Article

Hammill Institute on Disabilities

The Journal of Special Education
2019, Vol. 53(1) 3–14

DOI: 10.1177/0022466918798361
journalofspecialeducation.sagepub.com
SAGE

A Systematic Review of Single-Case Research on Video Analysis as Professional Development for Special Educators

Kristi L. Morin, PhD[1], Jennifer B. Ganz, PhD[2], Kimberly J. Vannest, PhD[2], April N. Haas, MEd[2], Sarah A. Nagro, PhD[3], Corey J. Peltier, PhD[4], Marcus C. Fuller, MEd[2], and Sara K. Ura, MSW[2]

Abstract

Studies using video analysis are being reported more frequently in the literature. Although the body of research suggests that video analysis is effective for changing educators' instructional practices, questions regarding for whom and under what circumstances it is most effective still remain. This meta-analysis reports on the overall effectiveness of video analysis when used with special educators, as well as on moderator analyses related to participant and instructional characteristics. Tau-U, a nonparametric effect size commonly used in single-case research, was used to aggregate the results from 191 AB phase contrasts across 12 dissertations and 18 peer-reviewed publications. A total of 111 participants across 69 single-case design experiments are reflected in the results. Overall, results indicate that video analysis is effective for changing the instructional practices of special educators and that it can be used across a variety of settings and with a diverse group of educators.

Keywords

video analysis, meta-analysis, single-case research, teachers, paraprofessionals

Effective teaching is strongly correlated with positive student outcomes (Wei, Darling-Hammond, & Adamson, 2010). Unfortunately, many educators feel unprepared to provide students, particularly those with disabilities, with high-quality instruction (Dicke, Elling, Schmeck, & Leutner, 2015). Many educators who work with students with disabilities seek out professional development because it can be overwhelming to demonstrate competence in the many facets of the profession, such as delivering research-based interventions (Brownell, Sindelar, Kiely, & Danielson, 2010). Professional development is intended to help educators learn specific skills, grow a deeper understanding of how to meet student needs, and change practices to improve student outcomes (Britton & Anderson, 2010). Unfortunately, many professional development opportunities provided to educators follow the traditional model, which includes the one-stop workshop, or short-term training with no follow-up, despite overwhelming evidence to discredit its use (Garet, Porter, Desimone, Birman, & Yoon, 2001; Wei et al., 2010). Delivering ineffective professional development has negative consequences, such as educators' frustration that they have not learned the requisite skills needed to effectively teach (Wei et al., 2010).

One way to achieve a coherent link between educator needs and professional development objectives is to make educators' classrooms part of the professional development activities, leading to a more authentic learning experience. Authentic professional development, or the reform-based approach, is an effective model of professional development (Boudah, Blair, & Mitchell, 2003) and includes several characteristics. One, it focuses on educators' needs when planning and implementing professional development opportunities (Boudah et al., 2003). This is important because effective professional development involves the educator in decision-making regarding goals, feasibility, and acceptability of the intervention selected (Burns & Ysseldyke, 2009). Two, effective, authentic professional

[1]The University of North Carolina at Chapel Hill, USA
[2]Texas A&M University, College Station, USA
[3]George Mason University, Fairfax, VA, USA
[4]The University of Oklahoma, Norman, USA

Corresponding Author:
Kristi L. Morin, Frank Porter Graham Child Development Institute, The University of North Carolina at Chapel Hill, 105 Smith Level Road, Chapel Hill, NC 27516, USA.
E-mail: klmorin@unc.edu

development gives educators opportunities to practice learned skills within educators' daily lives (Corcoran, 2007; Garet et al., 2001). That is, it is critical to provide professional development that not only increases teacher knowledge but also supports the transition from knowledge to instructional practice with high fidelity of implementation. Three, effective professional development is sustained over longer time periods with progress monitoring to bolster improved practice (Harwell, 2003).

Benefits of and Questions About Video Analysis

Video analysis, a professional development approach that requires educators to evaluate video evidence of their own teaching for purposes of self-confrontation and self-reflection, is one approach to effective, authentic professional development (Nagro, deBettencourt, Rosenberg, Carran, & Weiss, 2017). Although 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). Video analysis is flexible, so educators can choose behaviors to focus on, thus assuring the professional development activity aligns with the educators' needs, transforms existing beliefs and practices, and supports the acquisition of new teaching knowledge and skills (Wang & Hartley, 2003). Reviewing video-recorded lessons offers flexibility because a single lesson can be reviewed multiple times, creating authentic learning experiences for teachers who can focus on real classroom situations without having to simultaneously teach (Tripp & Rich, 2012). In addition, advances in technology have increased the feasibility of capturing video evidence in authentic settings, resulting in greater access to ongoing and authentic learning experiences (Wang & Hartley, 2003).

Video analysis has been identified as a promising teacher education approach (Nagro & Cornelius, 2013) because educators have demonstrated improved instructional practice after participating in video analysis activities (e.g., Milburn, Girolametto, Weitzman, & Greenberg, 2014). Video analysis has been implemented with a range of populations and experience levels, including with in-service special education teachers (e.g., Capizzi, Wehby, & Sandmel, 2010), preservice special education teachers (e.g., Alexander, Williams, & Nelson, 2012), and paraprofessionals (e.g., Bingham, Spooner, & Browder, 2007). Given the poor methodological quality of several studies, it is impossible to attribute any changes or growth to video analysis or to generalize the findings to a larger population of special educators.

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); however, less is known about possible relationships between professional experience and usefulness, or effectiveness, of video analysis as a professional development approach. Some research has been conducted in this area with preservice teachers who are new to the profession and lack exposure to classroom events, finding that novice educators typically have little awareness of factors of high-quality instruction (Wiens, Hessberg, LoCasale-Crouch, & DeCoster, 2013). Add in the complexities associated with reviewing video evidence of dynamic classrooms with the necessary prerequisite skills of understanding what to look for when reviewing a video and experience in the field may play a larger role than is known regarding the success of professional development approaches such as video analysis.

Similar questions exist relating to the impact age has on educators' comfort level with technology and, therefore, ability to engage in video analysis activities with fidelity. For example, in one quasi-experimental study, where participants recorded themselves 4 times to engage in video analysis, 33 of the 36 participants reported technical difficulties and many needed ongoing technical support throughout the project (Nagro et al., 2017). There was no attempt to identify possible connections between age of participants and level of comfort with technology in this study, but younger populations may be more apt to use technology (Black, 2010) and therefore may be more willing to engage in technology-dependent professional development activities, such as video analysis. A third unanswered question is the possible impact of education. Evaluating the education level of participants is also important to determine whether 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.

Other considerations that have not been fully explored include classroom setting, grade level, and student groupings. Video analysis has been successfully implemented in general education settings (e.g., Ahuja, 2000), self-contained classrooms (e.g., Bingham et al., 2007), resource rooms or pull-out settings (e.g., Capizzi et al., 2010), and inclusive classrooms (e.g., Bose-Deakins, 2006). Video analysis has also been successfully implemented in a variety of grade levels, including preschool (e.g., Bishop, Snyder, & Crow, 2015), elementary (e.g., Westover, 2011), and middle/secondary levels (e.g., Capizzi et al., 2010), as well as with different student groupings including during whole group (e.g., Englund, 2011), small group (e.g., Carnine & Finke, 1978), and individual instruction (e.g., Lindsey, 2014). The wide variation of video analysis implementation parameters further demonstrates the flexibility of this professional development approach, but clarity regarding best practices in video analysis that will lead to positive outcomes for educators and their students is lacking. Therefore, despite the strong support for the benefits of video analysis, much of the research on the topic is descriptive in nature and many questions remain (Nagro & deBettencourt, 2017), such as whether or not participant characteristics, instructional practice, or setting variables differentially impact the effectiveness of video analysis as a professional development approach.

Purpose for the Current Study

Meta-analysis, a method of aggregating and evaluating the results of a body of research on a topic, is one way to investigate the unanswered questions regarding best practices for implementing video analysis (Lipsey & Wilson, 2001). Considerations when framing a meta-analysis of the published video analysis research include research design, publication type, and participant and setting characteristics. It is also common when conducting a meta-analysis to consider the appropriateness of applying research quality standards to evaluate the body of research on a topic (Bernard, Borokhovski, Schmid, & Tamin, 2014). To improve the rigor of single-case research and provide readers with data on the methodological quality of studies, several sets of quality indicators have been developed (e.g., Council for Exceptional Children [CEC], 2014; Horner et al., 2005; U.S. Department of Education, Institute of Education Sciences, & What Works Clearinghouse, 2016). Although quality standards can be used to measure the strength of research once applicable research studies have been identified through a systematic search, they are not intended to be used as inclusion criteria (e.g., Nagro & deBettencourt, 2017).

When implementing meta-analytic methods, concerns have been raised about including studies that do not meet minimum quality standards because results may be artificially diluted or inflated (Cook et al., 2015). However, this approach raises questions about the appropriate application of quality standards specifically relating to how many indicators must be met for research to be deemed high quality (Cooper, 2010). In addition, it is important to present a body of research holistically and accurately by identifying potential variance in the quality of research design, implementation, and reporting across studies (Lipsey & Wilson, 2001). One option is 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. Although this seems the preferable approach, there is no consensus in the field regarding which approach to use; therefore, the decision was made to use both approaches in an effort to present both comprehensive and valid findings.

The purpose of this study is to use meta-analytic methods to investigate the effectiveness of video analysis on the instructional practices of special educators. Our primary research question was the following: “What is the omnibus magnitude of effect of video analysis on the instructional practices of educators?” Secondary research questions focused on the differential effects that publication type, methodological quality, participant characteristics (e.g., role, education level, experience level, age), and instructional characteristics (i.e., group size, type of instruction, grade level, setting) have on the effectiveness of video analysis.

Method

Study Identification

Search strategy. The following databases were searched by combining the terms *teacher*, paraeducator*, teach* assistant*, paraprofessional**, or *instructional assistant** with *video*, analy*, evaluat*, reflect**, and *feedback*: Educational Resources Information Center (ERIC), PsycARTICLES, PsycINFO, Education Source, Teacher Reference Center, Academic Search Complete*, and *Education Full Text*. The search was limited to peer-reviewed articles and dissertations and covers the years 1973–2018.

Ancestral/forward/first author search strategy. An ancestral, forward, and first author search was conducted on all documents (i.e., peer-reviewed articles and dissertations) identified through the primary search by entering the title or first author name in Scopus, an abstract and citation database. For documents not included in Scopus, a hand search of the reference lists was conducted.

Inclusion/exclusion criteria. The following inclusion criteria were applied: (a) The study included single-case experimental design data; (b) at least one participant was a teacher or paraprofessional, or was enrolled in a teacher preparation program, in an early childhood through 12th grade classroom; (c) recorded videos of the in-service teacher, preservice teacher, or paraprofessional were analyzed by the participant; (d) the intervention included an evaluation or feedback component; (e) the participant was the focus of the video; (f) at least one dependent variable was related to improving observed teacher behavior; and (g) the study was published in English. 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. A lead

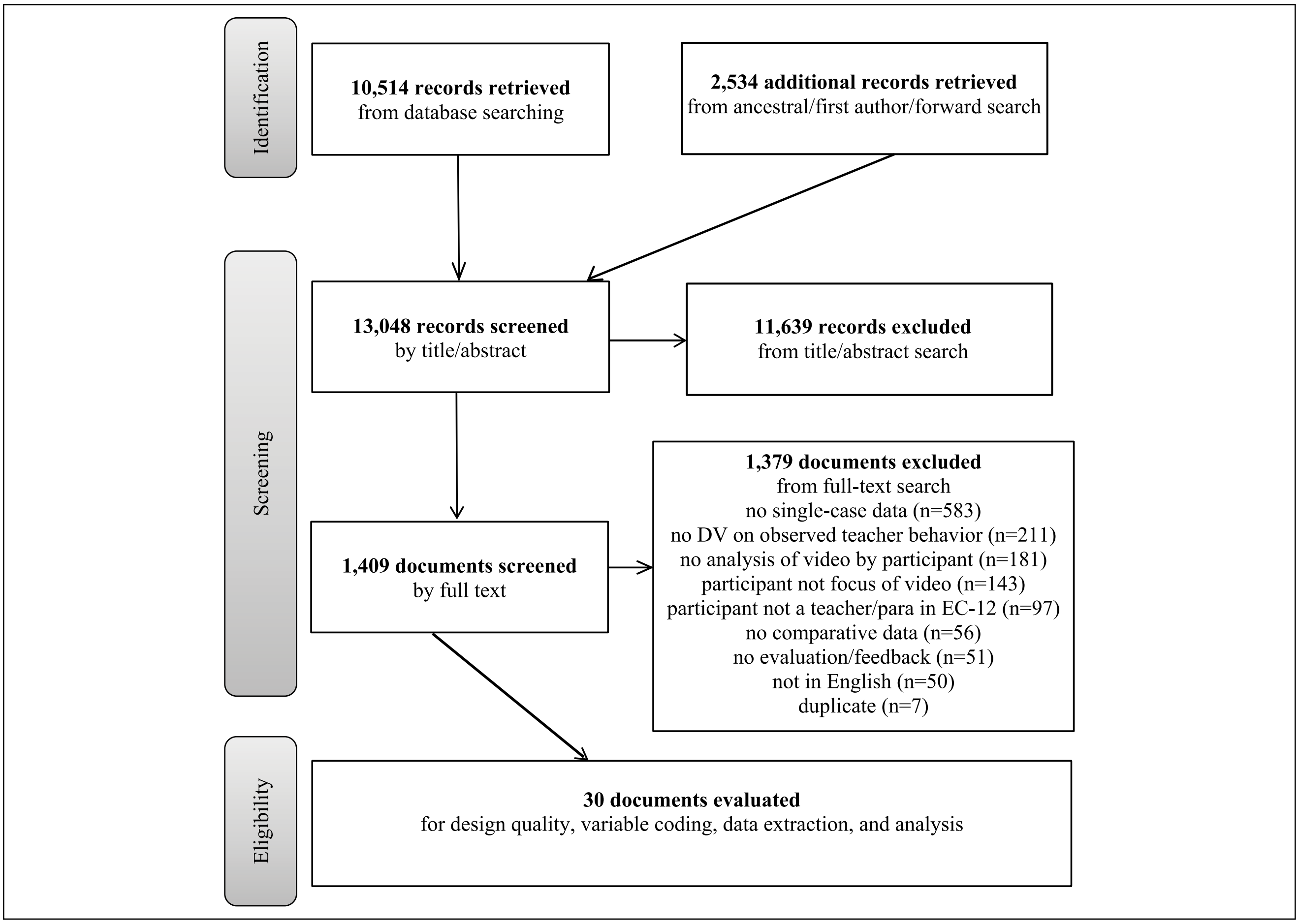

Figure 1. Flowchart depicting the records that were included/excluded during each phase.

teacher was defined as the person in the classroom who was primarily responsible for the instruction of students. In early childhood settings, the lead teacher was required to have some experience and/or credentialing in education to be coded as an in-service teacher. Early childhood was defined as students being at least 36 months of age.

Exclusion criteria included the following: (a) qualitative studies, reviews, group studies, 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 nonbehavioral 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.

Title/abstract and full-text review. After documents were identified and duplicates removed, the search resulted in 13,048 documents (primary search = 10,514; ancestral/first author/forward search = 2,534). Titles and abstracts of all identified documents were evaluated to determine whether they met inclusion criteria. If we could not determine 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, 11,639 documents were excluded, resulting in a total of 1,409 documents that were screened by full-text. Application of the inclusion/exclusion criteria to the full-text resulted in the exclusion of an additional 1,379 documents, leaving 30 documents to be evaluated for design quality, variable coding, data extraction, and analysis. See Supplemental Appendix A for a list of included documents and Figure 1 for reasons why documents were excluded.

Application of the What Works Clearinghouse (WWC) Pilot Single-Case Design Standards

Once all studies were identified, they were evaluated for design quality using the WWC Pilot Single-Case Design Standards (U.S. Department of Education, Institute of Education Sciences, & What Works Clearinghouse, 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. Each study was evaluated at the experiment level, defined as one single-case experimental design. 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, Institute of Education Sciences, & What Works Clearinghouse, 2016): (a) manipulation of the independent variable, (b) whether or not interobserver agreement was reported, (c) the percentage of data for which interobserver agreement was collected, (d) whether or not interobserver agreement scores met minimum quality thresholds, (e) whether or not the experiment included a minimum of three attempts to demonstrate treatment effects at three different points in time, and (f) the number of data points per phase. Multiple-probe designs were also evaluated on the presence of the following additional Basic Design Quality Standards (U.S. Department of Education, Institute of Education Sciences, & What Works Clearinghouse, 2016): (a) the number of data points within the initial baseline sessions, (b) the number of consecutive probe points prior to intervention, and (c) the collection of data points in subsequent levels when the previous level first received intervention.

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, Institute of Education Sciences, & What Works Clearinghouse, 2016). Each standard and the criteria used to evaluate each experiment are described in more detail in Supplemental Appendix B.

Variable Coding

A portion of the included articles were coded using descriptive data 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 contextual and participant 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).

Supplemental Appendix C provides operational definitions and subgroup categories for all variables. Some subgroups were initially coded separately—that is, experience: none, first year; age: 50 to 59, 60 and above; grade level: middle school, high school; type of instruction: reading/English language arts (ELA), math—but were later combined prior to moderator analyses due to a low number of contrasts in each category. In addition, although 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, if they fit into multiple categories, or if the study did not include enough information on the variable being coded, the study was coded as “0” for that particular variable and was excluded from further analysis.

Data Extraction and Effect Size Calculation

Data extraction. Data were extracted from each graph using the free, online software, GetData Graph Digitizer (2013). 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, which can be interpreted as the “percent of nonoverlapping data minus the percent of overlapping data” (Parker, Vannest, Davis, & Sauber, 2011, p. 285). Tau-U was selected as the effect size measure for this meta-analysis because it has several advantages over other nonparametric effect sizes (Parker, Vannest, & Davis, 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; (c) the ability to control for undesirable baseline trend; (d) the ability to calculate confidence intervals (CIs); (e) high sensitivity; and (f) simple calculation. In addition, Tau-U has been found to be consistent with visual analysis of data (Parker, Vannest, & Davis, 2011). Tau-U can be tentatively interpreted as follows: small effect = 0 to 0.62; medium effect = 0.63 to 0.92; large effect = 0.93 to 1.00 (Parker, Vannest, Davis, & Sauber, 2011).

Effect sizes were calculated by entering baseline and intervention data into the free, online Tau-U calculator (Vannest, Parker, Gonen, & Adiguzel, 2016) to obtain a Tau-U value for each AB contrast (i.e., baseline vs. intervention), and then combining these experiment-level effect sizes into one omnibus effect size per study. When combining Tau-U values, an inverse weighting scheme was used that gives more credit to studies with more data points and stability. Because trend was present in 81% of the baseline data and, of that, 39% was undesired trend, all results reflect corrected baseline data. Next, Tau-U effect sizes and their standard error (SD_{Tau}) were entered into the Comprehensive Meta-Analysis software program (Version 3; Borenstein, Hedges, Higgins, & Rothstein, 2005) to generate an omnibus effect size that represents the entire body of literature on video analysis. The omnibus effect size reflects a weighting scheme that assigns credit based on within-study variance and between-study variance. 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.

Although neither a fixed effects nor random effects model is an "exact fit" for single-case data, 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 (Borenstein, Hedges, Higgins, & Rothstein, 2009; Lipsey & Wilson, 2001). Statistical significance was determined using both the Q-statistic (Cochran, 1954) and the I^2 index (Higgins & Thompson, 2002). 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 et al., 2009). When the associated p value is less than the critical value, which we set to .05, there is evidence that the differences in the dispersion (i.e., range) of effect sizes between the subgroups are due to real differences and not random error (Borenstein et al., 2009). In addition, we calculated I^2 index (Higgins & Thompson, 2002) because the Q-statistic may have poor or excessive power, depending on the number of included studies, when determining true heterogeneity among studies. The I^2 index is interpreted as the percentage of variance between studies; thus, an $I^2 = 60$ means 60% of the variance is the "true variance" disregarding variance attributed to sampling error alone (Higgins & Thompson, 2002)

When conducting moderator analyses, the variance for the Q-statistic was partitioned into the variance within studies (i.e., Q_w) and the variance between studies (i.e., Q_b). The Q_w represents the amount of variance that is unaccounted for and remains within subgroups, and the Q_b represents the amount of variance the moderating variable is able to quantify (Lipsey & Wilson, 2001). Interpretation is similar to ANOVAS; thus, for each moderator analysis, the associated p value for the Q_w was used to determine whether the homogeneity of variance assumption within the groups was violated (Lipsey & Wilson, 2001). None of the p values for the associated Q_w values were statistically significant at the .05 level; therefore, we tested the statistical significance of the associated Q_b for each variable to determine whether they functioned as moderators.

Publication bias. To increase the validity in the interpretation of our results and to identify potential publication bias in this study, we computed *Rosenthal's Fail-safe N* (Rosenthal, 1979). The *Rosenthal's Fail-safe N* test identifies how many additional studies, with no effect, would need to be added to "nullify" the effects of a study. *Rosenthal's Fail-safe N* revealed that an additional 103 nonsignificant studies would need to be added to "nullify" the omnibus effect that included all designs and 3,736 nonsignificant studies would need to be added to "nullify" the omnibus effect that only included designs that met WWC Design Standards. This implies that publication bias should not be a concern.

Interrater reliability (IRR). A minimum of 20% of data for all phases of the study were independently coded for IRR by one of four doctoral students. All raters were trained to criterion for each stage of the process using a subset of data. If there was a disagreement, a third evaluator independently rated the studies or the first two evaluators discussed the disagreement until they came to a consensus. In cases where three evaluators were needed, the final decision was based on the agreement of two evaluators.

For data extraction 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. There was one disagreement, which was resolved by re-calculating the data to determine where the disagreement lay. For all stages of the study, IRR 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. IRR data were taken on an average of 45% of data (range = 20%–78%) and resulted in a mean IRR score of 95% (range = 86%–100%). Detailed IRR results for each stage of the study are presented in Supplemental Appendix D.

Results

The literature base on video analysis in educational settings includes diverse study, participant, setting, and student characteristics (see Tables 1 and 2 and Supplemental Appendices E-G). There were a total of 69 separate single-case designs, 191 AB phase contrasts, and 111 participants reflected across 30 documents included in this meta-analysis. Although there were a large number of dissertations in this data set ($n = 12$), most documents were peer-reviewed articles ($n = 18$). The studies fell short of WWC Pilot Single-Case Design Standards, with only one single-case design meeting the standards without reservations (i.e., Knowles, Massar, Raulston, & Machalicek, 2017) and less than half meeting them with reservations ($n = 30$). The remaining single-case designs did not meet the standards (see Table 1).

Table 1. Study Characteristics and Effect Sizes.

Study	Publication type	Design[a]	Design quality	Participants	Contrasts	Pairs	Tau-U (95% CI)	Tau-U effect
Ahuja (2000)	DIS	MBD (3)	MWR, DNM	7	7	165	0.99 [0.69, 1.00]	Large
Alexander, Williams, and Nelson (2012)	PR	AB (2)	DNM	2	2	106	0.76 [0.30, 1.00]	Medium
Bingham, Spooner, and Browder (2007)	PR	MPD (2)	DNM	3	6	318	1.00 [0.81, 1.00]	Large
Bishop, Snyder, and Crow (2015)	PR	MPD	MWR	2	2	432	0.39 [0.08, 0.70]	Small
Bose-Deakins (2006)	DIS	MPD	MWR	3	3	131	0.65 [0.26, 1.00]	Medium
Capizzi, Wehby, and Sandmel (2010)	PR	MBD (3)	MWR	3	9	312	0.67 [0.43, 0.91]	Medium
Carnine and Fink (1978)	PR	MBD (2)	MWR	3	6	1274	0.93 [0.75, 1.00]	Large
Digennaro-Reed, Codding, Cantania, and Maguire (2010)	PR	MBD	MWR	2	2	113	0.76 [0.33, 1.00]	Medium
Englund (2011)	DIS	MBD (2)	MWR	6	6	180	0.88 [0.57, 1.00]	Medium
Erbas, Tekin-Iftar, and Yucesoy (2006)	PR	MBD	MWR	6	6	4,144	0.98 [0.85, 1.00]	Large
Fedders (2012)	DIS	MBD (2)	MWR	3	6	120	1.00 [0.70, 1.00]	Large
Fullerton, Conroy, and Correa (2009)	PR	MBD	DNM	3	3	110	1.00 [0.59, 1.00]	Large
Hager (2012)	PR	MBD	DNM	1	2	79	0.98 [0.49, 1.00]	Large
Hawkins and Heflin (2011)	PR	Reversal (3)	MWR	3	6	235	0.85 [0.57, 1.00]	Medium
Kaiser, Ostrosky, and Alpert (1993)	PR	MBD (2)	MWR	1	2	264	0.87 [0.52, 1.00]	Medium
Knowles, Massar, Raulston, and Machalicek (2017)	PR	MBD	MET	1	3	333	0.95 [0.63, 1.00]	Large
Lambour (1976)	DIS	Reversal (2), ABA, ABAC, MBD	MWR, DNM	7	9	665	0.72 [0.53, 0.91]	Medium
Lindsey (2014)	DIS	MBD (3)	DNM	8	24	2123	0.88 [0.76, 1.00]	Medium
Lynes (2013)	DIS	MBD (2)	DNM	6	12	564	0.51 [0.28, 0.74]	Small
Morgan, Menlove, Salzberg, and Hudson (1994)	PR	MBD	MWR	5	5	420	0.98 (0.71, 1.00]	Large
Peck, Killen, and Baumgart (1989)	PR	MBD	MWR	3	6	614	0.99 [0.77, 1.00]	Large
Pelletier, McNamara, Braga-Kenyon, and Ahearn (2010)	PR	MBD	MWR	3	3	60	1.00 [0.79, 1.00]	Large
Pinter, East, and Thrush (2015)	PR	MBD	DNM	4	4	465	0.68 [0.39, 0.96]	Medium
Reamer (1996)	DIS	MBD (3), AB	MWR, DNM	3	15	565	0.92 [0.74, 1.00]	Medium
Robinson (2011)	PR	MPD	DNM	4	4	68	1.00 [0.60, 1.00]	Large
Saudargas (1973)	DIS	ABACBC (2), ABCACD, ABCDADED, ABABACD	DNM	5	9	1016	0.71 [0.53, 0.89]	Medium
Snyder (2013)	DIS	MBD (2)	DNM	4	8	290	0.81 [0.57, 1.00]	Medium
Stephenson, Carter, and Arthur-Kelly (2011)	PR	AB (2)	DNM	2	2	108	0.79 [0.35, 1.00]	Medium
Westover (2011)	DIS	MBD (3)	MWR	3	9	2259	0.93 [0.79, 1.00]	Large
Zheng (2017)	DIS	AB (10)	DNM	5	10	250	0.69 [0.45, 0.93]	Medium

Note. CI = confidence interval; DIS = dissertation; MBD = multiple-baseline design; MWR = met with reservations; DNM = did not meet; PR = peer-reviewed; MPD = multiple probe design; MET = met without reservations; 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. The references in this table can be found in Appendix A.
[a]Indicates the number of experiments, when more than one, included in the study

Effects of Video Analysis

Overall, Tau-U effect sizes for the use of video analysis to change special educators' instructional behavior ranged from 0.39 to 1.00, with a mean of 0.85 ($CI_{95\%}$ = [0.79, 0.91]) when all single-case design experiments were included, and 0.88 ($CI_{95\%}$ = [0.81, 0.96]) when only those meeting WWC Design Quality Standards were included (see Supplemental Appendix G). Pairwise comparisons for each study ranged from 60 to 4,144 (see Table 1). The overall Q-value was 46.63 (p = .02) when all single-case designs were included and 23.98 (p = .12) when only designs meeting WWC Design Standards were included. The I^2 value

Table 2. Effect Sizes by Potential Moderator.

Variable	Designs meeting WWC design standards only				All designs			
	Tau-U (95% CI)	Cases	Q_b	*p*	Tau-U (95% CI)	Cases	Q_b	*p*
Role			1.24	.54			4.60	.10
Paraprofessional	0.92 [0.80, 1.00]	21			0.90 [0.83, 0.97]	76		
Preservice	0.90 [0.68, 1.00]	8			0.80 [0.67, 0.92]	21		
In-service	0.84 [0.75, 0.93]	55			0.79 [0.72, 0.86]	80		
Education			2.51	.47			4.70	.20
High school/GED	0.93 [0.78, 1.00]	10			0.96 [0.84, 1.00]	21		
Some college	0.93 [0.74, 1.00]	11			0.80 [0.68, 0.91]	35		
Bachelor's	0.80 [0.68, 0.91]	27			0.84 [0.76, 0.91]	63		
Master's	0.86 [0.65, 1.00]	11			0.78 [0.62, 0.93]	18		
Experience			2.40	.30			6.52	.04*
None/first year	0.92 [0.81, 1.00]	27			0.93 [0.84, 1.00]	44		
Second or third year	0.77 [0.60, 0.93]	20			0.76 [0.67, 0.86]	53		
Fourth year or more	0.90 [0.81, 0.99]	27			0.85 [0.78, 0.92]	65		
Age			5.77	.12			4.86	.18
18–29	0.79 [0.67, 0.92]	26			0.84 [0.76, 0.93]	56		
30–39	0.64 [0.42, 0.87]	9			0.74 [0.60, 0.89]	22		
40–49	0.95 [0.79, 1.00]	11			0.94 [0.81, 1.00]	21		
50 and above	0.90 [0.71, 1.00]	6			0.94 [0.77, 1.00]	10		
Group size			1.03	.60			3.96	.14
One-to-one	0.92 [0.84, 1.00]	42			0.91 [0.85, 0.97]	92		
Small group	0.95 [0.83, 1.00]	24			0.83 [0.75, 0.91]	56		
Large group	0.84 [0.68, 1.00]	17			0.80 [0.66, 0.94]	23		
Type of instruction			1.08	.58			1.59	.45
Communication	0.88 [0.64, 1.00]	8			0.97 [0.81, 1.00]	18		
Academic	0.91 [0.82, 1.00]	39			0.88 [0.81, 0.94]	88		
Daily living skills	0.80 [0.64, 0.96]	15			0.84 [0.69, 0.98]	18		
Grade level			1.92	.38			0.74	.69
Preschool	0.86 [0.76, 0.96]	29			0.83 [0.75, 0.91]	63		
Elementary	0.94 [0.84, 1.00]	31			0.88 [0.81, 0.95]	71		
Middle/high	0.82 [0.64, 1.00]	15			0.84 [0.72, 0.96]	30		
Setting			2.32	.31			2.86	.24
Self-contained	0.94 [0.86, 1.00]	39			0.90 [0.84, 0.96]	87		
Resource	0.84 [0.68, 0.99]	20			0.85 [0.71, 0.99]	25		
Inclusion	0.85 [0.75, 0.95]	31			0.81 [0.73, 0.90]	57		

Note. WWC = What Works Clearinghouse; CI = confidence interval; GED = general educational development.
*Statistically significant at the .05 probability level.

was 37.81 when all designs were included and 29.12 when only those meeting WWC Design Standards were included.

Methodological Quality and Publication Type

One question of interest was the effect of methodological quality, as measured by WWC Pilot Single-Case Design Standards, on effects found for video analysis. One design across one document met the standards, 30 designs across 18 documents met the standards with reservations, and 38 designs across 15 documents did not meet standards (see Table 1). Some documents included multiple single-case design experiments, with some designs meeting standards with reservations and some not meeting standards (see Table 1). In these instances, two effect sizes were calculated for the document—one for the designs meeting standards and one for the designs not meeting standards. Documents with only designs meeting WWC Standards, or meeting them with reservations, produced a Tau-U effect size (ES) of 0.88 ($CI_{95\%}$ = [0.81, 0.96]), and documents with only designs not meeting WWC Standards produced a Tau-U ES of 0.81 ($CI_{95\%}$ = [0.72, 0.90]). These differences were not statistically significant (Q_b = 1.85; p = .174).

Also of interest was the potential difference between results produced in dissertations versus peer-reviewed journal articles. Analyses of all data for video analysis resulted in 73 AB phase contrasts from peer-reviewed literature (ES = 0.90, $CI_{95\%}$ = [0.83, 0.96]) and 118 AB phase contrasts from dissertations (ES = 0.83, $CI_{95\%}$ = [0.77, 0.88]); results from this analysis were not statistically significant ($Q_b = 2.80$; $p = .09$). Only including data meeting WWC Design Standards in the analyses resulted in 50 contrasts from peer-reviewed literature (ES = 0.90, $CI_{95\%}$ = [0.83, 0.97]) and 46 AB phase contrasts from dissertations (ES = 0.90, $CI_{95\%}$ = [0.80, 0.99]); results from this analysis were not statistically significant ($Q_b = 0.002$; $p = .962$).

Potential Moderators of Effect

Variables related to participant and instructional characteristics were examined for their potential influence on the size of an effect for video analysis. Table 2 provides detailed results of these analyses, both when all designs were included (i.e., all) and when only designs meeting WWC Design Standards were included (i.e., WWC only). Regarding participant characteristics, analyses were run for role, education level, experience level, and age of the participants. Tau-U results from the analyses for role were largest for paraprofessionals (ES = 0.92, WWC only; ES = 0.90, all); however, all variables produced moderate effects. When considering the education level of participants, those with a high school or general educational development (GED) level of education produced the largest effects (ES = 0.93, WWC only; ES = 0.96, all), although participants with some college also produced large effects when only designs meeting WWC Standards were included (ES = 0.93). Effect sizes for the experience level of educators were large for participants with no experience or those in their first year of teaching when all studies were included (ES = 0.93), with other experience levels producing moderate effect sizes. Age produced large effects for participants aged 40 to 49 years (ES = 0.95, WWC only; ES = 0.94, all) and 50 and above when all studies were included (ES = 94) and moderate effects for participants from other age groups. For all participant characteristics, the only variable that was statistically significant was experience ($Q_b = 6.52$, $p = .04$), but only when all studies were included in the analysis.

Variables related to instructional characteristics were also investigated as potential moderators, with analyses run on group size, type of instruction, grade level, and setting. Small groups produced large effects (ES = 0.95), but only when designs not meeting WWC Standards were excluded; all other group sizes produced moderate effects. When type of instruction was analyzed, communication produced large effects (ES = 0.97) when all studies were included; however, the effect dropped to a moderate effect size when designs not meeting WWC Standards were excluded (ES = 0.88); all other types of instruction produced moderate effects. For grade level, elementary produced a large effect size, but only when designs not meeting WWC Standards were excluded (ES = 0.94); all other grade levels produced moderate effects. Regarding setting, self-contained produced a large effect (ES = 0.94), but, as is the case with group size and grade level, these effects are only seen when designs not meeting WWC Standards were excluded; all other settings produced moderate effects.

Discussion

This meta-analytic review both supports and extends prior literature on the use of video analysis with educators. As a result of this work, video analysis can be deemed an effective practice when used with special education teachers. The current meta-analysis extends prior reviews by including more recent research and by using effect sizes to aggregate results across studies. The current work also supports policy and former findings that effective professional development includes active learning, coherence between the professional development and the educators' contexts, engagement of educators in the professional development, and longer term efforts (Corcoran, 2007; Garet et al., 2001).

When interpreting the results, characteristics of the educators included in the analyses are critical factors to consider. The educators in the included studies were mostly in-service teachers and paraprofessionals between the ages of 18 and 29 years. More research is needed with populations beyond these, such as with preservice teachers and older educators. Regarding reporting practices, a large number of studies either failed to report information on the characteristics of educators in the study (i.e., education, age, and experience level), or only provided general information. This is concerning considering participant demographic information is needed for generalizability and replication. If practitioners are not able to determine whether participants in the study have the same backgrounds as the educators with whom they plan to use the intervention, then the relevance of the study is reduced.

Of the four potential participant characteristic moderators analyzed, only experience level was found to have statistically significant differences among the subgroups, with educators in their first year of teaching or those with no experience (i.e., preservice teachers) showing the largest effects. One possible reason for this finding is that teachers with little to no experience may be more willing to engage in professional development efforts to improve their skills. Prior research indicates that novice teachers are typically unaware of effective teaching practices (Wiens et al., 2013); thus, video analysis, which requires teachers to develop an awareness of their instruction, appears to be a particularly effective intervention for this subgroup of educators. This conclusion, however, must be viewed with caution as statistically significant differences were only found when all studies were included. One possibility for this finding is that the variance attributed to design quality was being captured in this moderator analysis. To account for this, future researchers should consider running a meta-regression to simultaneously control for potential moderator variables to determine whether experience level truly does function as a moderator.

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). When evaluating setting characteristics, most participants taught in a preschool or elementary setting and delivered academic instruction, particularly in reading or language arts. The number of studies conducted in secondary settings was minimal, which is reflective of research in the field of special education (Wong et al., 2013). 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.

Implications for the Improvement of Research and Practice

Several implications result from this work. One, more rigorous single-case experiments are needed in this area of scholarship, to increase confidence in aggregated results. This will benefit practitioners by allowing for more confidence with regard to the legal mandate to implement evidence-based practices with students with disabilities. Two, more experiments are needed within each variable category, to determine whether or not there are truly no significant differences within variable categories. More fine-grained analyses are needed to enable practitioners to individualize video analysis based on educator characteristics, instructional characteristics, and settings. For example, the frequency with which participants viewed videos varied widely, as did the number of components the interventions included. Future research should investigate the possibility of moderation due to intervention length and other implementation characteristics. Three, more research is warranted to determine whether the strong effects across educator characteristics will maintain when used with a larger number of participants in each category, especially with older educators. Four, including studies with more diverse settings, including secondary education settings, in future single-case experiments and 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. That said, this work provides ample and promising evidence that video analysis is moderately effective for most educators, regardless of the setting in which they work.

Limitations

There are a few limitations which should be noted. First, only single-case research was included in this meta-analysis due to the lack of a commonly accepted effect size measure that can confidently be used to aggregate single-case and group research. Second, there is still debate in the field over the use of meta-analytic techniques with single-case experimental designs because they neither fit a random effects nor fixed effects model perfectly. Third, some moderator analyses included a small number of cases, particularly among those analyses that included only designs that met WWC Design Standards; thus, the results of these analyses should be viewed with caution. Finally, the use of Tau-U, or other nonparametric effect sizes, to quantify study effects has their limitations; idiosyncrasies in time series data are lost when converted to a metric to report results, which may lead to inaccurate conclusions (Burns, Zaslofsky, Kanive, & Parker, 2012).

Conclusion

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.

Authors' Note

Portions of this article were previously published as part of the first author's doctoral dissertation and were presented at the

American Educational Research Association's 2018 Annual Meeting.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education through Grant R324B160038 awarded to University of North Carolina–Chapel Hill; by the Office of Special Education Programs, U.S. Department of Education through Grant H325D110046 awarded to Texas A&M University; and by a Research Scholar Award awarded by the College of Education and Human Development at Texas A&M University. The opinions expressed within this article represent those of the authors and do not represent the U.S. Department of Education or Texas A&M University.

ORCID iD

Jennifer B. Ganz https://orcid.org/0000-0002-2080-102X

References

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*(4), 18–24.

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.

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]. Available 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.

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

Britton, L. R., & Anderson, K. A. (2010). Peer coaching and pre-service teachers: Examining an underutilized concept. *Teaching and Teacher Education*, *26*, 306–314.

Brownell, M. T., Sindelar, P. T., Kiely, M., & Danielson, L. C. (2010). Special education teacher quality and preparation: Exposing foundations, constructing a new model. *Exceptional Children*, *76*, 357–377. doi:10.1177/001440291007600307

Burns, M. K., & Ysseldyke, J. E. (2009). Reported prevalence of evidence-based instructional practices in special education. *The Journal of Special Education*, *43*(1), 3–11.

Burns, M. K., Zaslofsky, A. F., Kanive, R., & Parker, D. C. (2012). Meta-analysis of incremental rehearsal using phi coefficients to compare single-case and group designs. *Journal of Behavioral Education*, *21*, 185–202. doi:10.1007/s10864-012-9160-2

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

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.

Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, *10*, 101–129.

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.

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

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

Dicke, T., Elling, J., Schmeck, A., & Leutner, D. (2015). Reducing reality shock: The effects of classroom management skills training on beginning teachers. *Teaching and Teacher Education*, *48*, 1–12. doi:10.1016/j.tate.2015.01.013

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)

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 (Version 2.26) [Computer software]. (2013). Available from http://www.getdata-graph-digitizer.com/

Harwell, S. H. (2003). *Teacher professional development: It's not an event, it's a process*. Waco, TX: Center for Occupational Research and Development.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539–1558.

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.

Knowles, C., Massar, M., Raulston, T. J., & Machalicek, W. (2017). Telehealth consultation in a self-contained classroom for behavior: A pilot study. *Preventing School Failure: Alternative Education for Children and Youth*, *61*, 28-38. doi:10.1080/1045988X.2016.1167012

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). *Applied Social Research Methods Series: Vol. 49. Practical meta-analysis* (L. Bickman & D. J. Rog, Eds.). Thousand Oaks, CA: SAGE.

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

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

Nagro, S. A., & deBettencourt, L. U. (2017). Reviewing special education teacher preparation field experience placements, activities, and research: Do we know the difference maker? *Teacher Education Quarterly*, *44*, 7–33.

Nagro, S. A., deBettencourt, L. U., Rosenberg, M. S., Carran, D. T., & Weiss, M. P. (2017). The effects of guided video analysis on teacher candidates' reflective ability and instructional skills. *Teacher Education and Special Education*, *40*, 7–25.

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

Rice, J. K. (2010). *The impact of teacher experience: Examining the evidence and policy implications* (Brief No. 11). Retrieved from https://files.eric.ed.gov/fulltext/ED511988.pdf

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin*, *86*, 638–641.

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, Institute of Education Sciences, & What Works Clearinghouse. (2016, September). *What Works Clearinghouse: Procedures and standards handbook* (Version 3.0). Available from http://whatworks.ed.gov

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: Texas A&M University. Available from www.singlecaseresearch.org

Wang, J., & Hartley, K. (2003). Video technology as a support for teacher education reform. *Journal of Technology and Teacher Education*, *11*, 105-138.

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)

Wiens, P. D., Hessberg, K., LoCasale-Crouch, J., & DeCoster, J. (2013). Using a standardized video-based assessment in a university teacher education program to examine preservice teachers knowledge related to effective teaching. *Teaching and Teacher Education*, *33*, 24–33. doi:10.1016/j.tate.2013.01.010

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: Autism Evidence-Based Practice Review Group, Frank Porter Graham Child Development Institute, The University of North Carolina.