INTRODUCTION

Meta-Analysis of Single-Case Design Research: Introduction to the Special Issue

Matthew K. Burns

Abstract   Single-case design (SCD) research focuses on finding powerful effects, but the influence of this methodology on the evidence-based practice (EBP) movement is questionable. Meta-analytic procedures may help facilitate the role of SCD research in the EBP movement, but meta-analyses of SCDs are controversial. The current article provides an introduction to the special issue on meta-analyses of SCD research by discussing concerns regarding the internal and external validity of these designs. Specific considerations for increasing the validity of SCD meta-analyses are provided, as are brief overviews of the articles included in the special issue.

Keywords   Meta-analysis · Effect size · Evidence-based practice

Introduction

What effect has the evidence-based practice (EBP) movement had on actual practice in education? There seems to be an increased interest in research among practitioners, but the effects of the EBP movement remain mostly unknown (Cooper et al. 2009). The American Psychological Association (APA Presidential Task Force on Evidence-Based Practice 2006; APA Task Force on Evidence-Based Practice for Children and Adolescents 2008) and the field of special education (Cook et al. 2009) have clearly embraced the move toward stronger evidence for common practices. However, a recent survey of school psychologists and special education teachers found that the research base for some commonly implemented

M. K. Burns (✉)
University of Minnesota, 341 Education Science Building,
56 E. River Road, Minneapolis, MN 55455, USA
e-mail: burns258@umn.edu

practices was either questionable or clearly indicated the practice to be ineffective (e.g., modality instruction) (Burns and Ysseldyke 2009).

EBP is predicated on dissemination of intervention research, but the process of moving from research to practice remains ambiguous (Lochman 2003) at least partly due to the voluminous amount of research that is often difficult to interpret, disconnected, and occasionally contradictory (Wandersman et al. 2008). Hattie (2009) indicated that in order to close the research-to-practice gap in education, practitioners and policy makers need to summarize and compare all of the diverse types of evidence, which is likely best accomplished through meta-analytic research (Kavale and Forness 2000).

Meta-analysis is the statistical synthesis of results from a systematic review of original research related to a particular topic (Borenstein et al. 2009). Research *syntheses* focus on integrating empirical research in order to find generalizations in the data, which is different from research *reviews*, which focus on evaluating research (Cooper and Hedges 2009). Thus, the validity of conclusions drawn from meta-analytic research is only as strong as the methods used within each original study. However, meta-analytic research will likely be an important contributor to policy debates and the EBP movement because it allows for objective and comprehensive comparisons of the effectiveness of various practices (Cordray and Morphy 2009).

Applied behavior analysis (ABA) and EBP are nearly synonymous because ABA "is a discipline deliberately turning away from the detection of weak variables: it systematically filters from its discovery methods the ability to discover variables of less than powerful effect" (Baer 1977, p. 117). Thus, applied behavior analysts have been providing decades of research on effective strategies for promoting student learning. However, William Heward, the former President of the Association for Behavior Analysis International, reported that the field of ABA has been "at best, a bit player in our country's efforts to reform education" (Heward 2008, p. 1). Although there are many potential reasons for ABA's limited role in reforming educational practice (e.g., assumptions and beliefs that are different from most educators), the lack of systematic methods to synthesize research should be considered. A literature search with PsycINFO found 3,077 articles published in the *Journal of Behavioral Education* and the *Journal of Applied Behavior Analysis*, but combining the search results with the term "meta-analysis" identified only 2 articles. Certainly, behavior analysts can conduct narrative reviews of previous research, but narrative reviews tend to be susceptible to several errors inherent in the review process such as omitting important studies, misrepresenting conclusions, and treating all evidence within the synthesis as equal (Dunkin 1996); and it was in response to these criticisms that meta-analytic methods were first proposed (Glass 1976).

One potential difficulty with conducting meta-analyses of behavioral research is the prevalence of single-case design (SCD) studies, which are well suited for determining the effectiveness of academic interventions, but are controversial bases for meta-analyses. The strength of SCD studies is the strong internal validity inherent in the experimental designs, and meta-analyzing the data essentially turns

them into a group design with weaker causal claims (Baron and Derenne 2000). Moreover, meta-analyses rely on empirical estimates of effect, and doing so with SCD data is not likely to capture patterns in the data across time, could miss idiosyncracies in the data, and is overly affected by atypical baseline data (Salzberg et al. 1987; White 1987).

The exclusion of an entire class of experimental designs from meta-analyses is "dismaying, especially because potentially weaker designs, such as nonequivalent comparison group designs with just one pretest and posttest measure, are often included" (Shadish et al. 2008, p. 188). Moreover, it is necessary for SCD researchers to embrace meta-analytic approaches in order to fully join the EBP movement (Shadish et al. 2008). Myles et al. (1996) discussed procedures to conduct a meta-analysis of SCD studies over 15 years ago, but the prevalence of these studies has only recently increased, and we have learned much about synthesizing SCD data since then. Therefore, additional research and discussion are needed regarding the meta-analysis of SCD studies, which is the goal of this special issue. Below I will begin the special issue by briefly addressing the potential criticisms of meta-analyzing SCD data, such as decreased casual claims, inability to adequately describe the effect with one empirical estimate, and the priority on causality within SCDs by framing each within internal and external validity, and will provide an introduction to the special issue.

Internal Validity

The What Works Clearinghouse has recently accepted SCDs as adequate evidence of causal validity and provided standards to evaluate them (Kratochwill et al. 2010), which was a substantial step forward for SCD researchers and the EBP movement. Meta-analyses are essentially a survey of research reports (Lipsey and Wilson 2001) and should be interpreted and presented as such. Thus, behavior analysts should interpret conclusions within meta-analytic research as descriptive and should not automatically assume they are internally valid. However, the rigor of the experimental design can be coded and/or used as an inclusion criterion to enhance the confidence in findings of meta-analytic research.

A second consideration regarding the internal validity of meta-analyses of SCD studies is the use of an effect size metric to interpret the data, which is complicated by the lack of an accepted effect size for SCD data. Cohen's d is not appropriate for SCD research because it examines differences between groups rather than within subjects and because the data do not meet basic assumptions needed to conduct parametric analyses. Two of the studies (Burns, Zaslofsky, Kanive, and Parker; Petersen-Brown, Karich, and Symons) and one commentary (Parker and Vannest) in the special issue directly address effect sizes. Thus, I will not discuss effect sizes at length here, but will point out the need for additional research. Different metrics such as a no-assumptions effect size (NAES; Busk and Serlin 1992) and percentage of nonoverlapping data (PND, Scruggs et al. 1987) are frequently used, but they have difficulties related to the autocorrelation of data for NAES, and an inability to

compute confidence intervals and the potential influence of outlying data for PND (Riley-Tillman and Burns 2009).

Given the difficulties with PND and NAES, Parker and colleagues have recently provided multiple interesting options for effect sizes such as percentage of all nonoverlapping data (PAND; Parker et al. 2007) and nonoverlap of all pairs (NAP; Parker and Vannest 2009). Burns et al. (this issue) provide a discussion of these approaches, and Parker and Vannest (this issue) discuss the rationale for them. Although these are intriguing options for meta-analyzing data, effect sizes will never replace visual analyses, and reporting them for individual studies may ultimately do little to help us better understand the data. Providing numeric estimates of effect merely serves to help synthesize the data across studies.

Although the purpose of meta-analyses is to describe conditions under which the effects are larger, there are steps that meta-analytic researchers can take to strengthen the internal validity of their claims. Cooper (2010) provides an excellent summary of threats to internal validity that are inherent in meta-analytic research and suggests methods to overcome them. These descriptions and recommendations are included in Table 1, but are then extended to meta-analyses of SCD research. It should be noted that not every threat to validity is included in Cooper's recommendations and not every one of Cooper's recommendations is included in the table. However, the recommendations that are most salient to meta-analyses of SCD studies are included and extended and are consistent with suggestions made by Baron and Derenne (2000; that is, carefully consider the rules for including a study within any specific category and frankly recognize that the procedure has many of the characteristics of a between-group design).

External Validity

Data from a SCD study experimentally test a conceptual theory or intervention effectiveness for the given participants, but may have limited relevance for participants, settings, or behaviors other than those included in the study (Horner et al. 2005), which is an acceptable trade-off if the field is to focus on finding powerful effects (Baer 1977). SCD researchers may be more interested in internal validity than external validity, but generalizability and feasibility are essential dimensions of EBP in psychology (Levant et al. 2006). Thus, to move the EBP movement forward in education and psychology, SCD researchers should carefully consider external validity and the implications for SCD research.

Consumers of research often associate traditional randomized clinical trials with stronger external validity because the sample is representative of a larger population and the design includes multiple settings, but the only way to be confident in the external validity of between-group designs is to systematically replicate them (Berliner 2002). A SCD researcher could argue that future between-group research of the effect first noted in a SCD study with strong internal validity would be an acceptable approach to address external validity, but the generalizability of randomized between-group trials is often unclear (Rothwell 2005). Moreover,

Table 1 Threats to and protections of validity in meta-analyses based on Cooper (2010) and application to single-case design (SCD) meta-analyses

Threat	Description	Protecting validity	Additional considerations for SCD
Poorly defined constructs	Defining constructs in a manner that suggests greater generality than appropriate	Use broadly conceptualized definitions Reevaluate definitions as variables are identified Identify differences in study results that are attributed to study characteristics	Code as many aspects of the setting and participants as possible Code by study design and examine differences in effect
Publication bias	Studies found in literature search may be fundamentally different than the actual population of studies	Conduct an exhaustive search with broad search terms Hand search relevant journals, previous meta-analyses, and reference lists of selected articles Contact primary authors of selected articles Include indices of retrieval bias (e.g., Failsafe N)	Seek out studies that do not demonstrate an effect
Unreliability of coding	Coders are not consistent between coders or over time	Train coders to a high level of agreement before beginning Conduct interrater reliability for inclusion criteria and coding Discuss codes that lead to disagreement	Follow standards and procedures for interrater agreement associated with SCD research
Causal relationships incorrectly inferred	Selected studies were not designed to show causality	Use a justifiable weighting system Use a priori criteria to determine conclusion rather than results Exhaust design characteristics when coding articles	Include experimental design and quality of baseline data as inclusion criteria Code AB designs and compare effects to experimental designs Compute single effect size for multiple-baseline studies
Inappropriate summarizing of data	Nonindependent effects are treated as independent, incorrectly or not weighting effect sizes, and conducting homogeneity tests without sufficient power	Explicitly state assumptions that guide analyses Evidence for validity of interpretation rules should be presented if appropriate	Weight effect sizes by number of data points or number of participants Carefully consider if data presented within one study are independent or are assessments of the same construct

Table 1 continued

Threat	Description	Protecting validity	Additional considerations for SCD
Misinterpreting cumulative effects	Missing data leading to approximated effect sizes, restricted range in the data, misidentifying or missing a moderator variable, lack of statistical power, and overgeneralizing to persons, settings, treatments, outcomes, and times not contained in the studies	Explicitly state how incomplete reports were handled Use multiple effect sizes and statistical procedures with different assumptions Summarize the sample characteristics of individuals in the studies	Describe study participants in the detail that is available for SCD studies Include a nonparametric estimate of effect that can be computed with visual analysis of the data and does not require actual scores Compute multiple estimates of effect such as percent of all nonoverlapping data, nonoverlap of all pairs, and phi
Presenting the results	Synthesis procedures are missing or not described well enough to evaluate threats	Follow the Meta-Analysis Reporting Standards (APA Publications and Communications Board Working Group on Journal Article Reporting Standards 2008)	Present data as descriptions of conditions under which interventions are more effective Thoroughly discuss topics for future research

Campbell and Stanley (1963) stated that generalization is never fully justified and even a well-designed between-group study may have questionable external validity.

Issues regarding external validity may be one reason to consider conducting meta-analytic research because of the possibility to examine potential moderator variables, or the "classes of treatments, outcomes, settings, populations, or times across which the magnitude or direction of a causal effect differs" (Matt and Cook 2009, p. 552). It is often not possible to experimentally manipulate the setting, conditions, or time of day, week, or year in SCD research. Barlow and Hayes (1979) suggested that external validity within SCD research should be addressed by predicting which factors will affect generalizability and should systematically replicate the research to address those factors. However, the number of factors that could affect the effectiveness of a particular behavior or academic intervention is often too numerous to list and there may not be enough systematic replications of any one procedure (e.g., with different populations, in different settings, etc.) to examine levels of any given moderator variable.

Meta-analytic researchers can code the factors that could affect intervention effectiveness into broad categories and evaluate their moderating influence, which can then be tested as needed with additional research. As stated above, generalizability and feasibility are essential dimensions of EBP, and infusing

meta-analytic designs within the SCD research literature could help forward the role of SCD research within the EBP movement.

The Special Issue

The special issue includes six articles; two that empirically examine effect sizes for SCD meta-analyses, two that are examples of meta-analyses with SCD studies, and two commentaries. Petersen-Brown and colleagues extended previous research by examining NAP with multiple baseline design studies using visual analysis as the criterion. Although using visual analyses as the criterion is not novel, the application of receiver operating characteristics to determine a large effect is quite novel and suggested that a NAP of almost 1.00 was required to determine a large effect.

Burns et al. used PAND and NAP to compute phi in order to compare the effect sizes derived from SCD research and those from between-group studies with incremental rehearsal (IR). The data suggested that IR was an effective intervention for several conditions, but the question of efficiency was not adequately addressed. The novel contribution of this paper is the comparison of effect sizes from the two different designs, which did not result in a statistically significant difference. Thus, the phi coefficient may have promise in linking the two literatures either directly or by comparison, but additional research is needed. Moreover, the information presented on the comparison of PAND, NAP, and phi may help behavior analysts better interpret those metrics.

Perry, Albeg, and Tung, and Methe, Kilgus, Neiman, and Riley-Tillman present two examples of meta-analyses of SCDs. The former presented an interesting meta-analysis of self-regulation interventions for academic deficits. The paper relied on PAND to synthesize the SCD data and found that the interventions were effective for several different subject areas, disabilities, age groups, etc. However, it was noted that self-regulatory interventions were less effective for older elementary-aged children, which contradicts previous assumptions. Moreover, the meta-analysis pointed out the almost complete lack of research with students in middle school and high school. Finally, self-monitoring alone was not as effective as combining self-monitoring with strategy instruction, which is a potentially important finding with obvious implications for practice.

The latter example of a meta-analysis (Methe et al.) studied math computation interventions and used both PAND and the improvement rate difference (IRD). Several potential moderator variables were identified that suggested some conditions for effectiveness that practitioners could consider. However, the novel contribution of this study is the comparison of the size of the effect and the rating of experimental control that resulted in an inverse relationship. The decrease in the effect size as experimental control increased was a potentially important finding that the authors argue suggested a need for continued experimental rigor in the literature. Future meta-analysts might consider conducting a similar analysis to evaluate the influence that experimental control, and factors that influence it, has on the effect sizes.

Parker and Vannest provide the first commentary, which discusses the role of bottom–up approaches to synthesizing SCD data. A bottom–up approach relies on a

visually guided selection of phase contrasts to compute one estimate of effect for the entire data. A top–down approach relies on analyses of the continuous data as is conducted with hierarchical or multi-level modeling, or with ordinary least squares. The authors carefully point out that both approaches have merit, but make a convincing argument for bottom–up approaches. The major appeal of bottom–up approaches that is most relevant to this special issue is their consistency with the tradition of visual analysis within SCD research and behavior analysis. Although the other top–down approaches are interesting, the authors make a compelling case that bottom–up is preferable for meta-analyses of SCD research.

Finally, Horner and Kratochwill provide a commentary that moves the EBP debate beyond the size of the effect. The special issue focused largely on examining effect sizes for meta-analyses of SCD research, which adds to the literature on this important line of inquiry. However, there is more to being designated as an EBP besides a large median effect. Horner and Kratochwill propose the 5–3–20 rule in which there are at least 5 SCD studies with documented experimental control, conducted by at least 3 different research groups, with at least 20 total participants. The 5–3–20 rule is a method to evaluate the literature and could be prerequisite to conducting a meta-analysis. Five studies may or may not be enough of a literature to conduct a meta-analysis, but behavior analysts can have some confidence in an intervention that meets these criteria and future meta-analysts can later examine the conditions that enhance the effectiveness of the intervention.

Although the special issue presents a logical argument for conducting meta-analyses of SCD studies, multiple studies that examine effect sizes within meta-analyses of SCD studies, and multiple examples, the purpose of this issue is not to convince SCD researchers to conduct meta-analyses. Instead, we hope to persuade SCD researchers that meta-analysis is a topic worthy of additional consideration. Given the important movement toward EBP in education and the purported desire of behavior analysts to more directly influence that movement, additional research on meta-analytic procedures and resulting meta-analyses both seem warranted.

References

American Psychological Association Presidential Task Force on Evidence-based Practice. (2006). Evidence-based practice in psychology. *American Psychologist, 61*, 271–285.

American Psychological Association Task Force on Evidence-Based Practice for Children and Adolescents. (2008). *Disseminating evidence-based practice for children and adolescents: A systems approach to enhancing care*. Washington, DC: American Psychological Association.

Baer, D. M. (1977). Reviewer's comment: Just because it's reliable doesn't mean that you can use it. *Journal of Applied Behavior Analysis, 10*, 117–119.

Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis, 12*, 199–210.

Baron, A., & Derenne, A. (2000). Quantitative summaries of single-subject studies: What do group comparisons tell us about individual performances? *Behavior Analyst, 23*, 101–106.

Berliner, D. C. (2002). Educational research: The hardest science of all. *Educational Researcher, 31*, 18–20.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis: Statistics in practice*. New York: Wiley.

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

Busk, P. L., & Serlin, R. C. (1992). Meta-analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 187–212). Hillsdale, NJ: Lawrence Erlbaum.

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago: Rand McNally.

Cook, B. G., Tankersley, M., & Landrum, T. J. (2009). Determining evidence-based practices in special education. *Exceptional Children, 75*, 365–383.

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

Cooper, H., & Hedges, L. V. (2009). Research synthesis as a scientific process. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 3–18). New York: Sage.

Cooper, A., Levin, B., & Campbell, C. (2009). The growing (but still limited) importance of evidence in education policy and practice. *Journal of Educational Change, 10*(2–3), 159–171.

Cordray, D. S., & Morphy, P. (2009). Research synthesis and public policy. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 473–494). New York: Sage.

Dunkin, M. J. (1996). Types of errors in synthesizing research in education. *Review of Educational Research, 66*(2), 87–97.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3–8.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York: Routledge.

Heward, W. L. (2008). A place at the educational reform table: Why behavior analysis needs to be there, why it's not as welcome as it should be, and some actions that can make our science more relevant. *ABAI Newsletter, 31*(3). Portage, MI. Available online at http://www.abainternational.org/ABA/newsletter/vol313/Heward.asp.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165–179.

Kavale, K. A., & Forness, S. R. (2000). Policy decisions in special education: The role of meta-analysis. In R. Gersten, E. P. Schiller, & S. Vaughn (Eds.), *Contemporary special education research: Syntheses of knowledge base on critical instructional issues* (pp. 281–326). Mahwah, NJ: Lawrence Erlbaum Associates.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., et al. (2010). *Single-case designs technical documentation*. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Levant, R. F., Barlow, D. H., David, K. W., Hagglund, K. J., Hollon, S. D., Johnson, J. D., et al. (2006). Evidence-based practice in psychology. *American Psychologist, 61*, 271–285.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Lochman, J. E. (2003). Commentary: School contextual influences on the dissemination of interventions. *School Psychology Review, 32*, 174–177.

Matt, G. E., & Cook, T. D. (2009). Threats to the validity of generalized inferences. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 538–5578). New York: Sage.

Myles, F. S., Allison, D. B., & Gorman, B. S. (1996). Meta-analysis of single-case research. In R. D. Franklin, D. B. Allison, & B. S. Gorman (Eds.), *Design and analysis of single-case research* (pp. 245–277). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 40*, 357–367.

Riley-Tillman, T. C., & Burns, M. K. (2009). *Evaluating educational interventions: Single-case design for measuring response to intervention*. New York, NY: The Guilford Press.

Rothwell, P. M. (2005). External validity of randomized control trials: "To whom do the results of this trial apply?". *The Lancet, 365*(9453), 82–93.

Salzberg, C. L., Strain, P. S., & Baer, D. M. (1987). Meta-analysis for single-subject research: When does it clarify, when does it obscure. *Remedial and Special Education, 8*, 43–48.

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remediate and Special Education, 8*, 24–33.

Shadish, W. R., Rindskopf, D. M., & Hedges, L. V. (2008). The state of the science in the meta-analysis of single-case experimental designs. *Evidence-Based Communication Assessment and Intervention, 2*, 188–196.

Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., et al. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American Journal of Community Psychology, 41*, 171–181.

White, O. (1987). The quantitative synthesis of single-subject research: Method and validation: Comment. *Remedial and Special Education, 8*, 34–39.