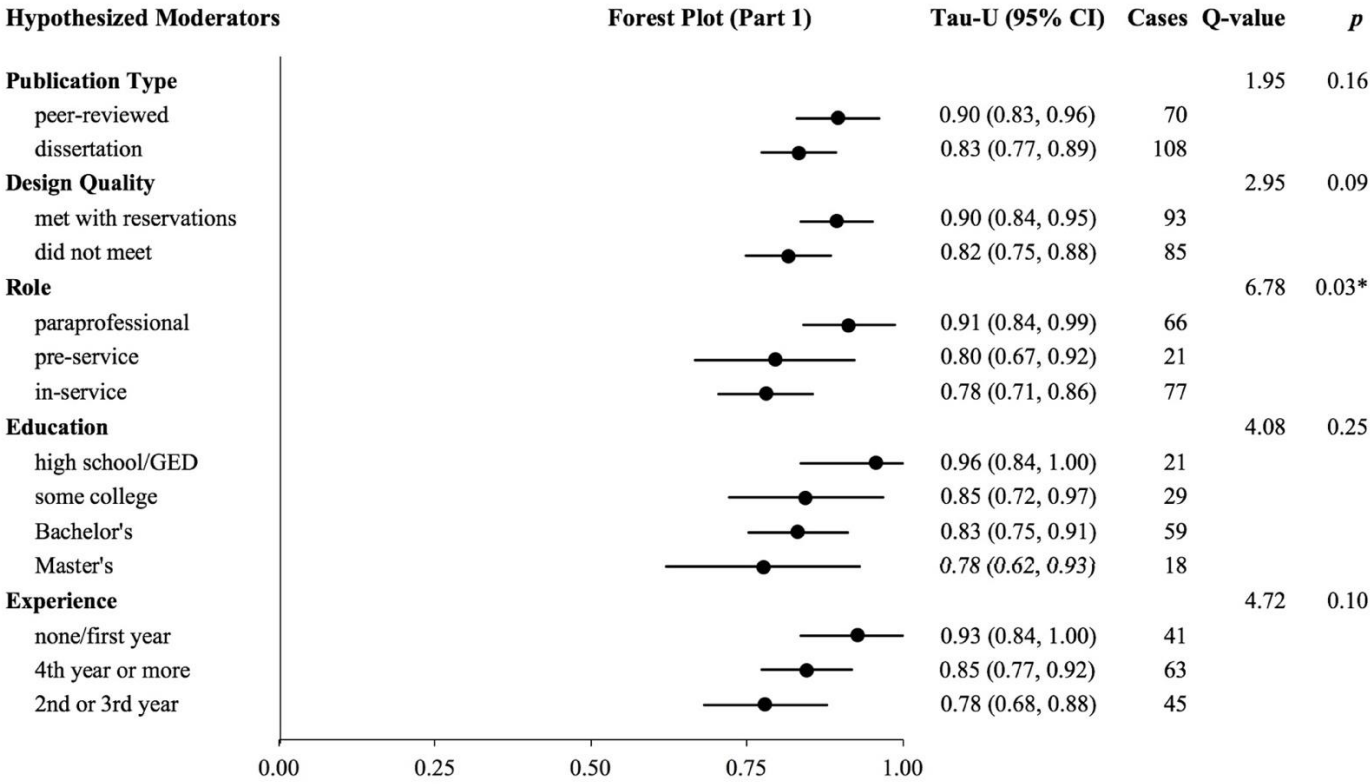

Figure H1. Forest plot (part 1) of Tau-U effect sizes, confidence intervals, the between *Q* value, and *p* values for the following potential moderators and their related subgroups: publication type, design quality, role, education, and experience level

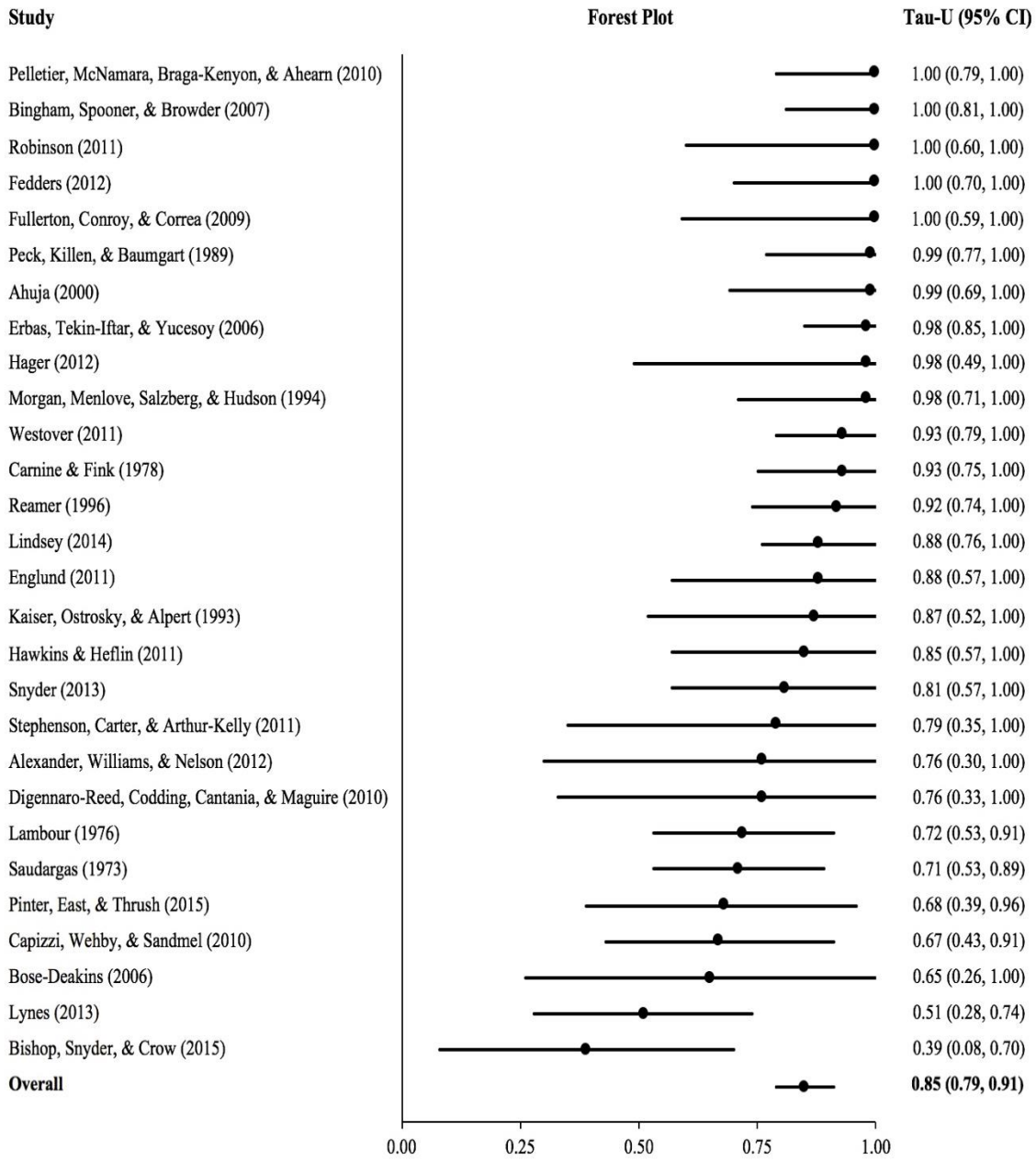

Figure H2. Forest plot of Tau-U effect sizes and confidence intervals by study

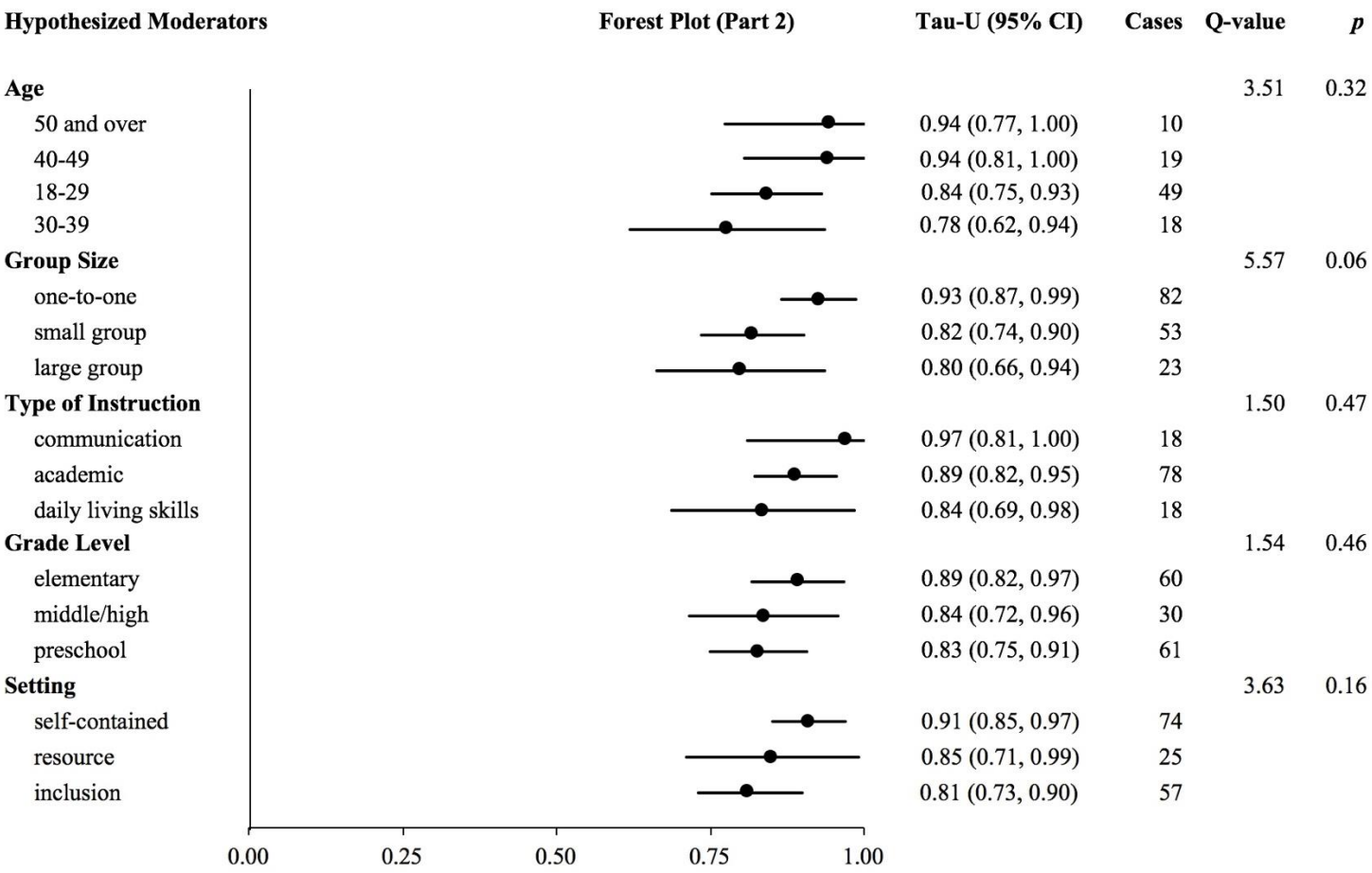

Figure H3. Forest plot (part 2) of Tau-U effect sizes, confidence intervals, the between *Q* value, and *p* values for the following potential moderators and their related subgroups: age, group size, type of instruction, grade level, and setting