Methodological standards in single-case experimental design: Raising the bar

Jennifer B. Ganz\textsuperscript{a,}\textsuperscript{*}, Kevin M. Ayres\textsuperscript{b}

\textsuperscript{a} Texas A&M University, 4225 TAMU, College Station, TX 77843-4225, USA
\textsuperscript{b} Center for Autism and Behavioral Education Research, The University of Georgia, 509 Aderhold Hall, Athens, GA 30602, USA

ARTICLE INFO

Number of reviews completed is 2

Keywords:
Single-case experimental design
Single-subject research
Methodological quality standards
Guidelines
Developmental disability

ABSTRACT

Single-case experimental designs (SCEDs), or small-n experimental research, are frequently implemented to assess approaches to improving outcomes for people with disabilities, particularly those with low-incidence disabilities, such as some developmental disabilities. SCED has become increasingly accepted as a research design. As this literature base is needed to determine what interventions are evidence-based practices, the acceptance of SCED has resulted in increased critiques with regard to methodological quality. Recent trends include recommendations from a number of expert scholars and institutions. The purpose of this article is to summarize the recent history of methodological quality considerations, synthesize the recommendations found in the SCED literature, and provide recommendations to researchers designing SCEDs with regard to essential and aspirational standards for methodological quality. Conclusions include imploring SCED to increase the quality of their experiments, with particular consideration regarding the applied nature of SCED research to be published in Research in Developmental Disabilities and beyond.

What this paper adds

This paper provides guidance to single-case experimental researchers with regard to the quality of the experiments they design and report, based on the most current literature. We highlight the relationship between increasing standards for single-case experimental design and the impetus to identify evidence-based practices for the education of individuals with developmental disabilities. This paper may serve as a guide to researchers who wish to publish single-case experimental research in RIDD.

1. Introduction

Single-case experimental designs (SCEDs), or N-of-1 experiments, are prominent in the disability literature, particularly given research on low-incidence disabilities, such as severe intellectual and developmental disabilities (IDD; Tate, Perdices, McDonald, Togher, & Rosenkoetter, 2014). Although previously not held in high regard, within the recent past, such designs have become more widely accepted (Wilson, 2011). However, increased acceptance should come with increased scrutiny. Despite efforts of SCED experts to increase methodological standards over the past decades, we, as editor and associate editor of high-quality journals on autism spectrum and developmental disabilities, continue to see submissions to Research in Developmental Disabilities (RIDD) and SCEDs that are published with questionable methodological quality. Beyond journal submissions, our own work conducting single-case research,
meta-analyses, and systematic reviews has led us to advocate for increased methodological rigor across the literature base, both for the sake of the reputation of SCED and for the purpose of allowing determination of evidence-based practices. That is, without strongly designed and relevant SCED research, researchers cannot possibly draw conclusions regarding whether or not interventions are evidence-based and for whom and under what conditions these interventions are most likely to be effective.

1.1. History of literature on methodological standards

For approximately the last decade and a half, there has been an increasing focus on identifying evidence-based practices in special education, which necessitates evaluation of the methodological quality of SCED literature and status of the evidence of said literature (see Fig. 1. Evolution of Methodological Standards). Primary drivers of this movement have included legislation (No Child Left Behind in 2001; the Individuals with Disabilities Education Improvement Act in 2004), which required that educators consider and, whenever possible, apply scientifically-supported interventions for all students, including those with disabilities.

The process of developing and revising standards for SCED is iterative. That is, research and educational organizations have published recommended standards, which have been followed by critiques of these standards, which have been followed by further revision of standards. Key researchers and organizations are provided in Fig. 1. Highlights include initial pressure on educational institutions to implement research-supported educational interventions; this pressure was driven by legislation and statements from national institutes (Council for Exceptional Children, 2014; National Research Council, 2002; What Works Clearinghouse, 2014). Further, individuals and groups of researchers have participated in development of these standards or have published critiques and additional suggestions (Horner et al., 2005; Kratochwill & Stoiber, 2002; Tate et al., revised in 2013, 2014, 2016; Horner et al., 2005; Smith, 2012; Wendt & Miller, 2012; Council for Exceptional Children, 2014), including developing research review protocols (Tate et al., 2008, 2014).

1.2. Highlights of recommended standards reported and evaluated in the literature

The question of which criteria among those proposed are necessary and critical to developing and reporting a high-quality SCED is unanswered. Reviewing the literature exposes a range of priorities. This discrepancy across authors and evaluation tools is problematic (Maggin, Briesch, Chafouleas, Ferguson, & Clark, 2014). Primarily, use of different evaluation tools results in disagreement regarding the standing of literature base and, thus, the state of the evidence for intervention practices and for whom and under what conditions they may be considered to be most effective. Furthermore, as standards and expectations evolve over time, research considered rigorously designed a decade ago may no longer be viewed as adequate. Thus, we have reviewed the literature to produce a sampling of the standards to enable us to best propose best practices in SCED quality based on this literature and our own experience and expertise as experienced single-case researchers. This summary of previously recommended standards is provided in Table 1.

While there appears to be some agreement across quality rubrics with regard to criteria that evaluate the ability of the design to document an experimental effects if it exists (e.g., number of data points per phase, systematic manipulation of the independent variable, collection of inter-rater reliability data); there is less agreement related to components related to generalizability of studies (e.g., description of participants, description of the procedures, collection of procedural integrity data) (Maggin et al., 2014; Moeller, Dattilo, & Rusch, 2015). Further, while some sources recommend the following requirements, most do not: stability or contratherapeutic trend of baseline data, detailed requirements regarding numbers of data points required per phase, inclusion of data measuring procedural integrity, description of recruitment procedures, blinding procedures, and supplementing of visual analysis with statistical analysis (CEC, 2014; Horner et al., 2005; Kratochwill et al., 2010; Logan, Hickman, Harris, & Heriza, 2008; Maggin et al., 2014; Reichow et al., 2008; Tate et al., 2014; Wolery, 2013). In fact, existing standards document have been criticized for ignoring elements of quality that allow for generalization of the results (Hitchcock, Kratochwill, & Chezan, 2015) or assurance of internal integrity (Wolery, 2013).

While systematic literature reviews and meta-analyses have increased the use of quality indicator analyses, there are still
significant gaps in the elements evaluated. SCED aggregate analyses should evaluate quality related to the ability of a design to demonstrate experimental control, systematic manipulation of the intervention, the number of data points per phase, and collection of interobserver-agreement data (Fallon, Collier-Meek, Maggin, Sanetti, & Johnson, 2015; Ganz et al., 2012; Ganz et al., 2017; Hong et al., 2015; Maggin, Zurheide, Pickett, & Baillie, 2015; Schlosser & Wendt, 2008; Smith, 2012). Less frequently, reviews have included evaluation of items related to procedural fidelity, social validity, and the description of who implemented the intervention (Ledford & Wolery, 2013; Westbrook et al., 2015).

The field’s disagreement regarding criteria for high-quality SCEDs leaves researchers without definitive guidance and journal editors and reviewers without set standards by which to judge the design quality methodology of papers submitted for publication in professional journals (Maggin, 2015). Thus, this article attempts to ameliorate this issue for authors submitting to RIDD, single-case researchers who design studies to evaluate interventions for people with IDD, and, at best, the field at large. The purpose of this article is to encourage single-case researchers to increase their methodological rigor when developing, implementing, and writing single-case experiments. Additionally, and more specifically, given the focus of this journal and our own research agendas, we are particularly concerned with the state of the SCED literature that focuses on individuals with developmental disabilities, including intellectual disability and autism spectrum disorder. Further, we urge researchers submitting manuscripts to RIDD to take particular

<table>
<thead>
<tr>
<th>Authors, in Chronological Order</th>
<th>Participant Description</th>
<th>Setting and Materials Description</th>
<th>Interventionist Description</th>
<th>Response Definition and Measurement</th>
<th>Collection &amp; Adequacy of Inter-observer Agreement</th>
<th>Maintenance &amp; Generalization Data Collection</th>
<th>Procedural Description &amp; Procedural Fidelity</th>
<th>Research Design</th>
<th>Social Validity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wolery &amp; Ezell (1993)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horner et al. (2005)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Logan, Hickman, Harris, &amp; Heriza (2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reichow, Volkmar, &amp; Cicchetti (2008)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kratochwill et al. (2010, revised by WWC 2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ledford &amp; Wolery (2013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ledford, Wolery, Meeker, &amp; Webby (2012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CEC (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maggin, Briesch, Chafouleas, Ferguson, &amp; Clark (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tate et al. (2014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
note of our proposed standards, as they are ones that are likely to be used by reviewers and editors in making decisions regarding submissions. By putting forth these standards, we desire to aid authors in the development of more rigorous single case experiments while strengthening analysis and improving reporting.

2. Proposed standards for single case experimental design studies

We provide, in Table 2, proposed increased standards, gleaned from the recent literature, including the literature highlighted in Table 1 (e.g., Council for Exceptional Children, 2014; Horner et al., 2005, Kratchwill et al., 2010; Reichow et al., 2008; Tate et al., 2008, 2014; What Works Clearinghouse, 2014). These sources were selected via our review of frequently cited commentary papers and reviews of the quality of the literature base related to identification of evidence-based practices for individuals with IDD and based in our experience reviewing submitted manuscripts as editors and our own experience conducting systematic reviews and meta-analyses of this literature.

Readers and authors should not take this set of standards to be a definitive synthesis of existing methodological standards and quality indicators. Rather, these standards draw on the strengths of existing standards and the careful consideration those authors dedicated to providing researchers a framework to design future studies. These standards are meant to help researchers and authors design and report studies in RIDD, and elsewhere, that have adequate rigor to support causal relations and refute other plausible explanations for changes in the dependent variable. In sum, these standards would ensure an acceptable level of protection against threats to internal validity. Our recommended standards are organized according to the typical structure of a research report to highlight key elements that should be reported in RIDD and cover (a) participant descriptions; (b) setting descriptions; (c) describing materials; (d) describing who is implementing the intervention; (e) dependent measures or outcomes and data collection; (f) critical data to collect and description of collection of any maintenance and generalization data (i.e., participant performance over time and in varied contexts); (g) describing procedures for baseline and implementation; (h) developing the research design; (i) attending to social validity-related issues; and (j) critical aspects in reporting results (see Table 2). As such, they serve as an outline of sorts for the development and reporting of SCED.

Participants in a study can be individuals or aggregated groups (e.g. classroom) but authors need to provide key information about participants to permit interpretation, replication, as well as application in genuine contexts. Failing to provide this information will inhibit researchers from putting results in context and impede practitioners from adapting interventions for practice. Uniformly, the key considerations authors should weigh relates to relevance of the information and the extent that it will help readers understand the participants. Relatedly, to allow for adequate information for practitioners and replication and thus further study of potential evidence-based practices, it is critical that researchers report, in replicable detail, the materials, settings, and interventionists involved in the study.

Applied research has largely arisen out of clinical and classroom contexts. Sometimes in these conditions practitioners may sacrifice accuracy of data collection for feasibility (e.g. using momentary time sampling instead of a continuous measure). While we will always harvest valuable results from these “real life” settings, when possible, researchers should strive for greater precision and more complete reporting of behavior in context. Measurement begins with clearly defining dependent variables of interest. Response definitions should be unambiguous, objective, and exclusive when possible (i.e., instead of “tantrums” count each individual behavior). Further, simply collecting and reporting data during the baseline and intervention phases is inadequate. That is, researchers should demonstrate whether their results are maintained beyond intervention cessation (i.e., maintenance of skills learned) and whether or not skills generalize to untrained contexts (e.g., new communicative partners, novel settings, varied materials). Learned skills are of little use if they are not maintained long term and demonstrated across all relevant contexts.

With similar clarity, researchers must report their procedures in a manner that will facilitate replication. The technical precision of these descriptions should include not only all procedural steps but also highlight fundamental differences between conditions; Wolery and Ezzell (1993) argued that when one wants to generalize findings from one study to another, baseline performance and the reason for baseline performance (i.e. problem behavior maintained by attention) were more important than any status variables and psychometrics. Authors must also report on their attempts to evaluate procedural fidelity. The relation between poor procedural fidelity and contratherapeutic results can be an interesting outcome itself. For these results to be meaningful, the data have to be collected in the context of a rigorous methodology that permits the evaluation of a functional relation. Researchers then have to communicate these results clearly so that readers can judge the outcomes for themselves. That table below provides a summary of the most basic elements a researcher should consider when designing a study. Alongside this, the table shows desirable study characteristics that would guide a researcher toward developing studies of the highest quality and rigor.

3. Conclusions

The researcher approaches his or her question with an aim to better understand some phenomena of interest. If the researcher wants to know “the” answer and not just “an” answer, they should apply the most rigorous methodology they can to rule out all other plausible explanations. This is what separates research from simply data collection. Then, to contribute to the research literature, that researcher must effectively communicate what they did, in what context they did it, and the effects of their manipulation of the independent variable. By doing so, the researcher contributes knowledge and information in a format that permits careful examination by the research community and possible replication. These essential and aspirational standards are an effort to set a contemporary threshold for manuscripts submitted to RIDD. By designing and reporting studies according to these essential standards, the articles published in RIDD stand to have a greater impact on the field.
Table 2
Recommended Methodological Standards.

<table>
<thead>
<tr>
<th>Essential</th>
<th>Aspirational</th>
</tr>
</thead>
</table>
| **Participant Descriptions** | • Adherence to American Psychological Association (APA) guidelines for referring to participants (in relation to race, gender or sex, sexual orientation, person first language)  
•Acknowledgement of informed consent  
•Demographic variables, including age and any other relevant status variable (e.g. gender) whose associated learning history might impact outcomes  
•Disability diagnosis/category (if any)  
•Familiarity between the researcher/interventionist and the participant  
•Diagnostic assessment and psychometric information that is relevant to the study and not just tangential (e.g. reading comprehension scores from a standardized test in a study about reading)  
•Information about the educational context in which the participants received services (if appropriate)  
•Reason for baseline performance (e.g. function of problem behavior or not previously exposed to the material), if the baseline data are not flat or do not demonstrate need for intervention  
•Description of how the materials worked (e.g. software)  
•If a commercially available product, the manufacturer and model  
•Description of the context and environmental arrangements of the physical space (e.g. how many chairs, desks)  
•Description of other people present and ratio of staff to participants/students/clients  
•Approximate dimensions of the space or spaces where the study took place  
•Arrangement of the physical space (e.g. how many chairs, desks)  
•Description of the context and environmental characteristics (e.g., classroom, grocery store, noise levels, other stimuli) in which the procedures took place  
•Size, color, dimensions, quantity  
•If the materials are not subject to copyright and can be shared on line, links to examples or copies  
•Occupation  
•Relationship to the participant(s)  
•Education level  
•Experience/training relevant to the study  
•When discontinuous systems are used (e.g., interval systems), authors report sufficient information regarding the percentage of total observations for samples of the data to permit readers to ascertain approximate error  
•Video record sessions  
•Description of qualifications of data collectors (e.g., relevant training and experience, educational level, relationship to the participant)  
•Description of training for data collectors  
•Report IOA based on data collector, phase, condition, and participant  
•IOA data were provided for procedural integrity and met the standards for IOA (continued on next page) |

(continued on next page)
The expectations of the academic community consuming research from RIDD will evolve over time. As researchers learn more about the limitations of our current measurement systems for example, new ones will arise. Technology will likewise play a greater role in data collection and in analysis of those data. While the aspirational methodological and reporting recommendations made above may appear to be something to strive for in the future, they will likely become the minimum expectations. While researchers will continually consider whether it is possible to implement studies in potentially less controlled, but more natural contexts, which will lead to more generalization, or applicability, to real-life contexts, including implementation by interventionists typically found in such contexts, rather than highly-trained interventionists, such as researchers and scientists we will always approach our results skeptically and avail ourselves of the built-in error correcting machinery (Sagan, 27), because this is what gives our credibility as a field. Continuously striving to refine our knowledge through better designs and communicating that new knowledge clearly and widely. Strong designs and careful methodology permit confidence in our work and allow us to have a greater impact on the field of developmental disability.

Finally, although oft not included in lists of critical standards for SCED (see Table 1), we encourage SCED researchers to seriously consider the application of their work in real life contexts; that is, researchers must consider the social validity, or importance of the controlled, but more natural contexts, which will lead to more generalization, or applicability, to real-life contexts, including implementation by interventionists typically found in such contexts, rather than highly-trained interventionists, such as researchers and scientists.
their graduate students. This involves careful consideration of the social validity, intervention fidelity, and feasibility of interventions, areas which are often ignored with regard to methodological quality. Ultimately, treatments that are better suited for natural contexts, and have been evaluated as such, are the ones that will be most useful for practitioners and other key stakeholders.

**Funding**

This work did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**References**


