Re-examining Evidence Based Practice in Special Education: A Discussion

Robert L. Williamson

Simon Fraser University

Andrea Jasper

Central Michigan University

Jeanne Novak

Bowling Green State University

Clinton Smith

University of Tennessee at Martin

William Hunter, Laura Casey, Kay Reeves

The University of Memphis

Abstract

The Council for Exceptional Children (CEC) recently released updated standards regarding how to determine whether any particular intervention may be deemed an evidence-based practice (EBP). As new criteria regarding the acceptance of any specific intervention as evidence-based become available, the question arises: Would the application of new standards to studies completed under the older guidelines result in changes to past conclusions? The current study examined if changes in EBP standards might change the classification of an exemplar practice that was previously designated as an EBP. In this case, we examined video modeling (VM), an accepted practice regarding skill acquisition in special education as an exemplar practice. In order to determine if the new CEC 2014 standards would impact a previously determined EBP finding, a re-examination of Bellini and Akullian's (2007) frequently cited meta-analysis that focused on VM as an intervention for individuals with autism spectrum disorders (ASD) was conducted. The results revealed that if Bellini and Akullian had conducted their review using updated CEC 2014 standards, VM applied to individuals with ASD would not have been classified as an EBP.

Key Words: Evidence-Based Standards; Special Education; International Special Education; International Standards; Standards-Based Education; Contextual Education; Individualized Education

Within the field of special education, much of the literature has consisted of studies utilizing single case designs (Freeman & Sugai, 2013). Single case methodology is particularly useful when examining groups of individuals who exhibit great variability of attributes. The use of single case designs within special education has led to an ongoing discussion within the field of exceptional children regarding how to define any particular body of literature as

pointing to an evidence-based practice (EBP) that should be adopted widely by practitioners. While several authors and entities have attempted to define EBPs regarding special education teaching techniques, each attempt contains limitations. This is especially true when such efforts have attempted to classify a literature base that consists largely of research involving the use of single case methodologies. More importantly, no single accepted set of criteria exists internationally to which practitioners can turn in order to find EBPs relevant to the contexts, languages, cultures, and educational goals of any specific instance of student need (Hornby, Gable, & Evans, 2013).

Legislation in the United States (U.S.) and elsewhere has made understanding EBPs crucial to both researchers and practitioners. Most applicable to U.S. special education professionals is the *Individuals with Disabilities Education Improvement Act* (2004) which suggests that peer-reviewed research be used to guide practice and, more recently, the *Every Student Succeeds Act* (2015) which includes the words "evidence-based" 45 times (ESSL, 2015).

Outside the U.S., calls echo those of the U.S. to improve the quality of learning through providing children with well-educated, quality practitioners who utilize evidence-based teaching practices. Furlong, Cochran-Smith, and Brennan (2009) comment concerning their review of a collection of internationally peer reviewed articles, that all education practitioners require training in defining and understanding teaching practices that have been scientifically verified as part of an international focus on improving the quality of education worldwide.

In Canada, this question is of particular interest. Efforts to respect and incorporate a multicultural approach to teaching and learning has challenged more traditional Western type practices still widely used and accepted within institutes of higher education. Teaching and learning professionals have been challenged to re-define and examine traditional epistemologies and ontologies regarding the practices of teaching and learning (Madden, 2015). Such efforts are reflected in new local and provincial policies regarding the incorporation of such multi-cultural foundational views to include First Nations' traditional ways of knowing (Truth and Reconciliation Commission of Canada, 2015) and others as valid and useful approaches to the ongoing formation and dissemination of knowledge.

Hornby et al. (2013) examined international literature from Australia, England, Hong Kong, New Zealand and the U.S. to analyze similarities and differences in the ways international locations approach the definition of EBPs within education. These authors noted several complicating factors in defining EBPs once the search expands beyond cultural, language, political and/or geographic borders. These authors noted that there "is an inadequate sharing of information between professionals...(that) has resulted in the development of isolated examples of successful practices that are not easily disseminated across cultural and political settings" (Hornby et al., 2013, p. 121). This is particularly problematic as the world becomes smaller through globalization, advanced transportation and geographic migration of peoples at a far greater rate than in previous times. The findings of these authors suggest that as classrooms become more and more diverse, it is likely not sufficient to have any one list of EBPs such as is listed in the What Works Clearinghouse (U.S. Dept. of Education: IES, 2018) and the Best Evidence Encyclopedia (Johns Hopkins University CRRE, 2018). Educators need to consider context as a foundational piece of methodology effectiveness (Hornby et al., 2013). Additionally, social validity is also likely foundational in choosing the best EBPs as applied to any particular individual student even after a particular methodology may have passed muster as an EBP by any individual set of criteria (Hornby et al., 2013).

Within the field of special education, the Council for Exceptional Children (CEC) devoted a special issue of their journal *Exceptional Children, vol. 71(2)* in 2005 to providing quality indicators for research studies and proposed standards with which to classify the quality and quantity of research in support of special education related treatments and/or instructional methods. Key within this issue, Horner et al. (2005) defined "single case research (as) experimental rather than correlational or descriptive, and its purpose (is) to document causal, or functional, relationships between independent and dependent variables" (p. 166). Additionally, Kennedy (2005) defined single case designs as being "used to demonstrate experimental control within a single participant" (p. 12) and defended the use of single case research designs to define evidence-based practices stating that one person is used as both the control and the participant to obtain internal validity. In single case designs, threats to internal validity are controlled through manipulation of a single independent variable and noting any effects on a single outcome dependent variable (Kennedy, 2005). Clearly, those engaged in single case research have attempted to define and solidify the usefulness of this methodology in identifying interventions and strategies as evidence-based at the level of the individual learner.

However, skepticism remains as to if a collection of single case studies may be sufficient to provide evidence that a particular intervention is effective for specific groups by those more accustomed to larger-scale, quantitative true or quasi-experimental study designs (Horner, Smolkowski, Sugai, & Swaminathan, 2012). Horner et al.'s (2005) article addressed this and outlined seven quality indicators to examine the methodological soundness of single case studies. These authors implied that only studies that employ these standards of quality should be examined through meta-analysis to reach conclusions as to the evidence base of a given educational or behavioral intervention. These seven quality indicators included: 1) a

thorough description of the study's participants and setting, 2) an operationalized definition of the dependent variable, 3) an operationalized definition of the independent variable, 4) a baseline time period sufficient to decipher a pattern, 5) the establishment of experimental control, 6) the establishment of external validity, and 7) the establishment of social validity (Horner et al., 2005). Key within these seven quality indicators is an understanding that population and setting play a significant role in understanding the context of any researched intervention and must be considered as much as the experimental rigor of the research designs when analyzing the effectiveness of any methodology under study. Following publication in 2005, Horner et al.'s seven quality indicators were largely adopted as the standard by which to judge the quality of single case research studies within the field of special education (Wendt & Miller, 2012).

In 2012, Horner et al. revisited the idea of how single case research designs could be used to meet the continuing demands that educational research be conducted using rigorous experimental designs and that a focus on meta-analysis be used to determine any single intervention as an EBP (Horner et al., 2012). Horner et al. (2012) called for the reconsideration of using statistical analysis to summarize and evaluate the outcomes of single case research. These authors pointed out that the size of experimental effects represents one essential element in ascertaining the evidence base for a practice while also suggesting that visual analysis of individual single case results represents another essential element. They suggested that the body of research evidence supporting a practice should 1) show an experimental or causal effect, 2) document the size of any experimental effect, and 3) validate the social importance of any findings. Again, this social importance element is listed as critical in interpreting a selection of any set of literature as being classified as an EBP.

Several prominent entities, including the What Works Clearinghouse (U.S. Dept. of Education: IES, n.d.), the Evaluative Method (Reichow, Volkmar, & Cicchetti, 2008), and the Certainty Framework (Simeonsson & Bailey, 1991), have attempted to develop standards for the use of single case research by placing scaled metrics around the quality of single case research designs (Wendt & Miller, 2012). While these standards and metric-based measurement tools offer a means of classifying the quality of single case designs, Wendt and Miller (2012) found that using multiple tools to evaluate individual studies often produces different results. With each tool offering a different quality assessment of any given single case study, questions remain as to how a body of quality single case studies can be identified and then examined through meta-analysis to establish the evidence base of a specific intervention that warrants the technique be employed on a larger scale (Wendt & Miller, 2012).

More recently, Freeman and Sugai (2013) again attempted to assist practitioners in choosing EBPs in the field of special education. They discussed typical methods of quality analysis involved in analyzing the ability of multiple single case research studies to point to EBPs. Still, these authors pointed out that "the body of special education research is varied and clear standards for EBPs are not yet clearly defined" (Freeman & Sugai, 2013, p. 12).

Important to current discourse, The Council for Exceptional Children (CEC) commissioned a group of leading researchers in the field of special education to work on the critical issue of defining EBPs. This work group was tasked with formulating a current set of standards with which to evaluate a body of literature on specific instructional practices as evidence-based. Providing details to the processes involved, Cook et al. (2015) described how the group utilized the foundational work of Horner, Gerston, Kratochwill, Cook and others to formulate and test 28 indicators of quality research and to establish standards to classify the evidence-base of an instructional practice based on methodologically sound studies. The CEC 2014 standards of EBPs (Cook et al., 2015) are separated into two sections – quality indicators to examine the methodological soundness of studies and standards for classifying the evidence base of practices on the basis of sound studies (Cook et al., 2015). The group adopted quality indicators of methodologically-sound studies to include a sufficient description of the study context and setting, a thorough description of participants, and a description of the intervention agents (Cook et al., 2015). Also included in the quality indicators was an assessment of implementation fidelity, evidence of internal validity, appropriate application of outcome measures, and appropriate data analysis (Cook et al., 2015). Based upon the quality indicators, the group then provided standards for classifying the effects of a practice. Within this classification system, a practice could be categorized as a) evidence-based, b) potentially evidence-based, c) practice with mixed effects, d) insufficient evidence, or e) practice with negative effects (Cook et al., 2015). There are 28 quality indicators total; however, 6 apply only to group comparison studies and 4 apply only to single-case type studies with 18 applying to both group comparison and single-case type studies. As a result, there are 24 quality indicators that apply to group comparison studies and 22 that apply to single-case studies.

Further complicating the dissemination of information regarding EBPs to practitioners is the abundance of web-based resources that propose to supply readers with information regarding which practices are evidence-based. Test, Kemp-Inman, Diegelmann, Hitt, and Bethune (2015) identified 47 web sites that professed to list EBPs regarding the teaching of children with exceptionalities. Upon probing the sites, the authors ultimately classified 43% as "Do Not Trust" (Test et al., 2015).

Clearly the practitioners and researchers within the field of education continue to refine and define criteria with which to judge any specific practice as evidence based. Also clear is that the commonly available 'lists' of evidence based practices in education are typically based on differing sets of criteria that in the best of cases concentrate on the meta-analytic analysis of scientifically valid studies with only a small regard for linguistic, cultural, social, religious or political context. In the worst of cases, such lists seem to be unrelated to any criteria of scientific rigor or review at all. It is the goal of this work to analyze how changing criteria may affect past conclusions as to if any specific method may be counted among those that are evidence based and thus to analyze the hypothesis that lists of EBPs may need to change as context and criteria change over time. This work also tests a secondary hypothesis that examines the validity of defining methods as evidence based through any analysis of single case research studies that by definition, only examine any intervention as it relates to a specific individual subject.

VIDEO MODELING AS EXEMPLAR

Video modeling (VM) represents one relatively common behavioral and instructional intervention that is used in part, to support the learning of students with autism spectrum disorders (ASD). Research dating back to the 1970s demonstrates the effectiveness of VM for students with ASD (e.g., Charlop & Milstein, 1989; Charlop & Walsh, 1986; Creer & Miklich, 1970; Dowrick & Raeburn, 1977). Thus, VM is an intervention whose effectiveness has been well-researched and documented in the literature.

When examining the effects of differences in criteria for EBPs, VM serves as a particularly good focus given that the method assumes that the videos are produced in culturally, linguistically, and socially valid ways across contexts. For example, a video model portraying appropriate joyous response (clapping after a performance) for a child with ASD in a Euro-centric culture would be socially valid for that context. By contrast, a similar video model used to show appropriate ways to greet or show support in a First Nations context in Western Canada may show the traditional raised hand gesture. Both would thus be equally socially valid in their relevant contexts while maintaining methodological standards required to implement VM as an evidence based instructional strategy. This technique of VM would seem somewhat suited for use as subject matter for this study precisely because the technique largely eliminates contextual differences in the ways the method is utilized.

Regarding the specific instructional/behavioral intervention of VM, several meta-analyses have been conducted in an effort to classify VM (in several variations) as an EBP to support the learning of individuals with ASD. Mason, Ganz, Parker, Burke, and Camargo (2012) analyzed peer and adult actor type video modeling in an effort to determine whether this method should be classified as an EBP. These authors identified 106 participants with ASD across 42 studies, 27 of whom had a comorbid diagnosis. The authors' methodological evaluation criteria did not include all of the CEC 2014 standards for EBPs, as this article was written prior to the standards being published. Still, the authors concluded at the time of publication that VM with other as model is "highly effective" (Mason et al., 2012, p. 1076) for instruction of individuals with ASD.

Another meta-analysis often cited when considering VM as an EBP was conducted by Shukla-Mehta, Miller, and Callahan (2010). The outcomes of interest to these authors were social and communication skills. Since this article also preceded the release of the CEC 2014 standards for EBPs, the authors did not utilize them in selecting articles to include in their meta-analysis. Of the 20 studies found to meet their inclusion criteria, only four examined VM in isolation (i.e.: no other support or intervention was used in addition to VM); other studies used confounding and extraneous instructional techniques in addition to VM bringing into question Kennedy's (2005) definition of single case research as involving a single independent and dependent variable. Still, the authors identified VM as essentially effective.

Bellini and Akullian (2007) conducted a widely cited meta-analysis in which they sought to discern the effectiveness of video-based instruction for students with ASD in terms of a) behavioral functioning, b) social-communication skills, and c) functional life skills. The authors examined 23 peer-reviewed, published works that utilized single-case research designs and that demonstrated experimental control through multiple-baseline, reversal, and/or alternating treatment methods. This methodology was true to the accepted guidelines at the time the article was written. The authors concluded that VM and video self-modeling (VSM) interventions for students with ASD should be considered EBPs. While this study was welcome news for researchers and practitioners examining video-based interventions and those well-versed in single-case research, the authors could not employ the study selection criteria advocated by authors in more recent years. Since Bellini and Akullians' (2007) article is overwhelmingly used as a citation for the assertion that VM is an EBP for individuals with ASD, the new CEC 2014 standards for EBPs should be applied so that the conclusions regarding the effectiveness of VM for this population can be updated appropriately.

PURPOSE

The purpose of this study was to re-examine Bellini and Akullian's (2007) highly influential meta-analysis to determine whether VM meets the CEC 2014 standards for EBPs outlined in Cook et al. (2015). All 23 studies included in the original Bellini and Akullian work were examined using the new CEC 2014 standards for EBPs (Cook et al.,

2015) that relate to single case studies to better understand how changing EBP criteria might lead to the classification or declassification of any particular practice (in this case VM) as an EBP. This re-examination is important not only to ensure practitioners are provided with the most up-to-date information regarding VM as an EBP, but also in analyzing how changes in standards might inform future discourse in continuing to define EBP standards that align with ever increasing diverse student populations in a world- wide multi-cultural context. Such review also represents an essential re-evaluation of a highly-cited source to either reinforce findings and further citation or to indicate a need for new meta-analytic work using newly refined standards be conducted.

METHOD

To begin the analysis, the original articles examined by Bellini and Akullian (2007) were obtained. Within their article, the authors outlined their method for searching and obtaining articles as well as how the articles were narrowed to 23 that met their inclusion criteria. Thus, these 23 articles that met Bellini and Akullian's (2007) original inclusion criteria were then re-examined to determine if they met the new CEC 2014 quality indicators for methodological soundness as it relates to single case studies.

Initial analysis was conducted through independent quality review of the 23 articles. Reviews were conducted by the authors of this work after careful review and discussion concerning the 22 relevant CEC 2014 quality indicators as published by Cook et al. (2015) and on the CEC website. Initial reviews were conducted by the first and third authors using the 22 quality indicators applicable to the single-case studies. Following the independent examination of each article by the first and third authors, any disagreement between these reviewers on one or more of the 22 quality indicators was examined independently by the fifth author. The fifth author's independent review of a disputed indicator determined its final disposition (i.e., two of three reviewers agreed as to the indicator being met or not met). The analysis of the original 23 articles in this way was sufficient to explore how the changing quality indicator criteria for examination of studies may affect the ultimate classification of the evidence base for VM found in Bellini and Akullian's original work.

Next, classification of articles that were determined to be methodologically sound by at least two of three reviewers in all 22 criteria areas related to single-case studies were examined using single-subject phase trend visual analysis and points of non-overlapping data analysis. These data analysis methods were used to determine if each study revealed a positive, negative, or neutral effect using the standards for classifying the evidence base of practices (Cook et al., 2015). The use of single-case studies and the

visual analysis of data is common in clinical settings to demonstrate a change over time when investigating phenomena on a small number of participants (Bobrovitz & Ottenbacher, 1998). Additionally, it represents the traditional means by which single-case data has been evaluated (Horner et al., 2005; Park, Marascuilo & Gaylord-Ross, 1990; Parker, Hagan-Burke, & Vannest, 2007).

Visual analysis consisted of reviewing each graph per participant and visually observing the trends between baseline and following periods to decipher any differences in slope and/or mean data levels between treatment conditions and baseline (Cook et al., 2015; Scruggs, Mastropieri, & Regan, 2006). Next, the variability of the data around a mean level by phase and an analysis of the "immediacy of any observed treatment effect" was noted (Cook et al., 2015, p. 229). This analysis was independently performed by the first and third authors. No tie breaker analysis was needed by the fifth author since the first and third author agreed as to the nature of the trends between baseline and intervention periods for the analyzed three studies. Although traditional visual analysis represents a relatively subjective approach to interpreting single-subject data (Alresheed, Hott, & Bano, 2013), it remains histori- cally recommended as a piece of treatment effect analysis (Franklin, Gorman, Beasely, & Allison, 1996; Robey, Schultz, Crawford, & Sinner, 1999) and is included within the new CEC 2014 standards for EBPs (Cook et al., 2015).

Although controversial, percentage of non-overlapping data (PND) analysis represents a common method of reporting effect size when analyzing single case data (Maggin, Briesch, & Chafouleas, 2013; Scruggs & Mas- tropieri, 2013). While many methods of analysis exist for evaluating effect size using variations of non-overlapping data calculations, each has limitations (Alresheed et al., 2013; Kratochwill et al., 2010). As PND represents one of the oldest and most well-known methods to calculate effect size (Alresheed et al., 2013) when analyzing a multitude of single case designs, it was used in this study to represent a non-regression approach to quantitative effect size analysis.

PND was calculated by observing how many data points collected during intervention exceeded the upper- most data point collected during baseline (Alresheed et al., 2013). That number was then divided by the total number of observations during the intervention period to obtain a PND score. This method was used on each of the skills performed from each of the participants in each study. Additionally, two authors independently calculated PND effect size to account for any visual judgement differences that might occur when working with graphs rather than raw data.

A note concerning effect size when analyzing single case research designed interventions is important. CEC's published Standards for Evidence Based Practices in Special

Education published online and noted earlier acknowledge many ways in which the effects indicated through analysis of single-case designs can be examined. These guidelines stress that a "functional relationship is established by reviewers' use of standard methods of visual analysis" (p.7). These guidelines continue to specifically indicate several examples of functional relationship analysis to include examining data across phases through observing or calculating: the "level of mean scores between phases... the slope of data within a phase,... (the) variability (or) range of scores around a level or trend line,... assessments of the immediacy of any observed treatment effect... and (or) overlap of data points across phases" (p.7). Specifically, three quality indicators relate to the calculation and interpretation of effect size within the new CEC 2014 standards. Quality indicator 6.0 (Internal validity: 6.1 & 6.5) mandates that the independent variable is under control of the experimenter and any changes in the dependent variable(s) are caused by the manipulation of the independent variable. Quality indicator 7.0 (Outcome measures: 7.2 & 7.3) suggest that outcome measures are able to gage effect size and are appropriately measured. Quality indicator 8.0 (Data analysis: 8.2 & 8.3) mandates that any analysis of data is conducted appropriately and that effect size is reported. Clearly, the CEC guidelines acknowledge that some disagreement remains as to how to uniformly assess the effect size of any treatment effect when examining multiple single case studies related to the variables of interest. While some argue that points of non-overlapping data should be used to confirm visual analysis when conducting meta-analysis of single case designed studies (Scruggs & Mastropieri, 2013), others suggest that the method does not represent a true effect size since the method may not indicate a magnitude of change (Wolery, Busick, Reichow & Burton, 2010). We report the effect size in this work using the points of non-overlapping data approach to replicate the original method of reporting provided in Bellini and Akullian's original work however it should be noted that overall visual analysis (to include PND analysis) as provided for in CEC's guidelines was used in the present re-evaluation to analyze any functional relationship in the studies meeting all of CEC's quality indicators.

RESULTS

A review of all 23 articles was initially conducted independently by the first and third authors. Agreement between these two authors concerning the CEC 2014 quality indicators across all articles was measured to be 87.15%. Because the first and third author agreed that all CEC 2014 quality indicators concerning single subject designs had been met for three of the articles reviewed, a third reviewer was not needed in order to meet the two of three reviewer agreement threshold for those three articles. All the remaining articles ($n=20$) were reviewed by the fifth

author in order to resolve each disagreement between the first and third author. Inter-observer agreement (IOA) between the first and fifth author regarding the 20 articles reviewed by the fifth author was calculated to be 97.25%. IOA between the third and fifth author regarding these same 20 articles was calculated to be 87.87%.

The percent with which the original 23 articles met Council for Exceptional Children (2014) standards for EBPs ranged from 100% ($n=3$) to 45.45% ($n=1$). Thus, only three of the original 23 articles were found to meet all of the CEC 2014 quality indicators to be considered methodologically sound (see Table 1). Only three of the new CEC 2014 quality indicators were met by all 23 articles (indicators 2.1, 4.1 and 7.1; participant demographics, detailed intervention procedures, socially important outcomes) while indicators 3.1 and 3.2 (describing the role and training of the intervention agent) received the least affirmative reviews (see Table 2).

Using Cook et al.'s (2015) approach, the three studies that met CEC 2014 methodological requirements were individually classified as having a positive, neutral/mixed, or negative effect. According to Cook et al., a positive effect can be concluded if "functional positive relations are established for at least three-fourths (75%) of the relevant cases...with a minimum of three total cases; and no functional negative relations are observed" (Cook et al., 2015, p. 229).

Only two of the three articles that had been found to be methodologically sound under CEC 2014 standards met the standards to be regarded as having a positive effect. Charlop-Christy, Le, and Freeman (2000) examined differences between VM and in-vivo (real life) modeling for five participants with ASD. The authors utilized a multiple-baseline across skills and participants design. Targeted skills included social communication, conversational skills, spontaneous greetings, and play behaviors. Some prompting and positive reinforcement were used during baseline, however, they were not used during the VM intervention. Overall intervention PND was calculated to be 82%. Visual analysis concluded overall positive effects for all participants during the intervention phases of the study. Therefore, this study met CEC 2014 standards for classification as having a positive effect.

Norman, Collins, and Schuster (2001) examined the effectiveness of VM to teach self-help skills to two participants with intellectual disability and one participant with ASD. The authors utilized a multiple probe design across behaviors replicated across participants coupled with a constant time delay procedure. Targeted skills included cleaning sunglasses, putting on a wrist watch, and zipping a jacket. Overall intervention PND was calculated to be 74%. Visual analysis concluded overall positive effects for all participants during the intervention phases of the study. Because only one participant of three participants was noted

Table 1
Bellini & Akullian Articles/CEC 2014 Standards Not Met

Article	Findings	
	CEC standard(s) not met	Percent agreement - CEC 2014 standards
Alcantara, P. R. (1994)	3.1, 3.2	91.30
Apple, A. L., Billingsley, F., Schwartz, I. S., & Carr, E. G. (2005)	6.5	95.65
Charlop, M. H., & Milstein, J. P. (1989)	3.1, 3.2, 6.1	86.96
*Charlop-Christy, M. H., Le, L., & Freeman, K. A. (2000)	0	100.00
D'Ateno, P., Mangiapanello, K., & Taylor, B. A. (2003)	3.1, 3.2	91.30
Haring, T. G., Kennedy, C. H., Adams, M. J., & Pitts-Conway, V. (1987)	3.1, 3.2, 5.1, 5.2, 5.3	78.26
MacDonald, R., Clark, M., Garrigan, E., & Vangala, M. (2005)	3.1, 3.2, 5.1, 5.2, 5.3	78.26
Mechling, L. C., Pridgen, L. S., & Cronin, B. A. (2005)	3.2	95.65
Nikopoulos, C. K., & Keenan, M. (2003)	3.1, 3.2, 5.1, 5.2, 5.3	78.26
Nikopoulos, C. K., & Keenan, M. (2004)	1.1, 3.2, 5.1, 5.2, 5.3,	78.26
*Norman, J. M., Collins, B. C., & Schuster, J. W. (2001)	0	100.00
Ogletree, B. T., & Fischer, M. A. (1995)	1.1	45.45
	4.2	
	5.1, 5.2, 5.3	
	6.2, 6.5, 6.6, 6.7, 7.3, 7.4, 7.5	
Sherer, M., Pierce, K. L., Paredes, S., Kisacky, K. L., Ingersoll, B., & Schreibman, L. (2001)	3.1, 3.2, 5.2, 5.3	82.61
Shipley-Benamou, R., Lutzker, J. R., & Taubman, M. (2002)	3.1, 3.2, 5.1, 5.2, 5.3	78.26
Simpson, A., Langone, J., & Ayres, K. M. (2004)	3.1, 3.2	91.30
Taylor, B. A., Levin, L., & Jasper, S. (1999)	3.1, 3.2, 5.1, 5.2, 5.3	78.26
Coyle, C., & Cole, P. (2004)	3.1, 3.2	91.30
Buggey, T. (2005)	5.1, 5.2, 5.3	86.96
Buggey, T., Toombs, K., Gardener, P., & Cervetti, M. (1999)	1.1	69.57
	3.2, 3.2, 5.1, 5.2, 5.3	
	6.3	
Hagiwara, T., & Myles, B. S. (1999)	3.1, 3.2, 5.1, 5.3	82.61
Lasater, M. W., & Brady, M. P. (1995)	3.1, 3.2, 5.1, 5.3	82.61
Thiemann, K. S., & Goldstein, H. (2001)	3.1, 3.2	91.30
*Wert, B. Y., & Neisworth, J. T. (2003)	0	100.00

as having an ASD, this study could not meet CEC 2014 standards for classification as having a positive effect. No negative effects were reported within this study thus the study could not be classified as having a negative effect. According to CEC classification guidelines, when a study neither meets positive or negative effect standards, the study should be classified as having neutral or mixed effects thus this article is classified as such in the present work.

The final study, by Wert and Neisworth (2003), utilized a multiple-baseline across four participants with ASD in an effort to examine the effects of video self-modeling (VSM) on spontaneous requesting behaviors. Positive trends were noted via visual analysis for all four participants between baseline and intervention phases. Overall PND was calculated to be 96%. Therefore, this study met CEC 2014 standards for classification as having a positive effect.

Table 2
Percent of EBP Standards Met

CEC standard #	Findings	
	%	*f*(n)
1.1	86.96	20
2.1	100.00	23
2.2	95.65	22
3.1	34.78	8
3.2	26.09	6
4.1	100.00	23
4.2	95.65	22
5.1	43.48	10
5.2	60.87	14
5.3	65.22	15
6.1	95.65	22
6.2	95.65	22
6.3	86.96	20
6.5	86.96	20
6.6	95.65	22
6.7	91.30	21
7.1	100.00	23
7.2	95.65	22
7.3	95.65	22
7.4	95.65	22
7.5	95.65	22
8.2	91.30	21

In total, two studies (i.e., Charlop-Christy et al., 2000; Wert & Neisworth, 2003) met the CEC 2014 methodological standards and the CEC 2014 standards to be classified as having a positive effect. However, only nine participants with ASD were examined across these two studies, missing the required 20 participants across five studies necessary to be considered an evidence-based practice. This is a significant change from Bellini and Akullian's original findings that utilized standards current at the time of their 2007 study.

These findings do, however, allow for VM to be classified as a *potentially EBP*. According to CEC 2014 standards for EBPs, the findings included "two to four methodologically sound single case studies with positive effects [and]...include no methodologically sound studies conducted with negative effects, and at least a 2:1 ratio of methodologically sound studies with positive effects compared to methodological sound studies with neutral/mixed effects" (Cook et al., 2015, p. 9). Despite the fact that all three studies examined the effect of different types of VM "packages" (with and without other co-occurring interventions) on various dependent variables, one can broadly

conclude that VM, in general, would have been deemed a potentially EBP when used to teach skills to participants with ASD based on the re-examination of Bellini and Akullian's original work. No specific conclusions can be made beyond this very general statement based on the application of CEC 2014 standards to only the studies included within Bellini and Akullian's original 2007 study.

DISCUSSION

The purpose of this study was to analyze how changing criteria may affect past conclusions as to if any specific method may be counted among those that are evidence based. This work also tested the secondary hypothesis that examined the validity of defining methods as evidence based through any analysis of single case research studies. With regard to the first inquiry, the re-examination included reviewing each of the 23 articles included in Bellini and Akuillan's (2007) meta-analysis using the quality indicators that focused on the methodological soundness of single-case studies to understand how the changes might affect the findings of their original work. Only three quality indicators (2.1 – participant demographics, 4.1 – detailed intervention procedures, and 7.1 – socially important outcomes) were met by all 23 studies. Additionally, only three studies (i.e., Charlop-Christy, Le, & Freeman, 2000; Norman, Collins, & Schuster, 2001; Wert & Neisworth, 2003) met all 22 quality indicators.

The indicators most often missed were 3.1 – describe the role of the intervention agent, 3.2 – describe training needed to implement the intervention, and 5.1 – assess and report intervention fidelity. These indicators were missed by 15, 17, and 11 of the 23 articles, respectively (see table 1). These findings suggest that many of the articles identified in the Bellini and Akullian's (2007) meta-analysis did not provide a significant amount of information on the role of the intervention agent, meaning the person responsible for implementing the intervention (e.g., teacher, researcher, paraprofessional, parent) or a sufficient description of said person. Further, many of the articles lacked a clear description of the training necessary or provided to intervention agents in implementing the intervention. This information relates to the ability of a practitioner to learn required information that would be needed in order to implement the intervention in an appropriate manner and in accordance with required parts to replicate any evidence based practice.

Most important, 11 of the articles lacked clear information as it related to implementation fidelity meaning the implementation procedures were not directly assessed or reported. Implementation fidelity is important as it relates to EBPs because implementation is the critical link between research and practice (Cook & Odom, 2013). The fidelity with which an intervention is implemented affects its level of effectiveness (Carroll et al., 2007). In order to

ensure that practitioners are engaged in practices that lead to improved outcomes for students, they must correctly and effectively implement the EBPs. The EBP cannot yield the intended outcome if it is not implemented correctly (Cook & Odom, 2013). Thus, the need for a direct assessment and report of implementation fidelity within studies is crucial. While findings did show that detailed intervention procedures were provided in all 23 examined studies, the lack of intervention fidelity information leave open the question as to if procedures were followed as indicated in the studies not meeting criteria in this re-evaluation of the 23 studies.

Of note regarding the quality indicators that were met by all 23 studies examined, participant demographics and socially valid outcomes point to historically acknowledging the effects of demographics and context when examining any given intervention as EBP. Still, the original work completed by Bellini and Akullian and even the present re-evaluation of that work within the new CEC 2014 guidelines only provide for the understanding that such information was included in the works analyzed. After indicating that the analyzed studies contained such indications, the findings of the original work and of this re-evaluation of the work report an outcome categorization of the method and mostly does not report any contextual caveats that may be relevant. In practice, the methodology is most typically categorized in general and not specific to any diverse set of differences within an overall population. For example, within the three articles used to indicate that VM be categorized as potentially EBP in the re-examination here, five children (ages 7-11) experienced their intervention in a clinical setting, one (aged 12) in a school setting and four (ages 3-6) in both a home and school setting. The question becomes: For whom is the finding based? For whom is the citation being used as evidence that the particular intervention of VM is potentially EBP? While surely the CEC 2014 standards imply that specific demographic limiters are specified within the ultimate classification findings to include only those represented in the included works, the diversity of the subjects included within single case designed studies involving children with disabilities may not be homogenous enough to make a claim of EBP that is useful for a population subset of reasonable size.

Of interest, findings of other recent applications of the CEC (2014) standards are worth review. Considering a recent application of the new CEC standards of EBPs to another method and population of subjects, Losinski, Wiseman, White, and Balluch (2016) conducted a meta-analysis of VM interventions for reducing challenging behaviors for students with EBD. These authors applied the CEC 2014 quality indicators to the body of found literature that was applicable to their population and method but found that "none of the 12 included studies (within their analysis) met all the quality indicators set forth

by CEC" (Losinski et al., 2016, p. 243). Faced with this reality, the authors were forced to conclude that VM as applied to the reduction of challenging behaviors for students with EBD "requires more systematic replications before (the method) can be considered an evidence-based practice" (Losinski et al., 2016, p. 250). Interestingly, Losinski et al. noted that CEC quality indicators 3.1 (description of intervention agent), 3.2 (description of training and qualifications), and 5 (implementation fidelity) were most frequently missing from their analyzed studies. These are the very same indicators with which the majority of studies analyzed in this re-analysis of Bellini and Akullian's work also did not meet. This would suggest the possibility that these quality indicators may be lacking in research designs across populations of focus, at least with regard to video modeling interventions. Further analysis of different interventions and different populations is needed to identify if these trends go beyond what is being revealed in this work.

LIMITATIONS

Important to this work is that one should not view these findings as calling into question VM as an EBP when applied to support students with ASD. Several articles have been published previous to and since the 2014 standards were adopted that examine VM in various forms and as applied to various populations and outcome skills across contexts. The body of work published since the Bellini and Akullian (2007) is extensive and thus current publications regarding VM as an instructional approach to teaching students with ASD should be used in determining the current evidence base of VM for this population of students. This work sought to examine changes in EBPs standards and how changing standards may impact the classification and possible choice to use any specific practice. VM was used only as an exemplar instructional practice upon which to explore the concepts and practicality of employing the notion of EBP within special education's predominant single case designed literature base.

We further acknowledge that debate exists concerning the utility of a meta-analysis (a deductive approach) being applied to single case research (an inductive approach) at all (Horner et al., 2005). This current work adds to this debate by examining changing meta-analytic protocols and finding that such changes may affect meta-analytic findings over time. This work does not support or detract from this debate but rather seeks to add information to the relevant literature.

FUTURE DIRECTIONS

The findings from Losinski et al. (2016) and the current work may cause one to question what impact *new* standards have on an *old* evidence-base. Are those practices that were once considered EBPs no longer evidence-based if they do

not meet the new standards? In the case of video modeling, clear positive outcomes over many studies and many years would not intuitively appear to be invalid. A clear shift in the knowledge base as it relates to EBPs has occurred and indeed represents the essence of ongoing attempts to further define best practices regarding research to practice over time. It is critical to point out that because a practice does not meet the standards to be considered an EBP by any given set of criteria, it does not mean the practice is necessarily ineffective (Cook & Odom, 2013). On the contrary, it suggests that more high-quality research needs to be conducted on that practice to further demonstrate its effectiveness.

Some questions still remain: Are we possibly restricting teaching practices that might be effective at the level of the individual learner by moving to any mandate to use only EBPs? Does any specific EBP label provide for a list of acceptable teaching practices that result in greater student growth than other practices? In practice, are guidelines regarding criteria for EBP too loose regarding which demographic context to which a given label is applicable? Should EBP standards apply across cultures and nationalities? Surely use of ineffective fad methodologies with no evidence of effectiveness should be eliminated within teacher tool kits but the question remains as to how exactly to define each practice in a way that draws an appropriate line between useful and non-useful practices at the level of single diverse individual learners within individual contexts.

At its' core, single case designs are intended to compare reactions to an intervention at the level of the individual subject. Legislation in multiple forms and from multiple nations have placed an emphasis on defining and refining easy to use 'lists' of teaching methods that are evidence based. This work suggests that a great deal of further research, discussion and thought should be conducted that should examine the following: 1) How do different published criteria regarding evidence based practices affect the use of any particular practice by practitioners? 2) To what extent do current meta-analysis of predominantly single-case research literature bases go beyond just classification of a methodology and refine such classifications by context and demographics of the studied population? and 3) To what extent are practice informed strategies affected by mandated use of evidence based strategies? These questions represent a need to investigate not only classification paradigms but the pros and cons of classification itself regarding EBPs as applied to highly diverse populations and across cultural and geopolitical contexts. A great amount of further research that focuses not only on the *use* of standards to define individual EBPs but also that examines the *effects* of the EBPs designation on the selection of teaching methodologies needs to occur. Examination of any relationship between the designation of a practice as EBP by any particular criteria and actual use of and effectiveness of a method by practitioners within widely diverse contexts remains to be examined.

REFERENCES

Alresheed, F., Hott, B. L. & Bano, C. (2013). Single case research: A synthesis of analytic methods. *Journal of Special Education Apprenticeship*, 2(1), 1-18.

Bellini, S., & Akullian, J. (2007). A meta-analysis of video modeling and video self-modeling interventions for children and adolescents with Autism Spectrum Disorders. *Exceptional Children*, 73(3), 264-287. doi: 10.1177/001440290707300301

Bobrovitz, C. D., & Ottenbacher, K. J. (1998). Comparison of visual inspection and statistical analysis of single case data in rehabilitation research. *American Journal of Physical Medicine and Rehabilitation*, 77(2), 94-102. doi: 10.1097/00002060-199803000-00002

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science,* 2(1), 40-49. doi: 10.1186/1748-5908-2-40

Charlop, M. H., & Milstein, J. P. (1989). Teaching autistic children conversational speech using video modeling. *Journal of Applied Behavior Analysis,* 22, 275-285.

Charlop, M. H., & Walsh, M. E. (1986). Increasing autistic children's spontaneous verbalizations of affection: An assessment of time delay and peer modeling procedures. *Journal of Applied Behavior Analysis,* 19, 307-314.

Charlop-Christy M. H, Le, L., & Freeman K. A. (2000). A comparison of video modeling with in vivo modeling for teaching children with autism. *Journal of Autism and Developmental Disorders*, 30(6), 537-552.

Cook, B. G., Buysse, V., Klingner, J., Landrum, T., McWilliam, R. A., Tankersley, M. & Test, D. (2015). CEC's standards for classifying the evidence base of practices in special education. *Remedial and Special Education,* 34(4), 220-234. doi: 10.1177/0741932514557271

Cook, B. G., & Odom, S. L. (2013). Evidence-based practices and implementation science in special education. *Exceptional Children,* 79(2), 135-144.

Council for Exceptional Children. (2014). *Council for Exceptional Children Standards for Evidence-Based Practices in Special Education*. Retrieved from http://www.cec.sped.org/~/media/Files/StandardsEvidence%20based%20Practices%20and%20Practice/CECs%20Evidence%20Based%20Practice%20Standards.pdf)

Creer, T. L., & Miklich, D. R. (1970). The application of a self-modeling procedure to modify inappropriate behavior: A preliminary report. *Behavior Research and Therapy, 8*, 91-92. doi:10.1016/0005-7967(70)90040-9

Dowrick, P. W., & Raeburn, J. (1977). Video editing and medication to produce a therapeutic self-model. *Journal of Consulting and Clinical Psychology, 45,* 1156-1158. doi: 10.1037/0022-006X.45.6.1156

ESSA (2015). Every Student Succeeds Act of 2015, Pub. L. No. 114-95 § 114 Stat. 1177 (2015-2016).

Franklin, R. D., Gorman, B. S., Beasely, T. M., & Allison, D. B. (1996). *Graphical display and visual analysis. Design and analysis of single-case research*. Mahwah, NJ

Freeman, J., & Sugai, G. (2013). Identifying evidence-based special education interventions from single case research. *Teaching Exceptional Children, 45*(5), 6-12. doi: 10.1177/004005991304500501

Furlong, J., Cochran-Smith, M., & Brennan, M. (2009). *Policy and politics in teacher education: An international perspective*. New York, NY: Routledge.

Hornby, G., Gable, R. A., & Evans, W. (2013). Implementing evidence-based practice in education: What international literature reviews tell us and what they don't. *Preventing School Failure*, *57*(3), 119-123. doi: 10.1080/1045988X.2013.794326

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The Use of Single-Subject Research to Identify Evidence-Based Practice in Special Education. *Exceptional Children*, *71*(2), 165-179. doi: 10.1177/001440290507100203

Horner, R. H., Smolkowski, K., Sugai, G., & Swaminathan, H. (2012). Considerations for the systematic analysis and use of single-case research. *Education & Treatment of Children, 35*(2), 269-290. doi: 10.1353/etc.2012.0011

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004)

Johns Hopkins University, Center for Research and Reform in Education, Best Evidence Encyclopedia, *retrieved from: http://www.bestevidence.org*

Kennedy, C. H. (2005). *Single case-designs for educational research*. Boston, MA: Pearson Education, Inc.

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

Losinski, M., Wiseman, N., White, S. A. & Balluch, F. (2016). A meta-analysis of video-modeling based interventions for reduction of challenging behaviors for students with EBD. *The Journal of Special Education, 49*(4), 243-252. doi: 10.1177/0022466915602493

Logan, L. R., Hickman, R. R., Harris, S. R., & Heriza, C. B. (2008). Single case research design: Recommendations for levels of evidence and quality rating. *Developmental Medicine & Child Neurology*, *50*(2), 99-103. doi: 10.1111/j.1469-8749.2007.02005.x

Madden, B. (2015). Pedagogical pathways for Indigenous education with / in teacher education. *Teaching and Teacher Education, 51*, 1-15. doi: 10.1016/j.tate.2015.05.005

Maggin, D. M., Briesch, A. M., & Chafouleas, S. M. (2013). An application of the What Works Clearinghouse standards for evaluating single case research: Synthesis of the self-management literature-base. *Remedial and Special Education, 34*(1), 44-58. doi: 10.1177/0741932511435176

Mason, R. A., Ganz, J. B., Parker, R. I., Burke, M. D., & Camargo, L. P. (2012). Moderating factors of video-modeling with other as model: A meta-analysis of single-case studies. *Research in Developmental Disabilities*, *33*(4), 1076-1086. doi: 10.1016/j.ridd.2012.01.016

Norman, J. M., Collins, B. C., & Schuster, J. W. (2001). Using an instructional package including video technology to teach self-help skills to elementary students with mental disabilities. *Journal of Special Education Technology, 16*(3), 5-18

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practice. *Exceptional Children*, *71*(2), 137-148. doi: 10.1177/001440290507100201

Park, H., Marascuilo, L., & Gaylord-Ross, R. (1990). Visual inspection and statistical analysis of single-case designs. *Journal of Experimental Education*, *58*(4), 311-320. doi: 10.1080/00220973.1990.10806545

Parker, R. L., Hangan-Burke, S., & Vannest, K. (2007). Percent of all non-overlapping data (PAND): An alternative to PND. *The Journal of Special Education*, *40*(1), 194-204.

Reichow, B., Volkmar, F. R., & Cicchetti, D. V. (2008). Development of the evaluative method for evaluating and determining evidence-based practices in autism. *Journal of Autism and Developmental Disorders*, *38*(7), 1311-1319. doi: 10.1007/s10803-007-0517-7

Robey, R. R., Schultz, M. C., Crawford, A. B., & Sinner, C. A. (1999). Single-subject clinical outcome research: Designs, data, effect sizes, and analysis. *Aphasiology*. *13*(6), 445-473. doi: 10.1080/026870399402028

Scruggs, T. E., & Mastropieri, M. A. (2013). PND at 25: Past, present, and future trends in summarizing single-subject research. *Remedial and Special Education, 34*(1), 9-19. doi:10.1177/0741932512440730

Scruggs, T. E., Mastropieri, M. A., & Regan, K. S. (2006). Statistical analysis for single subject research designs. In T. E. Scruggs & M. A. Mastropieri (Eds.), *Applications of*

Research Methodology: Advances in Learning and Behavioral Disabilities (Volume 19, pp. 33-53). Bingley, United Kingdom: Emerald Group Publishing Limited. doi: 10.1016/S0735-004X(06)19002-3

Shukla-Mehta, S., Miller, T., & Callahan, K. J. (2010). Evaluating the effectiveness of video instruction on social and communication skills training for children with autism spectrum disorders: A review of the literature. *Focus on Autism and Other Developmental Disorders, 25*(1), 23-36. doi:10.1177/1088357609352901

Simeonsson, R., & Bailey, D. (1991). Evaluating programme impact: Levels of certainty. In D. Mitchell & R. Brown (Eds.), *Early intervention studies for young children with special needs* (pp. 280-296). London, United Kingdom: Chapman and Hall.

Test, D. W., Kemp-Inman, A., Diegelmann, K., Hitt, S. B., & Bethune, L. (2015). Are online sources for identifying evidence-based practices trustworthy? An evaluation. *Exceptional Children, 82*(1), 1-23. doi: 10.1177/0014402915585477

Truth and Reconciliation Commission of Canada. (2015). *Truth and Reconciliation Commission of Canada: Calls to action.* Winnipeg, MB: Author. Retrieved from http://www.trc.ca/websites/trcinstitution/File/2015/Findings/Calls_to_Action_English2.pdf

U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse (n.d.). *Retrieved from: http://ies.ed.gov/ncee/wwc/*

Wendt, O., & Miller, B. (2012). Quality appraisal of single case experimental designs: An overview and comparison of different appraisal tools. *Education and Treatment of Children*, 35(2), 235-268. doi: 10.1353/etc.2012.0010

Wert B. Y., & Neisworth J. T. (2003). Effects of video self-modeling on spontaneous requesting in children with autism. *Journal of Positive Behavior Interventions*, 5(1), 30-34. doi: 10.1177/10983007030050010501

Wolery, M., Busick, M., Reichow, R., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *Journal of Special Education, 44,* 18-28. doi:10.1177/0022466908328009

CORRESPONDENCE

Robert L. Williamson, Ed.D
Simon Fraser University Faculty of Education
Education Bldg #8568
8888 University Drive
Burnaby BC Canada V5A 1S6
robert_williamson@sfu.ca