Review of Systematic Reviews and Meta-analyses of Single-Subject Experimental Studies

Laleh Jamshidi, Mieke Heyvaert, and Wim Van den Noortgate[1]

Abstract

Based on the increasing interest in systematic reviews and meta-analyses of Single-Subject Experimental Designs (SSEDs), the aim of the present review is to determine the general characteristics of these meta-analyses, including design characteristics of the primary studies and the meta-analyses, the kind of data, and the kind of analysis. After a systematic search for studies, 178 studies were examined for various features. The results indicated that the number of SSED meta-analyses and reviews has increased remarkably in recent years. The most frequently used effect size metric was percentage of non-overlapping data (PND) and the most commonly utilized analysis for synthesizing the results of primary SSEDs was a simple average of effect sizes. Based on the findings of the present review, some implications for future SSED meta-analyses are proposed.

Keywords: Single-subject experimental design, meta-analysis, effect sizes, review

Single-Subject Experimental Designs (SSEDs) have been frequently applied in different disciplines to examine the effects of interventions or treatments (Riley-Tillman & Burns, 2009; Schlosser, Lee, & Wendt, 2008; Shadish, 2014a, 2014b; Shadish & Rindskopf, 2007; Smith, 2012). One of the main issues of SSEDs is generalizability. To enhance generalizability, researchers replicate experimental designs across cases. Meta-analytic procedures allow researchers to quantitatively synthesize past research results and provide evidence for best practices (Beretvas & Chung, 2008; Petit-Bois, Baek, Van den Noortgate, Beretvas, & Ferron, 2016; Tincani & De Mers, 2016). The interest in the meta-analysis of SSEDs has increased in the past decade (Shadish, 2014a; Shadish, Hedges, & Pustejovsky, 2014). The use of meta-analysis indeed offers multiple opportunities, including drawing inferences on the overall treatment effect without losing information about the individual cases, or the inclusion of moderator variables (Van den Noortgate & Onghena, 2003). By aggregating several SSED studies together, researchers would be able to study generalizability of the results (Owens, 2011). Tincani and De Mers (2016) noted that when individual studies in the same field have yielded different and inconsistent findings, meta-analyses can be particularly helpful to identify moderating effects of an intervention.

It is crucial that the meta-analysts and reviewers try to apply the most appropriate and efficient procedures. Shadish and Rindskopf (2007) indicated that despite the importance of SSEDs to provide the basis for causal inferences, the ways to meta-analyze these designs is fraught with controversy. An important prerequisite for researchers to select the most appropriate procedures is being aware of different meta-analytic approaches and methods. In addition, getting further insight in the characteristics of primary studies is important for further developing methodological techniques for the (meta-)analysis of SSED data.

Several studies examined published meta-analyses and systematic reviews of SSED studies (Farmer et al., 2010; Beretvas & Chung, 2008; Maggin, O'Keeffe, & Johnson, 2011;

[1] Faculty of Psychology and Educational Sciences & imec-itec, KU Leuven, University of Leuven, Belgium
Corresponding Author:
Laleh Jamshidi, KU Leuven, Etienne Sabbelaan 53, 8500 Kortrijk, Belgium
Email: laleh.jamshidi@kuleuven.be

Schlosser et al., 2008; Shadish & Rindskopf, 2007;). Each of these studies reviewed some characteristics and have some limitations. In general, these previously conducted reviews have some limitations, like 1) focusing on just some aspects of the SSED meta-analyses such as determining the used metrics or the way of handling dependency (Beretvas and Chung, 2008a), 2) reviewing only studies which applied PND (Schlosser et al., 2008), 3) examining only reviews in a certain field (Maggin, O'Keeffe, & Johnson, 2011), or 4) investigating just a few studies done in only one decade (Farmer et al., 2010). It is a fact that the number of SSED studies has grown and plenty of methodological works on SSED (meta-)analytical methods have been done in last decades.

One interesting question is to examine whether, in practice, the SSED (meta-)analyses follow the methodological advancements in the field of meta-analysis regarding designs, effect sizes, and meta-analytic models and techniques. In this study, we tried to do a more comprehensive review to figure out how typical meta-analytic datasets look like from a methodological and analytical point of view. In current paper, we discuss a new review of SSED meta-analyses from the recent three decades (1985-2015), which is updating and extending previous reviews, and provides a general image of the (evolution in) analyses done in this period. It can also help in determining whether the SSED (meta-)analyses have followed methodological and analytical advancements in this field. We answer three main questions: (1) What are the general characteristics and study designs of the primary SSED studies? (2) Which kinds of data have been provided in the primary studies and the reviews? (3) Which kinds of analysis have been done in the primary studies and the reviews?

1. Methods

1.1. Search procedure

Relevant reviews of SSEDs were identified by systematically searching electronic databases, studies in previously published reviews of SSED meta-analyses, relevant journals, bibliographies of relevant articles, citation indexes, and by contacting experts. This search process was done by three people. In a first step, six electronic databases were searched: PsycINFO, Web of Science, Science Direct, Medline PubMed, ERIC, and CINAHL. Afterwards, a search for meta-analyses in previously published reviews of SSED meta-analyses was conducted (Beretvas & Chung, 2008; Farmer et al., 2010; Maggin et al., 2011; Schlosser et al., 2008; Shadish & Rindskopf, 2007). To avoid publication bias, we searched grey literature (i.e. non-published literature) as well. To that end, ten databases were searched: the CORDIS Library, EdITLib, the Grey Literature Database of the Canadian Evaluation Society, the Index of Conference Proceedings, the Index to Theses in Great Britain and Ireland, IBSS, ProQuest Dissertations & Theses, SSRN eLibrary, SIGLE, and Theses Canada. The search string used in all these databases consisted of the following combination of terms: "single case" OR "single subject" OR "N of 1" OR "small N" OR "multiple baseline design" OR "alternating treatments design" OR "reversal design" OR "withdrawal design") AND ("meta-analysis" OR "synthesis" OR "review").

Next, we conducted a hand search of journals that showed up at least three times in the previous search steps (*Advances in Learning and Behavioral Disabilities, American Journal of Mental Retardation, American Journal of Speech-Language Pathology, Augmentative and Alternative Communication, Behavioral Analyst Today, Behavior Modification, Behavior Therapy, Behavioral Disorders, Education and Treatment of Children, Evidence-Based Communication Assessment and Intervention, Exceptional Children, Journal of Autism and Developmental Disabilities, Journal of Behavioral Education, Journal of Clinical Epidemiology, Journal of School Psychology, Journal of Special Education, Learning Disability Quarterly, Neuropsychological Rehabilitation, Remedial and Special Education,* *Research in Autism Spectrum Disorders, Research in Developmental Disabilities, Review of Educational Research, School Psychology Review*).

Afterwards, the titles, abstracts and in some cases the full texts of the retrieved studies were reviewed to identify additional studies which met all inclusion criteria. In the next step, an ancestral search was conducted in the reference lists of eligible studies to identify other additional relevant studies for inclusion in our review. After that, we consulted citation indexes in order to see which studies cited the studies already retrieved in the previous search steps. Finally, the authors whose names occurred frequently in the eligible studies were contacted, and we asked them whether they knew any other eligible studies.

1.2. Inclusion/exclusion criteria and data extraction

We searched for published and unpublished studies between 1985 to 2015, written in English, which met the following inclusion criteria: (1) Only studies with human participants were included. (2) The study should be a quantitative review of at least two SSED studies (possibly in addition to group studies). Any kind of commentaries, reviews of qualitative case studies, reviews of group studies (including randomized controlled trials), conceptual and methodological studies without an empirical illustration were excluded. (3) The study had to report a numerical summary of the effect over all included SSED studies. From the reviews of both SSED and group studies, only those were included that did separate analyses on the SSED data. (4) The study had to be a journal article, conference article, book, book chapter, dissertation, or report. Abstracts and slide presentations were excluded. We retrieved 178 studies that met our inclusion criteria.

In order to extract the required data, 22 items (some of which contained sub-items) related to general characteristics and study designs, the kind of data, and the kind of analysis were examined in each of the included studies. In terms of the study characteristics and study designs, we examined some characteristics such as the type of publication, the number of SSED studies, the reporting the length of phases, the number of participants in SSEDs, the study design of primary studies, and the type of synthesized studies. The next results section gives an overview of some characteristics related to the kind of data provided in primary SSED studies and reviews, namely the type(s) of the outcome scale for primary studies, the existence of multiple effects per study, the reasons for multiple effects, the handling of dependency, the kind of data reported in primary studies, and the kind of used data from primary studies in the review or synthesis. Finally, some characteristics related to the kind of analysis are examined: the analyses done in primary studies, the reported results in the review, the moderator variables, the analyses done in the review to synthesize the included SSED studies, the reporting of the reasons for conducting particular analysis, whether the possibility of linear / nonlinear time trends in the data (within studies) is taken into account, whether the possibility of autocorrelation, checking statistical assumptions is taken into account, and whether sensitivity analysis has been conducted. The following sections outlines the most important findings of our present review.

Twenty-one percent (38 out of 178 studies) of the included studies were randomly selected for data extraction by two independent coders. In total, the studies were rated by five coders. Interrater agreement was calculated for each study variable by dividing the number of agreements by the number of agreements plus disagreements. Disagreements were solved by discussion between coders. If no consensus was reached, the last author decided on the code given. The final agreement rates ranging from 96% to 100% for all extracted items with an average of 98.5 %. In the cases of discrepancies, the last author checked the studies and made a decision.

2. Results

In an attempt to identify what kind of study designs, what kind of data, and what kind of analyses were done in SSED meta-analyses, systematic reviews, and syntheses during the last three decades, we reviewed the 178 included studies in details.

2.1. Study characteristics and designs

As shown in Table 1, large majority of the included studies (*n*=166) has been published in peer-reviewed journals. The findings reveal that the total number of included primary studies was reported in the majority of the reviews (*n*=173). The results show that less than 20 percent of the reviews (*n*=31) reported the number of measurement occasions, and the sessions or duration of baseline and intervention phases for each or some of the primary studies or over all primary studies were provided in almost 30 percent of studies (*n*=56).

According to acquired results, 65 percent of the reviews (*n*=115) reported the total number of the participants over all included studies and almost half of the studies (*n*=87) provided the number of the participants of each primary study individually.

In terms of the kind of study design of the included primary studies, nearly one third of the reviews (*n*=55) did not provide any information about the kind of study design of aggregated primary studies and the rest of the studies reported a variety of different designs. Most of the included reviews (*n*=126) synthesized only SSED studies and the other reviews were combinations of SSED and group studies and reported the results of each kind of these studies separately.

Table 1 - General characteristics and Study design

Characteristics	**N**	**%**
Type of publication		
Journal article	166	93
Book chapter	4	2
Dissertation	7	4
Project report	1	1
Number of SSED studies		
Reported number of studies in review	173	97
Reported number of studies in quantitative synthesis	171	96
Reported the length of phases		
Reported only number of measurement occasions	16	9
Reported only duration of phases	40	22.5
Reported both number of measurement occasions and duration	15	8
Not reported	107	60
Number of participants in SSED		
Reported only total number of participants in all included studies	60	34
Reported only number of participants per primary study	32	18
Reported both total number of participants in all included studies and number of participants per primary study	55	31
Not reported	31	17
Study design of primary studies (multiple are possible)		
PCR	99	56
PCNoR	49	28
MB	118	66
AT	72	40
CC	12	7
Combined design	43	24
Not reported	55	31
Type of synthesized studies		
Only SSED studies	126	71
Combination of SSEDs and group designs	52	29

Note: **PCR = phase change with reversal designs; PCNoR = phase change without reversal designs; MB = multiple-baseline designs and multi-element designs; AT = alternating treatment designs; CC = changing criterion designs (see Shadish & Sullivan, 2008, for more information).**

2.2. Kind of data

In Table 2, some information related to kind of data in primary studies and meta-analyses is provided. Most of the reviews (*n*=162) did not provide the type of outcome scale of primary studies. Almost ten percent of the reviews reported percentages, count data, interval scale or a combination of these for the outcomes in synthesized data. A considerable number of reviews (*n*=168) had multiple effects per primary study and only 10 studies did not explicitly state whether there were multiple effects in synthesized studies. There were several reasons why primary studies reported multiple effects and some studies reported multiple reasons for existence of multiple effects. The inclusion of multiple participants was the most common reason for the presence of multiple effects in most of the studies (153 out of 168 studies) and approximately half of the studies mentioned multiple outcomes as the main reason.

Out of the 168 studies with multiple effects, 111 reviews explicitly stated that they took into account the dependency resulting from multiple effects. In total, 113 studies reported taking into account dependency, but two of them did not report explicitly whether there were any multiple effects in primary studies. A closer look at the data indicates that from these studies (n=113) that took into account the dependency, averaging the effect sizes per primary study was applied in 71 studies (65%); 34 studies (30%) did separate meta-analyses for each outcome, treatment or setting, and 19 studies (17%) used multilevel analysis for considering dependency. Some of the studies used the combination of different methods for synthesizing dependent data. Only five reviews used other procedures such as generalized least squares regression, calculating the median of effects per study, or randomly selecting one effect size for each included study.

The techniques used for handling dependent data differed for the various types of dependencies observed in the reviewed studies. In the case of multiple participants per study, the most frequently applied method for handling this kind of dependency was averaging effect sizes per study. From the 152 review studies with multiple participants per primary study, 61 studies averaged the effect sizes for the participants of each study, 19 studies applied multilevel analysis, and the rest of the studies did not report any procedure for taking into account the dependency. In total, from the 75 studies that reported multiple outcomes, 58 took into account the dependency. Taking a simple mean of effect sizes was the most commonly described method for handling the multiple outcome dependence in 33 of the studies. Almost half of the studies which considered the dependency of multiple outcomes summarized the results separately for each of the outcomes (*n*=28) and a small proportion of them applied multilevel analysis (*n*=8). Most of the studies with multiple treatments (*n*=15) employed a similar set of techniques to handle the multiple treatments dependency. Two thirds of these studies averaged the effect sizes of different treatments per study. The rest of the studies did separate meta-analyses or a multilevel analysis. In the studies with multiple time points, the most frequently used methods for taking into account dependency were the use of a simple mean per each study (11 out of 17 studies) or the use of separate meta-analysis for different time points (nine out of 17 studies).

Over 60 percent of the studies stated that they retrieved the data from the graphs in the primary SSED studies. A notable amount of included reviews reported that they used effect sizes from primary studies for synthesizing.

Table 2 - Kind of data

Characteristics	n	%

Type(s) of outcome scale for primary studies		
Percentages	11	6
Count	11	6
Interval scale	4	2
Not reported	162	91
Existence of multiple effects per study		
Yes	168	94
Not reported	10	6
Reasons for multiple effects		
Multiple participants	152	90
Multiple outcomes	75	45
Multiple treatments	19	11
Multiple time points	17	10
Multiple settings	3	2
Not reported	5	3
Taking into account dependency		
Yes	111	66
No	2	1
Not reported	55	33
Kind of data reported in primary studies		
Graphs	122	69
Raw data	9	5
Effect sizes	0	0
Other summary statistics	2	1
Not reported	54	30
Kind of used data from primary studies in the review or synthesis		
Raw data	22	12
Effect sizes	161	90
***p*-values**	0	0
Standardized	35	20
Not reported	2	1

2.3. Kind of analysis

A considerable number of reviews did not report which kind of analyses were done in the primary studies (Table 3). The results indicate that the majority of the reviewed studies provided the overall effect size such as a mean or a median across all primary studies. Many studies (*n*=139) just discussed whether the overall effect size was significant or not, but less than twenty percent of the studies mentioned the statistical significance of these overall effect sizes. A few studies (less than fifteen percent) examined the heterogeneity of effect sizes and the significance of this test. A moderator analysis has been done in many studies (*n*=130) and the significance of moderator variables has been provided in most of them. The findings illustrate that the majority of the studies which did a moderator analysis considered intervention characteristics and participant characteristics as moderators.

In terms of synthesizing the effect sizes of primary SSED studies, the most frequently employed approach was the calculation of a simple average of effect size. Around half of the reviews provided the range of the effect sizes among all included studies. Across all of the studies, only 22 studies applied a multilevel analysis for aggregating the SSED studies. In other studies, the calculation of a median of effect sizes, the calculation of a weighted average of effect sizes, or a regression analysis of effect sizes were reported as the kind of analysis. Some of the studies did multiple analyses in their reviews. Three-quarter of the studies (*n*= 133) explained their choice for a specific kind of analysis.

A few studies took the possibility of linear / nonlinear time trends in the data (within studies) into account. Among them, twelve studies included time trends in the model, seven studies used Tau-U for controlling time trend, and three studies applied de-trending by Allison-MT method. The results indicate that some of the studies (n=19) took into account the possibility of autocorrelation. Eight of these studies (42%) used multilevel analysis, from which one modeled Lag 1 autocorrelation. Four studies (21%) that did not use multilevel analysis explicitly mentioned they modeled Lag 1 autocorrelation.

In 19 reviews, statistical assumptions were checked. Most of them (n=12) checked for normality and outliers, five studies checked for homogeneity of variances of residuals, and two other studies checked assumptions for parametric analyses of mediating/moderating variables and checked whether they had a minimum of 10 data points per phase to obtain reliable results. Some of the studies (n=21) did sensitivity analyses. In this respect, 11 studies compared the results with and without outliers and four studies compared the results of overall effect size including all studies with results of overall effect size leaving out low-quality studies.

Table 3 - Kind of analysis

Characteristics	N	%
Analyses done in primary studies		
Regression analysis	2	1
Visual analysis	3	2
Calculating effect sizes	3	2
Descriptive analysis	4	2
Not reported	170	96
Reported results in review		
Overall effect size	170	96
Range of effect sizes	83	47
Significance of overall effect	139	78
Statistical significance of overall effect	32	18
Confidence interval for overall effect	56	32
Examining heterogeneity of effect	26	15
Exact value of heterogeneity	17	10
Examining significance of heterogeneity	25	14
Exact value for significance of heterogeneity	20	11
Moderator analysis	130	73
Exact value for moderator analysis	113	63
Significance of moderator analysis	119	67
Exact value of the significance of moderators	78	44
Reported other results	24	14
Moderator variables		
Participant characteristics	91	70
Intervention characteristics	107	82
Setting characteristics	58	45
Design characteristics	39	31
Study characteristics	19	15
Publication type	14	11
Outcome effect	56	43
Publication year	5	4
Place of publication	1	1
Analysis done in the review to synthesize the included SSED studies		
Simple average of effect sizes	116	65
Median of effect sizes	31	17
Weighted average of effect sizes	34	19
Range of effect sizes	83	47
Regression analysis of effect sizes	7	4
Multilevel analysis	22	12
Explanation given for conducting this analysis	133	75
Taking into account the possibility of linear / nonlinear time trends in the data (within studies)	21	12

Taking into account the possibility of autocorrelation	19	11
Checking statistical assumptions	19	11
Conducting sensitivity analyses	21	12

A significant portion of the reviewed studies (*n*=161) used effect sizes for synthesizing the results of SSED studies. More than half of the studies calculated percentage of nonoverlapping data (PND), some studies calculated the standardized mean difference (SMD), and the rest of the studies calculated effect size metrics.

We further explored the relation between the used effect size metric and the kind of analyses. In general, as presented in Table 4, a considerable amount of the studies which calculated effect sizes for synthesizing the results of SSED studies reported the simple average of effect sizes as an overall effect. Reviews that calculated PND, PZD, PEM, IRD, NAP, and Tau-U as effect sizes, mostly used a simple average of the effect sizes and the range of effect sizes as the kind of analysis for synthesizing. The results also indicate that studies that used SMD or other kinds of effect sizes frequently reported the simple average and the weighted average of the effect sizes. Some reviews reported multiple analyses for the overall effect size such as simple average, median, the range of effect sizes, etc.

Multilevel analysis has been applied for synthesizing the SSED studies by only a small portion of the studies (*n*=22). Most of these studies (*n*=13) did a three-level analysis, mostly (*n*=12) with measurement occasions, participants, and studies as the units at the three levels; only one study employed measurement occasions, effect sizes, and participants as units at the three levels. Seven other studies did a two-level analysis (three reviews considered measurement occasions and participants as levels, three studies used participants and studies as levels, and one study considered effect sizes and studies as two levels). Of these 22 studies which applied multilevel analysis, 19 studies did a moderator analysis as well. Out of these studies, 17 reviews reported using raw data for synthesizing, five studies used standardized data points, and one study did not report which kind of data they used for combining the results from SSED studies.

Other kinds of analysis have been employed in a small portion of the studies. In total, seven studies did a regression analysis of effect sizes. The remaining studies (*n*=12) reported analyses such as identification of outliers, visual analysis, Bonferroni correction to control the Type I error rate or a calculation of the quartiles of the distribution of effect sizes.

Table 4 - The effect size metrics and kind of analyses used for synthesizing

Effect size metric	Frequency (Percentage) (k=161)	Kind of analysis used for synthesizing			
		simple average of effect sizes	median of effect sizes	weighted average of effect sizes	range of effect sizes
PND	99(61)	76(77)	25(25)	7(7)	52(53)
SMD	32(20)	23(72)	3(9)	13(41)	10(31)
IRD	26(16)	17(65)	1(1)	7(27)	11(42)
NAP	11(7)	10(91)	2(18)	1(9)	4(36)
PZD	10(6)	8(80)	-	2(20)	5(50)
PAND	15(9)	7(47)	6(40)	5(33)	4(27)
PEM	11(7)	10(91)	-	1(7)	4(36)
Phi	9(5.5)	5(56)	4(44)	3(33)	2(22)
Tau-U	10(6)	5(50)	-	3(30)	4(40)
Other effects	15(9)	10(67)	1(7)	7(47)	6(40)

*** The numbers mentioned in the table are frequencies (Percentages)**

We looked closer at the trend of using different metrics and methods for synthesizing the results of SSED studies and providing a quantitative overview of overall effect size during the last three decades. The findings indicate that before 2000, only 16 studies met our inclusion criteria and in this period PND, SMD, and PZD were the most frequently used metrics for calculating effect sizes. Between 2001 and 2005, there were 12 reviews that calculated different nonregression-based metrics such as PND, SMD, PZD and other kinds of effect sizes. Almost one-third (*n*=56) of the included reviews done in 2006 to 2010 used different effect size metrics. More than half of the studies (*n*=94) done in 2011 to 2015 used nonregression-based metrics as well for calculating the effect sizes. Over all periods, the PND and SMD were the most commonly used metrics in primary SSEDs for combining the results. The usage of other kinds of metrics such as IRD, NAP, PAND and Tau-U increased in 2011-2015 (Table 5).

Table 5 - The evolution in the use of effect sizes for synthesizing SSED data

Time range	Total studies	Using effect sizes for synthesizing										
		Total	PND	SMD	IRD	NAP	PZD	PAND	PEM	Phi	Tau-U	Other ES
1985-2000	16	14(87)	11(85)	3(15)	-	-	1(8)	-	-	-	-	-
2001-2005	12	12(100)	8(67)	3(25)	-	-	3(2)	-	-	-	-	2(13)
2006-2010	56	53(95)	44(83)	9(17)	1(2)	1(2)	5(9)	2(4)	6(11)	2(4)	-	3(6)
2011-2015	94	82(87)	36(43)	17(22)	25(30)	10(12)	1(1)	13(16)	5(6)	7(8)	10(12)	10(12)
Total	178	161(90)	99(61)	32(20)	26(16)	11(7)	10(6)	15(9)	11(7)	9(6)	10(6)	15(9)

*** The numbers mentioned in the table are frequencies (Percentages)**

As indicated in Table 6, during the last three decades, some of the studies employed regression-based methods for synthesizing the results of SSED studies. Before 2000, only three out of 16 studies applied regression analyses for combining effect sizes. Between 2001 and 2005, there were twelve reviews of which only one study used multilevel analysis. During 2006 to 2010 only one study used regression analysis of effect sizes and two studies applied multilevel analysis. From more than half of the reviews that are done from 2011 to 2015, only three studies employed regression analysis of effect sizes and in this period more studies (*n*=19) applied multilevel analysis for synthesizing the results of SSED studies. In total, only 12 percent of studies applied multilevel analysis and only four percent of studies applied regression analysis of effect sizes.

Table 6- The evolution of other kind of analyses for synthesizing SSED data

Time range	Total studies	Regression analysis of effect sizes	Multilevel analysis
1985-2000	16	3(20)	-
2001-2005	12	-	1(8)
2006-2010	56	1(2)	2(4)
2011-2015	94	3(3)	19(20)
Total	178	7(4)	22(12)

*** The numbers mentioned in the table are frequencies (Percentages)**

3. Conclusion

Meta-analysis is one of the quantitative statistical methods for aggregating the findings of a body of researches with the same objectives and can provide larger statistical power and a more precise overall intervention effect for evidence-based research and practice. For this purpose, it is crucial that meta-analysts and reviewers try to present more accurate results by applying the most appropriate procedures and techniques based on the various characteristics of the SSED studies in their syntheses. Selecting the appropriate method for providing the overall effect in aggregation of SSEDs is still a crucial issue for meta-analysts and reviewers. Various quantitative approaches and methods for synthesizing these experimental designs have been proposed by different researchers, but superiority and appropriateness of these methods are still being debated and no consensus exists.

Because of the growing number of SSED meta-analyses and the advances in (meta-)analytical approaches and methods, our findings of a systematic review of the literature between 1985 and 2015 indicate a large variety in procedures that were applied and illustrates the lack of consensus about the most appropriate methods, both for expressing the size of the effect in primary studies and for synthesizing these over studies. Despite the methodological advancements in SSED (meta-)analysis, a considerable number of meta-analyses still apply rather basic methods. Some changes have taken place during three last decades in employing various effect size measures and applying different techniques for combining the results across primary studies, although the evolution in the use of advanced methods is rather slow. It can be useful for meta-analysts to become aware of different approaches which have been applied in their field of interest and of the methodological and statistical advancements which have been occurred in the SSED meta-analysis. This will leave them capable of selecting the most appropriate methods and procedures for calculating the overall effect across a bunch of SSED studies and of providing more reliable and valid results. Because every particular metric has its unique strengths and weaknesses, the best way to select the most adequate (meta-)analysis method is to formulate plausible arguments based on present data characteristics. When meta-analysts and systematic reviewers have sufficient knowledge about different effect size indices and synthesis approaches and procedures for meta-analyzing SSEDs and when they are well aware of all data features of their primary SSED studies, they can provide stronger evidence across studies.

We propose some recommendations to future SSED meta-analysts and reviewers, based on our findings and the existing limitations: (1) consider more advanced meta-analytical approaches such as regression-based methods or multilevel modeling instead of just simple analytical approaches in order to take into account some critical characteristics of single-subject data such as dependency, autocorrelation and time trends, and (2) report more descriptive details and characteristics of primary SSED studies such as the number of participants, the study design, the number of measurement occasions, the duration of different phases, the analyses done in the study, the kind of outcome scale, etc.

Further research is needed to examine the justifications, considerations, assumptions, procedures, advantages and disadvantages, and strengths and weaknesses of methods applied in SSED meta-analyses.

References

Beretvas, S. N., & Chung, H. (2008). A review of meta-analyses of single-subject experimental designs: Methodological issues and practice. *Evidence-Based Communication Assessment and Intervention*, *2*(3), 129–141. http://doi.org/10.1080/17489530802446302

Farmer, J. L., Owens, C. M., Ferron, J. M., & Allsopp, D. H. (2010, April). *A methodological review of single-case meta-analyses*. Paper presented at the American Educational Research Association. Denver, CO.

Maggin, D. M., O'Keeffe, B. V, & Johnson, A. H. (2011). A quantitative synthesis of methodology in the meta-analysis of single-subject research for students with disabilities: 1985–2009. *Exceptionality: A Special Education Journal, 19*(2), 109–135. http://doi.org/10.1080/09362835.2011.565725

Owens, C. M. (2011). *Meta-analysis of single-case data: A monte carlo investigation of a three level model*. Graduate Thesis and Dissertation. http://scholarcommons.usf.edu/etd/3273

Petit-Bois, M., Baek, E. K., Van den Noortgate, W., Beretvas, S. N., & Ferron, J. M. (2016). The consequences of modeling autocorrelation when synthesizing single-case studies using a three-level model. *Behavior Research Methods*, *48*(2), 803–812. http://doi.org/10.3758/s13428-015-0612-1

Riley-Tillman, T. Ch., & Burns, M. K. (2009). *Evaluating educational interventions: single-case design for measuring response to intervention*. Canada: The Guilford Press.

Schlosser, R. W., Lee, D. L., & Wendt, O. (2008). Application of the percentage of non-overlapping data (PND) in systematic reviews and meta-analyses: A systematic review of reporting characteristics. *Evidence-Based Communication Assessment & Intervention*, *2*(3), 163–187. http://doi.org/10.1080/17489530802505412

Shadish, W. R. (2014a). Analysis and meta-analysis of single-case designs: An introduction. *Journal of School Psychology*, *52*(2), 109–122. http://doi.org/10.1016/j.jsp.2013.11.009

Shadish, W. R. (2014b). Statistical analyses of single-case designs: The shape of things to come. *Current Directions in Psychological Science*, *23*(2), 139–146. http://doi.org/10.1177/0963721414524773

Shadish, W. R., Hedges, L. V., & Pustejovsky, J. E. (2014). Analysis and meta-analysis of single-case designs with a standardized mean difference statistic: A primer and applications. *Journal of School Psychology*, *52*(2), 123–147. http://doi.org/10.1016/j.jsp.2013.11.005

Shadish, W. R., & Rindskopf, D. M. (2007). Methods for evidence-based practice: Quantitative synthesis of single-subject designs. *New Direction for Evaluation*, *113*, 95–109. http://doi.org/10.1002/ev.217

Smith, J. D. (2012). Single-case experimental designs: A systematic review of published research and current standards. *Psychological Methods*, *17*(4), 510–550. http://doi.org/10.1037/a0029312.Single-Case

Tincani, M., & De Mers, M. (2016). Meta-analysis of single-case research design studies on instructional pacing. *Behavior Modification*, *40*(6), 799–824. http://doi.org/10.1177/0145445516643488

Van den Noortgate, W., & Onghena, P. (2003). Combining single-case experimental data using hierarchical linear models. *School Psychology Quarterly*, *18*(3), 325–346. http://doi.org/10.1521/scpq.18.3.325.22577