Archives of Scientific Psychology 2016, 4, 10–31
DOI: http://dx.doi.org/10.1037/arc0000027

2169-3269

American Psychological Association

Archives of Scientific Psychology

www.apa.org/pubs/journals/arc

The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016: Explanation and Elaboration

Robyn L. Tate
The Kolling Institute of Medical Research, St Leonards, New South Wales, Australia, and The University of Sydney

Michael Perdices
Royal North Shore Hospital, St Leonards, New South Wales, Australia, and The University of Sydney

Ulrike Rosenkoetter
The Kolling Institute of Medical Research, St Leonards, New South Wales, Australia, and The University of Sydney

Skye McDonald
University of New South Wales

Leanne Togher
The University of Sydney

William Shadish
University of California, Merced

Robert Horner
University of Oregon

Thomas Kratochwill
University of Wisconsin—Madison

David H. Barlow
Boston University

Alan Kazdin
Yale University

Margaret Sampson
Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada

Larissa Shamseer
Ottawa Hospital Research Institute, Ottawa, Ontario, Canada, and University of Ottawa

Sunita Vohra
University of Alberta

(For the SCRIBE Group)

ABSTRACT

There is substantial evidence that research studies reported in the scientific literature do not provide adequate information so that readers know exactly what was done and what was found. This problem has been addressed by the development of reporting guidelines which tell authors what should be reported and how it should be described. Many reporting guidelines are now available for different types of research designs. There is no such guideline for one type of research design commonly used in the behavioral sciences, the single-case experimental design (SCED). The present study addressed this gap. This report describes the Single-Case Reporting guideline In BEhavioural interventions (SCRIBE) 2016, which is a set of 26 items that authors need to address when writing about SCED research for publication in a scientific journal. Each item is described, a rationale for its inclusion is provided, and examples of adequate reporting taken from the literature are quoted. It is recommended that the SCRIBE 2016 is used by authors preparing manuscripts describing SCED research for publication, as well as journal reviewers and editors who are evaluating such manuscripts.

SCIENTIFIC ABSTRACT

Single-case experimental design (SCED) studies in the behavioral sciences literature are not only common, but their proportion has also increased over past decades. Moreover, methodological complexity of SCEDs and sophistication in the techniques used to analyze SCED data has increased apace. Yet recent reviews of the behavioral sciences literature have shown that reporting of SCED research is highly variable and often incomplete. Explicit, precise and

transparent reporting is crucial not only for critical evaluation of the study methodology and conclusions, but also to facilitate exact replication of investigations, and ascertain applicability and possible generality of results. Accordingly, we developed the SCRIBE 2016 (Single-Case Reporting guideline In BEhavioural interventions) by a consensus process by experts in SCED methodology and research in the behavioral sciences, as well as experts in reporting guideline development. The SCRIBE 2016 Explanation and Elaboration article describes a set of 26 items to guide and structure the reporting of SCED research. A rationale and minimum reporting standards that stipulate what needs to be reported are provided for each item. In addition, examples of adequate and clear reporting drawn from the literature are included for each item. It is recommended that the SCRIBE 2016 Explanation and Elaboration article is used in conjunction with the complementary SCRIBE 2016 Statement (Tate et al., 2016) by authors preparing manuscripts for publication and journal reviewers and editors considering manuscripts for publication.

Keywords: single-case design, methodology, reporting guidelines, publication standards
Supplemental materials: http://dx.doi.org/10.1037/arc0000027.supp

> Essentially, without publication, the research remains invisible to the world. And yet, too often, reading these articles leaves us unable to determine exactly how the research was conducted, what was found, how reliable the findings are and how they fit into the wider context of existing knowledge. Many published articles are not fit for purpose.
>
> —(Simera, Moher, Hoey, Schulz, & Altman, 2010, p. 35)

Single-case experimental designs (SCEDs) are used frequently in the behavioral sciences. Shadish and Sullivan (2011) surveyed the contents of 21 journals in psychology and education for the calendar year 2008 and found that 44% of intervention studies used single-case methods. Similarly, 39% of records archived on the PsycBITE evidence database (www.psycbite.com), representing all published nonpharmacological interventions for psychological consequences of acquired brain impairment, used single-case methods (Perdices et al., 2006). This result is comparable to Beeson and Robey's (2006) findings for the specific domain of aphasiology (41%). Both Hammond and Gast (2010) and Maggin, O'Keeffe, and Johnson (2011) demonstrated an accelerating trend for an increased number of single-case reports published over recent decades. SCEDs are also used in medicine (Gabler, Duan, Vohra, & Kravitz, 2011), where they are specifically referred to as *N*-of-1 trials. This variety of SCED consists of multiple cross-overs (or phase changes) using the withdrawal A-B-A-B paradigm in a single participant who serves as his or her own control, often incorporating randomization and/or blinding (Kravitz, Duan, & the DEcIDE Methods Center *N*-of-1 Guidance Panel, 2014). Well-designed and conducted SCEDs provide a strong level of evidence, and in particular the randomized *N*-of-1 trial provides Level 1

This article was published April 14, 2016.

Robyn L. Tate, John Walsh Centre for Rehabilitation Research, The Kolling Institute of Medical Research, St Leonards, New South Wales, Australia, and Sydney Medical School Northern, The University of Sydney; Michael Perdices, Department of Neurology, Royal North Shore Hospital, St Leonards, New South Wales, Australia, and Discipline of Psychiatry, The University of Sydney; Ulrike Rosenkoetter, John Walsh Centre for Rehabilitation Research, The Kolling Institute of Medical Research, and Sydney Medical School Northern, The University of Sydney; Skye McDonald, School of Psychology, University of New South Wales; Leanne Togher, Discipline of Speech Pathology, The University of Sydney; William Shadish, School of Social Sciences, Humanities and Arts, University of California, Merced; Robert Horner, Department of Special Education and Clinical Services, University of Oregon; Thomas Kratochwill, School of Educational Psychology, University of Wisconsin—Madison; David H. Barlow, Centre for Anxiety Related Disorders, Boston University; Alan Kazdin, Department of Psychology, Yale University; Margaret Sampson, Library and Media Services, Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada; Larissa Shamseer, Clinical Epidemiology Program, Ottawa Hospital Research Institute, Ottawa, Ontario, Canada, and School of Epidemiology, Public Health and Preventive Medicine, University of Ottawa; Sunita Vohra, Department of Pediatrics, University of Alberta.

The SCRIBE Group wishes to pay special tribute to our esteemed colleague Professor William Shadish (1949–2016) who passed away on the eve of publication of this article. His contribution at all stages of the SCRIBE project was seminal.

Funding for the SCRIBE project was provided by the Lifetime Care and Support Authority of New South Wales, Australia. The funding body was not involved in the conduct, interpretation or writing of this work. Members of the SCRIBE Group are as follows: Richard Albin (University of Oregon), Catherine Backman (University of British Columbia), David H. Barlow (Boston University), Jacinta Douglas (La Trobe University), Jonathan J. Evans (University of Glasgow), David Gast (University of Georgia), Robert Horner (University of Oregon), Alan Kazdin (Yale University), Thomas Kratochwill (University of Wisconsin-Madison), Rumen Manolov (University of Barcelona), Skye McDonald (University of New South Wales), Geoffrey Mitchell (The University of Queensland), Lyndsey Nickels (Macquarie University), Jane Nikles (The University of Queensland), Tamara Ownsworth (Griffith University), Michael Perdices (Royal North Shore Hospital, New South Wales, Australia, and The University of Sydney), Miranda Rose (La Trobe University), Ulrike Rosenkoetter (The Kolling Institute of Medical Research, St Leonards, New South Wales, Australia, and The University of Sydney), Margaret Sampson (Children's Hospital of Eastern Ontario, Ottawa, Ontario, Canada), Christopher H. Schmid (Brown University), William Shadish (University of California, Merced), Larissa Shamseer (Ottawa Hospital Research Institute, Ottawa, Ontario, Canada, and University of Ottawa), Robyn L. Tate (The Kolling Institute of Medical Research, St Leonards, New South Wales, Australia, and The University of Sydney), Leanne Togher (The University of Sydney), Sunita Vohra (University of Alberta), and Barbara Wilson (Oliver Zangwill Centre, Ely, Cambridgeshire, United Kingdom). For further discussion on this topic, please visit the *Archives of Scientific Psychology* online public forum at http://arcblog.apa.org.

Correspondence concerning this article should be addressed to Robyn L. Tate, John Walsh Centre for Rehabilitation Research, The Kolling Institute of Medical Research, Level 9, Royal North Shore Hospital, St Leonards NSW 2065, Australia. E-mail: robyn.tate@sydney.edu.au

evidence[1] for treatment decision purposes (Guyatt, Jaeschke, & McGinn, 2002; OCEBM, 2011).

Designs using single-case methods can be complex and sophisticated (e.g., a combination of a multiple-baseline with alternating-treatments design, this particular method comprising 10% of designs in the Shadish and Sullivan (2011) survey). Recently, a class of single-case designs has been proposed that involves a variety of randomization procedures and can increase the internal validity of these designs (Kratochwill & Levin, 2010), as well as provide options for data analysis (see Kratochwill & Levin, 2014). The range of single-case designs, however, is a potential source of misunderstanding because not all reports in the literature that study a single participant also apply single-case methodology, as defined.[2] Figure 1 depicts the common designs using a single-participant as reported in the literature. Components of the figure and their interrelationships are described in the companion Single-Case Reporting guideline In BEhavioural interventions (SCRIBE) 2016 Statement article (Tate et al., 2016). In brief, surveys of the literature (e.g., Perdices & Tate, 2009; Shadish & Sullivan, 2011) have identified multiple designs using a single participant. The figure presents nine prototypical designs, but not all of these use single-case methodology, as defined (i.e., containing multiple phases during each of which the dependent variable is measured repeatedly; see Footnote 2). In particular, the three designs below the solid horizontal line (B-phase training study, pre–post intervention evaluations alone, and case description) do not meet these criteria and they are not SCEDs.

The SCRIBE 2016 applies to those designs above the solid horizontal line, which all use single-case methodology, but differ fundamentally in terms of their structure. The four prototypical designs above the dotted horizontal line comprise the withdrawal/reversal, multiple-baseline, alternating/simultaneous-treatments, and changing-criterion designs. Reporting requirements differ among the designs, and are described in the SCRIBE 2016 Item 5. The medical *N*-of-1 trial falls within the withdrawal/reversal paradigm, and a separate reporting guideline is available for that design (see Shamseer et al., 2015 and Vohra et al., 2015 for the CONSORT Extension for *N*-of-1 Trials [CENT 2015]). Randomization of elements in all of the foregoing designs is feasible (see Kratochwill & Levin, 2010), albeit not common practice, and is covered in the SCRIBE 2016 Item 8.

Each of the designs also varies in terms of methodological rigor and even experimental designs above the dotted horizontal line may not meet design standards, such as those of Horner et al. (2005) and Kratochwill et al. (2013). The reason that the biphasic A-B design is separated from the other designs above the dotted horizontal line is because of its poor control of threats to internal validity. It is for this reason that the A-B design, although using single-case methodology, is regarded as quasiexperimental (Barlow, Nock, & Hersen, 2009). Scientific quality for both internal and external validity can be measured with methodological quality rating scales designed for single-case methodologies, such as those described in Maggin et al. (2014) and Tate et al. (2013b).

The behavioral sciences literature is highly variable with respect to the adequate conduct and complete reporting of single-case research. The systematic review of Maggin, Chafouleas, Goddard, and Johnson (2011) into SCEDs to evaluate token economy interventions for students with challenging behaviors found evidence of incomplete reporting. Of 24 eligible studies, a significant proportion failed to report on basic demographic features of the participants, such as age (42%) and sex (33%). Moreover, 21% of the studies did not provide information on who implemented the intervention and 42% failed to specify the method used to record the data. Didden, Korzilius, van Oorsouw, and Sturmey (2006) conducted a meta-analysis of 80 single-case reports on challenging behavior in mild-to-moderate intellectual disability, finding that only 27% reported on procedural fidelity. Tate et al. (2014) reported on a random sample of 35% of reports archived on the PsycBITE database published between 1990 and 2010 in the neurological conditions of dementia, stroke and traumatic brain injury that used a single participant ($n = 253$). Only 14% reported using an assessor who was independent of the therapist, 54% reported on interrater reliability of the dependent variable, and 62% provided a session-by-session data record in graphed or tabular format. Smith (2012) systematically reviewed the psychology and education literature between 2000 and 2010 specifically to identify SCEDs, with 409 reports meeting eligibility criteria. Twenty-two percent of studies did not report baseline data and 52% did not report either statistical or visual analyses of the data. These data from different populations suggest that problems with the conduct and/or reporting of fundamental elements of single-case research in the behavioral sciences are common and highlight the need for a reporting guideline.

Reporting guidelines in the CONSORT tradition (see www.equator-network.org) improve the clarity and transparency of reporting of randomized controlled trials (RCT) published in journals that endorse them. A systematic review based on evaluations of more than 16,000 RCTs found that 25 of 27 CONSORT-related items were more completely reported in journals that endorse CONSORT compared to those that do not, with five items being statistically significant ($p = .01$; Turner et al., 2012). It is expected that the development and implementation of similarly structured guidelines to cover common research designs used in SCEDs will assist (a) researchers to report on the requisite items that foster completeness, clarity, transparency and accuracy of reporting and (b) readers to know exactly what was done and what was found. The present report, referred to as the SCRIBE 2016 Explanation and Elaboration article, provides description of and rationale for 26 reporting items, along with examples of adequate reporting from the published literature. A separate SCRIBE 2016 Statement (Tate et al., 2016) describes the methodology underlying development of the SCRIBE 2016.

The genesis of the SCRIBE 2016 derives directly from the foundation work and development of the CENT 2015 (Shamseer et al., 2015; Vohra et al., 2015). Frequently, the medical interventions being evaluated in *N*-of-1 trials use pharmacological or nonpharmacological substances that are injected, ingested, inhaled, or topically applied. The design for these types of interventions sometimes involves run-in/wash-out periods, which are consequently highly relevant to adequate reporting of such trials and specific items are required for the purpose of reporting. By contrast, the SCRIBE 2016 is intended to apply to the broader variety of experimental single-case interventions used in the behavioral sciences, including, but not restricted to, health conditions. Behav-

[1] Levels of evidence refer to the hierarchy of strength or credibility of evidence that different research designs yield. The hierarchy is frequently used in medicine to critically evaluate the available evidence for different clinical questions (e.g., interventions, harms, diagnosis, prognosis). There is some variation among different classification systems (see, e.g., websites of the American Academy of Neurology, www.aan.com; Oxford Centre for Evidence-Based Medicine, www.cebm.net).

[2] Single-case methodology is defined as the intensive and prospective study of the individual in which (a) the intervention/s is manipulated in an experimentally controlled manner across a series of discrete phases, and (b) measurement of the behavior targeted by the intervention is made repeatedly (and, ideally, frequently) throughout all phases. Professional guidelines recommend that the experimental effect be demonstrated on at least three occasions by systematically manipulating the independent variable (Horner et al., 2005; Kratochwill et al., 2010, 2013). This criterion helps control for the confounding effect of extraneous variables that may adversely affect internal validity (e.g., history, maturation), and allows a functional cause and effect relationship to be established between the independent and dependent variables.

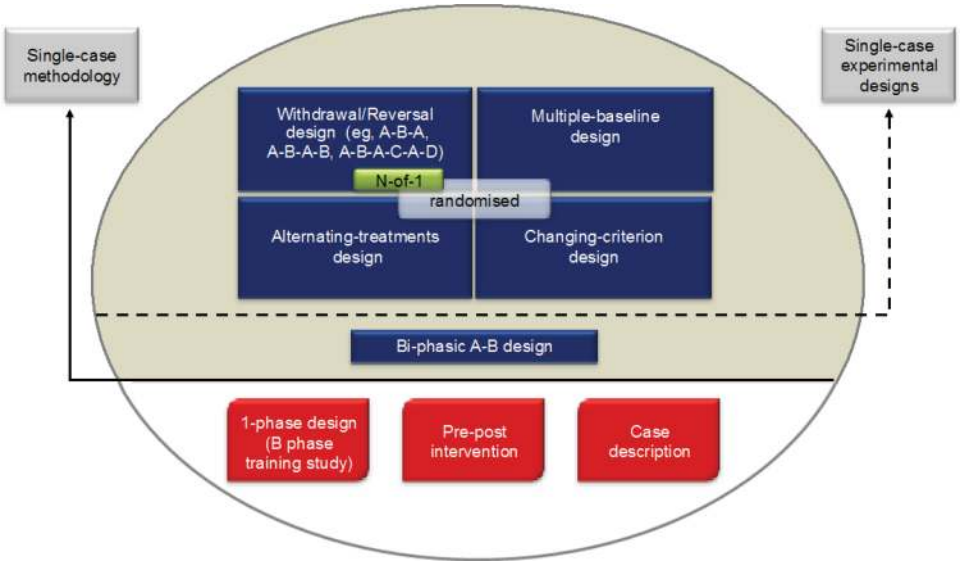

Figure 1. Common designs in the literature using a single participant. Reproduced from the expanded manual for the Risk of Bias in *N*-of-1 Trials (RoBiNT) Scale (Tate et al., 2015) with permission of the authors; an earlier version of the figure, taken from the original RoBiNT Scale manual (Tate et al., 2013a), was also published in 2013 (Tate et al., 2013b).

ioral interventions are often multicomponential and complex, and some strategies to minimize bias, such as blinding, are difficult to implement. For these reasons, two sets of reporting guidelines were deemed necessary to cater to the different types of interventions and single-case methodologies used in the respective fields. Similar reasoning drove the development of an extension of the CONSORT Statement for nonpharmacological RCTs (Boutron, Moher, Altman, Schulz, & Ravaud, 2008). The CENT 2015 guideline is intended for use in medical *N*-of-1 trials, whereas the SCRIBE 2016 guideline is intended for SCEDs in the behavioral sciences.

As with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement for observational studies (Vandenbroucke et al., 2007), the SCRIBE 2016 is intended to cover multiple research designs (specifically, the four most common prototypical experimental designs: withdrawal/reversal, multiple-baseline, alternating/simultaneous treatments, and changing-criterion designs, along with their variants, as well as adaptive designs). Although we did not intend the SCRIBE 2016 to apply to non-SCEDs using a single participant (designs shown below the solid horizontal line in Figure 1), authors of such articles may find it useful to follow the guidance of those SCRIBE 2016 items as may apply to their study. Authors may also wish to consult the reporting guide for clinical CAse REports (CARE; Gagnier et al., 2014), the CONSORT for Social and Psychological Interventions (CONSORT-SPI; Montgomery et al., 2013), and the Template for Intervention Description and Replication (TIDieR) guideline (Hoffmann et al., 2014).

We prepared the SCRIBE 2016 within the CONSORT tradition of guideline development (Moher, Schulz, Simera, & Altman, 2010b). Accordingly, the focus of this article concerns the *reporting* of studies, rather than *education* about the design of single-case experiments. Moreover, in our use of examples from the literature to illustrate adequate reporting of each of the SCRIBE 2016 items, we adopt the position of Boutron et al. (2008), wherein use of an example for a specific item is not intended to imply that the study also provides adequate examples of other items, nor even that the study per se is methodologically sound. Thus, although some of the examples may not meet design standards, nonetheless they meet reporting standards in that they clearly describe what was done. The suggested locations in the article for the SCRIBE 2016 reporting items are not prescriptive and authors should use their discretion about the most suitable location.

Methodology

The first phase to develop the SCRIBE 2016 consisted of two rounds of an online Delphi survey completed by SCED authors and methodology experts, resulting in 44 items to be discussed at a consensus conference. At the meeting, held in Sydney, Australia, in December 2011, participants reworked the items, resulting in a final set of 26 items for the reporting guideline, which are described in this Explanation and Elaboration article. The SCRIBE 2016 Statement (Tate et al., 2016) provides a detailed description of the methodology of this process.

Results

This article provides examples of adequate reporting from the literature for each of the 26 items, along with a rationale for inclusion of the item and, where available, evidence of bias resulting from incomplete reporting (the SCRIBE checklist of items appears in Table 1).

Section 1: Title and Abstract (Items 1 and 2)

Item 1—Title: Identify the research as a single-case experimental design in the title.

Example.

> Graded exposure in vivo in the treatment of pain-related fear: a replicated single-case experimental design in four patients with chronic low back pain. (Vlaeyen, de Jong, Geilen, Heuts, & van Breukelen, 2001, p. 151)

Explanation. Although journals may place word limits on the title, it is important to include as much information as possible in the title, such as the intervention, the target behavior and the population. In particular, the title should explicitly mention that the study is a

Table 1
The Single-Case Reporting Guideline In BEhavioural Interventions (SCRIBE) 2016 Checklist

Item number	Topic	Item description
TITLE and ABSTRACT		
1	Title	Identify the research as a single-case experimental design in the title
2	Abstract	Summarize the research question, population, design, methods including intervention/s (independent variable/s) and target behavior/s and any other outcome/s (dependent variable/s), results, and conclusions
INTRODUCTION		
3	Scientific background	Describe the scientific background to identify issue/s under analysis, current scientific knowledge, and gaps in that knowledge base
4	Aims	State the purpose/aims of the study, research question/s, and, if applicable, hypotheses
METHOD		
	DESIGN	
5	Design	Identify the design (e.g., withdrawal/reversal, multiple-baseline, alternating-treatments, changing-criterion, some combination thereof, or adaptive design) and describe the phases and phase sequence (whether determined a priori or data-driven) and, if applicable, criteria for phase change
6	Procedural changes	Describe any procedural changes that occurred during the course of the investigation after the start of the study
7	Replication	Describe any planned replication
8	Randomization	State whether randomization was used, and if so, describe the randomization method and the elements of the study that were randomized
9	Blinding	State whether blinding/masking was used, and if so, describe who was blinded/masked
	PARTICIPANT/S or UNIT/S	
10	Selection criteria	State the inclusion and exclusion criteria, if applicable, and the method of recruitment
11	Participant characteristics	For each participant, describe the demographic characteristics and clinical (or other) features relevant to the research question, such that anonymity is ensured
	CONTEXT	
12	Setting	Describe characteristics of the setting and location where the study was conducted
	APPROVALS	
13	Ethics	State whether ethics approval was obtained and indicate if and how informed consent and/or assent were obtained
	MEASURES and MATERIALS	
14	Measures	Operationally define all target behaviors and outcome measures, describe reliability and validity, state how they were selected, and how and when they were measured
15	Equipment	Clearly describe any equipment and/or materials (e.g., technological aids, biofeedback, computer programs, intervention manuals or other material resources) used to measure target behavior/s and other outcome/s or deliver the interventions
	INTERVENTIONS	
16	Intervention	Describe the intervention and control condition in each phase, including how and when they were actually administered, with as much detail as possible to facilitate attempts at replication
17	Procedural fidelity	Describe how procedural fidelity was evaluated in each phase
	ANALYSIS	
18	Analyses	Describe and justify all methods used to analyze data
RESULTS		
19	Sequence completed	For each participant, report the sequence actually completed, including the number of trials for each session for each case. For participant/s who did not complete, state when they stopped and the reasons
20	Outcomes and estimation	For each participant, report results, including raw data, for each target behavior and other outcome/s
21	Adverse events	State whether or not any adverse events occurred for any participant and the phase in which they occurred
DISCUSSION		
22	Interpretation	Summarize findings and interpret the results in the context of current evidence
23	Limitations	Discuss limitations, addressing sources of potential bias and imprecision
24	Applicability	Discuss applicability and implications of the study findings
DOCUMENTATION		
25	Protocol	If available, state where a study protocol can be accessed
26	Funding	Identify source/s of funding and other support; describe the role of funders

SCED because this differentiates the study from a case description. The abstract is sometimes copyrighted by the journal, so the title may be the only searchable information. Identifying the study as a SCED in the title will ensure that the article is appropriately indexed for bibliographic databases (such as PsycINFO or Medline). Note that using "SCED" as a key word will not be sufficient for this purpose because author key words are different from database key words and may therefore not be searchable in electronic databases.

Item 2—Abstract: Summarize the research question, population, design, methods including intervention/s (independent variable) and target behavior/s and any other outcome measures (dependent variable), results, and conclusions.

Example.

> This study tested the effectiveness of Imagery Rescripting (ImRs) for complicated war-related PTSD [posttraumatic stress disorder] in refugees. Ten adult patients in long-term supportive care with a primary diagnosis of war-related PTSD and Posttraumatic Symptom Scale (PSS) score >20 participated. A concurrent multiple baseline design was used with baseline varying from 6 to 10 weeks, with weekly supportive sessions. After baseline, a 5-week exploration phase followed with weekly sessions during which traumas were explored, without trauma-focused treatment. Then 10 weekly ImRs sessions were given followed by 5-week follow-up without treatment. Participants were randomly assigned to baseline length, and filled out the PSS and the BDI [Beck Depression Inventory] on a weekly basis. Data were analyzed with mixed regression. Results revealed significant linear trends during ImRs (reductions of PSS and BDI scores), but not during the other conditions. The scores during follow-up were stable and significantly lower compared to baseline, with very high effect sizes (Cohen's $d = 2.87$ (PSS) and 1.29 [BDI]). One patient did clearly not respond positively, and revealed that his actual problem was his sexual identity that he couldn't accept. There were no dropouts. In conclusion, results indicate that ImRs is a highly acceptable and effective treatment for this difficult group of patients. (Arntz, Sofi, & van Breukelen, 2013, p. 274)

Explanation. The abstract needs to provide an accurate, informative and unambiguous overview of the study. It is important that all relevant information is included because many readers may not have access to the full article, or may choose to limit their reading of the study to the abstract. The CONSORT guideline for abstracts for randomized trials (Hopewell et al., 2008) provides useful information about how to write an abstract which, although written for RCTs, has applicability to SCEDs. A structured abstract can be useful and make the abstract easy to follow. It is important that the abstract clearly describes relevant features of the participant/s (including clinical details where appropriate), defines the dependent and independent variables, along with the SCED design used to examine their relationship. The target behaviors and any additional outcome measures (e.g., for generalization) used in the study should be specified, along with the way in which the target behavior is measured. An accurate summation of the outcomes of the study needs to be clearly detailed, along with disclosure of any harms or adverse events, and conclude with a brief, cautious appraisal of the significance of the research.

Section 2: Introduction (Items 3 and 4)

Item 3—Scientific background: Describe the scientific background to identify issue/s under analysis, current scientific knowledge, and gaps in that knowledge base.

Example (abbreviated).

> Verb production problems are an extremely common and pervasive aphasic deficit following stroke. . . . Past research into word retrieval and production has mostly focused on nouns. . . . More recently there has been increased interest in verb retrieval and verb processing disturbances [*ref*]. Unfortunately . . . One potentially useful speech pathology treatment for word production deficits involves the use of arm and hand gestures. . . . What remains to be developed is empirical evidence to support or refute the suggestion that gesture is a potent treatment for verb retrieval deficits. This paper presents evidence . . . (Rose & Sussmilch, 2008, pp. 692–693)

Explanation. The introduction is normally a discursive text that overviews the relevant literature and identifies the gaps in knowledge that the current study aims to address. Ideally the text is succinct and targeted to the main issues that frame the context of the study. The introduction should commence with what is known about the problem area and its interventions, what is yet to be understood, and how this study can address this gap.

Item 4—Aims: State the purpose/aims of the study, research question/s, and, if applicable, hypotheses.

Example.

> The purpose of the present study was to examine the effects of a written cueing treatment programme on verbal naming ability in two adults with aphasia. Treatment involved using a written cueing hierarchy, which was modelled after CART [Copy and Recall Treatment] and included verbal and writing components. (Wright, Marshall, Wilson, & Page, 2008, p. 524)

Explanation. The purpose and aims of the study need to be clearly described, normally at the conclusion of the Introduction. These should take the form of research questions that define the independent variable, the dependent variable, and report if a formal relationship was assessed. Statement of aims and, if applicable, hypotheses, provide the reader with explicit directions regarding the way in which the design, methods and results should be read, given that these should all follow from the aims/research questions being asked.

Section 3: Method—Design (including both design structure, as well as broader aspects of internal and external validity), Participant/s, Context, Approvals, Measures and Materials, Interventions, Analysis (Items 5–18)

Item 5—Design: Identify the design (e.g., withdrawal/reversal, multiple-baseline, alternating-treatments, changing-criterion, some combination thereof, or adaptive design) and describe the phases and phase sequence (whether determined a priori or data-driven) and, if applicable, criteria for phase change.

Example 1: Withdrawal/reversal design.

> An A-B-A-B single-subject design evaluated a token economy for increasing exercise in children with CF [cystic fibrosis] . . . Two advantages of this design are that it provides two evaluations of the treatment compared to baseline and it ends on a treatment phase, which is important from a clinical standpoint . . . The exercise diary data were used to determine when study phase changes were made. The specific criteria were the following: (a) there were three or more stable data points, (b) there was a predictable pattern in the data, or (c) there was no pattern, but the data points were predictably random. (Bernard, Cohen, & Moffett, 2009, p. 354–357)

Example 2: Multiple-baseline design.

> A randomized, concurrent, multiple-baseline single-case design was applied.[13] Participants completed repeated measurements during a baseline phase (phase A), an intervention period (phase B, 12 weeks) and a postintervention period (phase A′). Phase A acted as a control and was therefore compared with phases B and A′. (Hoogeboom et al., 2012, p. 2)

Example 3: Alternating-treatments design.

> The study used a multielement design with no baseline [*ref*] and a final "best treatment" phase [*ref*] to compare the effects of three contingent consequent events (Treatment A, adapted toys and devices; Treatment B,

> cause-and-effect commercial software; and Treatment C, instructor created video programs) on the frequency of stimulus activations. . . . Students received intervention in alternating treatments followed by the best treatment phase, in which only the most effective intervention was delivered. (Mechling, 2006, p. 98)

Example 4: Changing-criterion design.

> [This study evaluated] a prompting and shaping intervention, in the form of a changing criterion design . . . [to] assist this athlete in the technical development of his vault and corresponding height cleared . . . A photoelectric beam . . . was set across the runway at a height of 2.30 m (5 cm above the mean height obtained during baseline) . . . Intervention at this height continued until the participant displayed stability in his performance at a 90% level. Stability in performance referred to the successful repetition of three or more 90% performances. Following successful completion, intervention was also administered at the following heights: 2.35 . . . 2.52 m. (Scott, Scott, & Goldwater, 1997, pp. 573–574)

Explanation. In a broad sense, the design of a SCED encompasses all components of the methodology. For the purpose of this item, it specifically refers to the basic structure of phase sequencing and phase onset used in the investigation. That is, the design defines how the independent variable is manipulated, what the baseline conditions are, and when/how the independent variable is introduced/changed. Manipulation of these parameters allows the investigator to systematically control the independent variable in order to demonstrate the experimental effect. Moreover, using multiple phase changes in the design provides opportunity for repeated demonstration of the experimental effect on the same participant. This constitutes direct intrasubject replication (see Item 7 for intersubject replication and systematic replication).

The design of SCEDs is crucial for determining the adequacy of control of threats to internal validity and the experimental effect (Horner et al., 2005; Kratochwill et al., 2013). Design characteristics need to be clearly and specifically reported, otherwise it is difficult for the reader to (a) determine if the study has sufficient experimental control to adequately establish a functional cause-effect relationship between the dependent and independent variables and (b) evaluate the reliability of the results. In a small random sample of 20 reports using a single participant archived on the PsycBITE database, 45% of reports did not provide any information on the type of design used and an additional 20% incorrectly described the design (Tate et al., 2013b).

Specific reporting requirements will vary among the experimental designs. A major strength of SCEDs is their flexibility and adaptability. In addition to the "classic" designs described below, elements of these designs can be creatively combined in a multitude of ways depending on the research/clinical question being addressed and the intervention being considered (e.g., see Hayes, Barlow, & Nelson-Gray, 1999). A fundamental requirement in reporting SCEDs is that the basic structure of the design is explicitly and accurately described. This consists of reporting the following eight invariant features that apply to all designs:

Basic information required for all designs, including withdrawal/reversal (A-B-A-B) design.

i. The type of design (e.g., withdrawal/reversal)

ii. The number of phases (including baseline, experimental, maintenance and follow-up phases)

iii. The duration (length) of each phase

iv. The order in which the phases are sequenced (e.g., randomized, counterbalanced, data-driven)

v. The number of sessions in each phase

vi. The number of trials within each session in a phase (i.e., occasions when the dependent variable is being measured)

vii. The duration of sessions

viii. The time interval between sessions (see also Item 20)

Additional descriptive information needs to be provided for specific design types.

Additional information required for multiple-baseline designs.

i. The number of different (i.e., multiple) baselines (also referred to in the literature as data series, tiers, levels or legs) that the design contains. A graph alone is insufficient, because a graph represents the results that were obtained, rather than the design as planned, which may change in response to the intervention (see Item 6).

ii. Whether the baselines are across participants, behaviors or settings

iii. The method for determining treatment onset (e.g., response guided, randomization), or, if necessary, that there was no specific rationale or empirical basis

iv. Sometimes multiple-baseline designs also incorporate either a follow-up phase after the initial intervention phase or an "embedded" design (e.g., alternating-treatments). When such a variant is utilized, the complete sequence of phases should be clearly stated in the design description.

v. Whether or not the onset, and subsequent continuance, of data collection in each of the baselines occurred concurrently (i.e., at the same points in time) or nonconcurrently. If nonconcurrent, provide a rationale for this choice.

Additional information required for alternating-treatments designs.

i. Whether interventions were administered on the same day/session (e.g., Intervention 1 in the morning, Intervention 2 in the afternoon) or different days/sessions

ii. The way in which the order of the interventions was determined (e.g., randomized, counterbalanced, Latin square)

iii. The detailed phase sequence (e.g., inclusion of a baseline preceding the intervention; a final "best treatment" phase following the intervention)

Additional information required for changing-criterion designs.

i. All criteria or decision rules used to determine when a phase change occurs

ii. Whether the criteria are set a priori or are response guided

Additional information required for adaptive designs. In these designs, the structure of the investigation (e.g., phase sequence/duration, interventions, variations in intervention) is not fixed a priori, but depends, on an ongoing basis, on characteristics of the data (or responses) from early (or preceding) phases, Authors should also clearly describe the following:

i. Features or characteristics of the data (operationally defined) that are used to regulate the study design

ii. Which specific aspects of the study design, at each and all steps in the study, are determined by which specific features of the data

Item 6—Procedural changes: Describe any procedural changes that occurred during the course of the investigation after the start of the study.

Example.

BST [behavioral skills training] was implemented with each child individually in two 15- to 20-min sessions. . . . If the child did not obtain a score of 3 during an assessment after the initial BST sessions and two booster sessions, an additional assessment session was turned into a training session. . . . For Jake, who did not exhibit the correct behavior following BST or in situ training, an incentive phase was added. (Miltenberger et al., 2004, pp. 514–516)

Explanation. There are occasions when, for a variety of reasons, the proposed implementation of a study changes from the original plan. Changes may occur to any component of the study, including (a) methodological design; (b) setting in which the study takes place; (c) target behavior and any other outcome measures, with respect to content, method or frequency of administration; (d) equipment or materials to deliver the intervention; (e) use of practitioners and assessors; or (f) the intended intervention. Changes involving participant attrition or early termination of phases are addressed in Item 19.

The researcher may actively initiate changes that are a departure from protocol or they may be thrust upon the researcher as a result of external factors. With respect to changes in the intervention, one of the reported strengths of single-case methodology is the flexibility of implementation of the intervention (Connell & Thompson, 1986; Gravetter & Forzano, 2009). If adverse events occur or the intervention is not working sufficiently, then it is acceptable for the researcher to make alterations without necessarily compromising experimental control.

Authors need to report any changes or departure from the original plan or protocol, such as those listed above, along with reasons. They should also provide a statement about their impact on the interpretation of the results. Places to describe procedural changes in the report will depend on the type of change/s, but either the Method or Discussion sections are appropriate.

Item 7—Replication: Describe any planned replication.

Example 1: Systematic replication.

We employed a multiple-baseline design to test the efficacy of a recently developed approach for reducing school refusal behavior. . . . To maximize external validity, the intervention was tested using a systematic replication strategy, whereby only the major conceptual elements of the intervention were retained from previous applications. . . . (Chorpita, Albano, Heimberg, & Barlow, 1996, p. 281)

Example 2: Direct intersubject replication.

A single-subject ABAC design with replication across three participants was employed. (Leon et al., 2005, p. 96)

Explanation. Replication is a key feature of single-case methodology and is important because it has the capacity to inform whether and the extent to which an intervention is generalizable and hence is important for external validity. In spite of its central significance, replication is not commonly reported, with just over half of the studies (54%) in the neurorehabilitation survey of Tate et al. (2014) describing replication. In addition, although it may be evident that there is replication in a study, it is not always explicitly stated.

Three types of replication are described in the literature (Barlow et al., 2009; Gast, 2010a; Horner et al., 2005; Kratochwill et al., 2010, 2013; Sidman, 1960). Direct intrasubject replication refers to replication of the experimental effect within the design and addresses issues of internal validity (see Item 5). The other two types of replication are relevant to this item: (a) systematic replication (i.e., repeating the experiment with the same intervention but systematically changing characteristics of the individuals, setting, interventionists and/or behaviors) and (b) direct intersubject replication (i.e., repeating the same study but with additional individuals).

Using a series of replications, Horner et al. (2005) and Kratochwill et al. (2010, 2013) note that it is possible to provide a strong basis for causal inference. They have proposed criteria for the purpose of establishing evidence-based treatments: (a) a minimum of five methodologically strong research reports, (b) conducted by at least three different research teams at three different geographical locations, and (c) with the combined number of cases being at least 20.

Authors should clearly and specifically state the number and type of replications in the Abstract and the Method section. If the study is using systematic replication to build an evidence-base, authors may consider using the term *systematic replication* in the title. Authors need to clearly indicate whether the replication refers to (a) the replication of a previously published study or (b) intersubject replication within the current study.

Item 8—Randomization: State whether randomization was used, and if so, describe the randomization method and the elements of the study that were randomized.

Example 1: Randomized sequence.

Two treatments, one imitative and one cognitive-linguistic, were employed and treatment order was determined randomly. (Leon et al., 2005, p. 96)

Example 2: Randomized sequence (with restricted randomization).

A single case randomised experimental design with 12 phases was used. Each phase lasted for one week. During six treatment phases, participants wore the equipment and received cues (contingency electrical stimulation). During six no treatment phases, participants wore the equipment, but no cues were received. The phases were administered in a random order but always starting with a treatment phase, and no more than two consecutive phases were the same. (Wenman et al., 2003, p. 449)

Example 3: Randomized onset.

The start of the treatment phase was determined randomly for each participant, given the restriction that the baseline phase should last for at least 6 weeks (42 days) and at most 12 weeks (84 days) . . . This means that the treatment phase could start on any day between the 42nd and the 84th days, resulting in a total of 43 possible assignments. (ter Kuile et al., 2009, p. 151)

Explanation. The concept of randomization in SCEDs differs from its application in between-groups designs. In between-groups designs, randomization exclusively refers to allocation of participants to intervention groups (i.e., experimental vs. control). By contrast, in SCEDs, it refers to (a) the random sequencing of baseline and intervention phases, (b) the random determination of the commencement time for each phase, and (c) the combined randomization of both phase order and phase starting point (Kratochwill & Levin, 2010). If more than one participant is being studied in withdrawal/reversal designs, individuals can also be randomly assigned to intervention conditions. In multiple-baseline designs, random allocation of participants, behaviors or settings to each baseline of the design can also be implemented. Ferron and Levin (2014) and Kratochwill and Levin (2010) provide descriptions of these options.

The sequencing of baseline and intervention phases in SCEDs may be randomized using either simple (unrestricted) or blocked (restricted) randomization strategies (hence, randomized order de-

signs). The commencement point in time of each phase may also be randomized using simple randomization (hence, randomized start point designs). Randomization of both phase order and phase start point can also be combined in any given design (Kratochwill & Levin, 2010). Random allocation in SCEDs provides control of potential confounders related to time (Edgington, 1996; Kratochwill & Levin, 2010; Onghena & Edgington, 2005) which addresses at least two potential sources of experimental bias in SCEDs: history and maturation.

Researchers can also use randomization to assign specific stimulus items to different stimulus sets (e.g., treated and nontreated stimuli), although this type of randomization does not control for experimental bias related to history and maturation.

When randomization is used, authors should provide a reason for why it was used. They should also report specific details of (a) the basic randomization strategy (i.e., simple or restricted) and (b) those aspects of the design that were randomized (e.g., phase order, phase commencement, allocation of participants to interventions, allocation of participants, behaviors or settings to baselines). Authors need to describe any restrictions or modifications to the randomization process, which may be necessary for clinical or ethical reasons.

If randomization was not used, authors need provide a reason why it was not used. They also need to report any decision criteria used to determine phase sequencing (e.g., counterbalancing), time points for phase onset (e.g., data driven), allocation of participants to interventions (if applicable), and/or allocation of participants, behaviors or settings to tiers. If decision criteria are based on participant-related reasons (e.g., clinical considerations, severity of behavior, participant needs), these need to be reported.

Item 9—Blinding: State whether blinding/masking was used, and if so, describe who was blinded/masked.

Example 1: Blind assessor.

> The PQS [Psychotherapy Process Q-Sort] raters were not involved in any other aspect of the study procedures, and had no prior information regarding intended treatment approaches, design, or hypotheses. The raters were blind to the session number, treatment, and phase. Ratings of study tapes were made as part of PQS ratings of sessions from a larger sample [*ref*], and therefore were not rated consecutively or in comparison to one another. (Satir et al., 2011, p. 406)

No suitable examples of blinding of participant or practitioner in the SCED behavioral sciences literature were identified.

Explanation. Blinding (or masking) "refers to keeping trial participants, investigators (usually healthcare providers), or assessors (those collecting outcome data) unaware of an assigned intervention, so that they are not influenced by that knowledge" (Schulz & Grimes, 2002, p. 696). Lack of blinding or inadequate blinding in RCTs can inflate estimates of intervention effects by up to 17%, especially if trial outcomes involve subjective measures (Wood et al., 2008). It is not unreasonable to expect that this also applies to SCEDs. Blinding in SCEDs is reported very rarely, particularly for participants and practitioners (Tate et al., 2013b).

Blinding is difficult to achieve in nonpharmacological trials involving surgery, psychological interventions or rehabilitation (Bang, Ni, & Davis, 2004; Boutron, Tubach, Giraudeau, & Ravaud, 2004). Given that the majority of SCEDs involve nonpharmacological interventions, this is a pertinent issue. Although difficult, blinding of persons providing the intervention in nonpharmacological trials is not insurmountable (e.g., Edinger, Wohlgemuth, Radtke, Marsh, & Quillian, 2001). Boutron et al. (2007) provide a selection of strategies that can be used for blinding participants. By contrast, blinding/masking of assessors is usually feasible.

In reporting SCEDs, if blinding was not implemented, authors should state the reasons. When blinding was implemented, authors should clearly report who was blinded, and how the blinding was achieved. Going beyond basic reporting standards, authors may wish to consider reporting on procedures to assess the effectiveness of the blinding procedures and the outcome.

Item 10—Selection criteria: State the inclusion and exclusion criteria, if applicable, and the method of recruitment.

Example.

> We advertised the project in community newsletters and notices sent to local hospitals and rehabilitation professionals who typically worked with individuals with brain injuries. In the advertisements, we stated the following inclusion criteria: parents must have a documented brain injury, the children should be under the age of 10, and they must be demonstrating behavioral difficulties with the injured parent. Through subsequent observation of the parent and child, we determined whether the child was demonstrating a serious level of oppositionality with the parent (e.g., noncompliance to more than approximately 50% of parent-delivered requests). (Ducharme, Davidson, & Rushford, 2002, pp. 586–587)

Explanation. Readers of SCEDs need to know as much as possible about the participant(s), within the boundaries of anonymity, because, until generality has been demonstrated, results are only representative of the conditions under which the investigation was conducted and for the individual/s who participated.

In situations where participants were actively recruited into the study, inclusion and exclusion criteria should be provided. This information will assist with the identification of factors that may influence a participant's response to the intervention. It will also give an indication of the extent of the replicability and generalizability of research findings (see also Item 11). The description of the selection process should provide detail regarding who was recruited, and also the way in which the participants were recruited (such as newspaper advertisements, online recruitment targeting specific users, snowballing methods via other study participants or relevant professionals, distribution of brochures and leaflets).

Information should be provided about the way in which selection criteria were applied (e.g., by use of diagnostic instruments including questionnaires and interviews). Details of methods, instruments administered, and classification or assessment criteria need to be clearly defined. Readers need to be able to unambiguously identify how participant selection was accomplished for successful replication of research (Horner et al., 2005).

Method of recruitment may not be applicable in SCEDs in those situations where an individual presents with an issue that needs to be addressed. Such issues may be clinical (e.g., increasing a child's food intake) or nonclinical (e.g., improving an athlete's performance). In these circumstances, authors should instead state the reasons that the individual presented to the service and the reason for the intervention.

Item 11—Participant characteristics: For each participant, describe the demographic characteristics and clinical (or other) features relevant to the research question, such that anonymity is ensured.

Example.

> Three individuals with aphasia participated in the study. All had acquired aphasia secondary to a left hemisphere stroke. . . . See Table 1 for participant's demographic data. . . . Based on test performance and clinical judgment, all participants had nonfluent aphasia with good audi-

> tory comprehension. All participants exhibited significant word retrieval difficulties. Participants' performances on the BNT [Boston Naming Test] were reviewed . . . P1's word retrieval errors consisted of a mix of semantic (e.g., boat for canoe) and phonemic errors (e.g., fesmask for mask) . . . (Rider, Wright, Marshall, & Page, 2008, pp. 162–163)

Explanation. Inclusion of standard baseline participant characteristics, including demographic information and functional status, ensures that the reader understands the presentation of the participants and will be able to interpret the findings (Higginbotham & Bedrosian, 1995). It is also important for generalization (Barlow et al., 2009) and facilitates meta-analysis of multiple studies (Robey, Shultz, Crawford, & Sinner, 1999). In spite of their relevance and importance, the systematic review of Maggin, Chafouleas, et al. (2011) found that participant characteristics were incompletely reported, even for very basic demographic features such as sex (not reported in 33% of reports) and age (not reported in 42% of reports).

The description of the participants of a study should include basic demographic information such as age, sex, ethnicity, socioeconomic status, geographic location, as well as diagnoses where indicated, and functional or developmental abilities (Wolery & Ezell, 1993). Any diagnosis used should include instrumentation and scores. Lane, Wolery, Reichow, and Rogers (2007) also recommend that baseline or environmental factors which serve to influence or maintain the participant's behavior during the initial baseline should be evaluated and reported. These features will go beyond simple description of sociodemographic, medical and functional status variables.

Authors need to ensure that information provided does not lead to identification of the participant. This is a particular risk in the treatment of rare conditions. It is also important to protect an individual's privacy if the study involves stigmatized conditions.

The participant in a SCED typically, but not always, represents an individual person. The participant, however, can also be a group whose performance generates a single score per measurement period (such as the rate of a particular behavior performed by all students within a classroom during a set period of time). In that situation they are generally referred to as a "case" or "unit." It is important that the authors operationally define what constitutes the group and provide criteria for its selection (see Item 10). Authors need to specify the baseline characteristics (to the same level of detail) for each participant (Wolery & Ezell, 1993) to allow the reader to ascertain the extent to which the results are generalizable (see also Item 24).

Item 12—Setting: Describe characteristics of the setting and location where the study was conducted.

Example.

> The children and teacher comprised the full membership of a third grade general education classroom in a rural postindustrial Northeast elementary school. . . . All observations were conducted in their homeroom class during the SSR [sustained silent reading] period . . . During SSR, students sat at their desks, which were arranged in two rows of desks facing toward the teacher in the front of the room. (Methe & Hintze, 2003, p. 618)

Explanation. It is critical to report the setting and location of the study because these factors have implications for the generalizability and applicability of the findings (see Item 24). The context of the intervention will vary according to whether it is provided in primary, secondary or tertiary health care, classroom/educational facility or community settings. The location of the study will also vary according to whether the intervention is offered in urban, rural or remote locations, and this will have an impact on the service delivery model.

The setting is of particular interest for the reporting of SCEDs for two reasons. First, it may be an inherent a priori feature of the design to introduce the independent variable across a range of settings in a controlled manner. Multiple-baseline designs across settings are a case in point. Second, the detailed description of the relevant location and setting is central to the replicability of the study (see also Item 7). An independent researcher may be interested in varying the location of the study as one step toward systematic replication. In this scenario, a sufficiently detailed description of the setting is requisite information.

Authors need to provide detailed information on the location of the study, including the number and type of settings, as well the practitioners or providers involved. There should be sufficient detail regarding the location and setting of an intervention to enable others to evaluate how different this is from their own situation.

Item 13—Ethics: State whether ethics approval was obtained and indicate if and how informed consent and/or assent were obtained.

Example.

> Approval for this research study was obtained from the human subjects research Internal Review Board at participating institutions. . . . If a woman was interested, she was given a pamphlet with information about the intervention and proposed research project before leaving the hospital, or at a follow-up appointment. Women then contacted the primary investigator (SMB) for more information and/or to schedule an initial assessment appointment. An informed consent form was read, discussed and signed before beginning the initial assessment and a copy of this consent form was given to the participant for their records. (Bennett, Ehrenreich-May, Litz, Boisseau, & Barlow, 2012, p. 166)

Explanation. It is virtually a universal requirement that research involving human participants requires review by and approval from an Institutional Ethics Committee. If the SCED was implemented as part of clinical care, ethics approval might not be required, as is the case in *N*-of-1 trials (Punja, Eslick, Duan, Vohra, & the DEcIDE Methods Center *N*-of-1 Guidance Panel, 2014). Reporting whether ethics approval has been obtained is not, however, a feature of all research reporting guidelines. For example, the CONSORT Statement (Moher et al., 2010a) does not include this as a checklist item. Following the CENT 2015 guidelines (Shamseer et al., 2015; Vohra et al., 2015), the SCRIBE 2016 also includes the reporting of ethical approval as a checklist item for reporting SCEDs.

Written informed consent should always be secured (Mechling & Gast, 2010). There may be instances when the participant cannot provide informed consent (e.g., if the participant is a minor or otherwise legally unable to provide informed consent). In this situation, their assent to participate should be sought, and consent also obtained from legal guardians or parents. It is not sufficient to merely state informed consent/assent was obtained. Rather, the process by which consent/assent occurred needs to be described so that it is clear who provided informed consent/assent. This is particularly important with vulnerable populations or where limited disclosure is necessary (National Health and Medical Research Council, 2009).

Item 14—Measures: Operationally define all target behaviors and outcome measures, describe reliability and validity, state how they were selected, and how and when they were measured.

Example 1: Operational definition of the target behaviors and how they were measured.

> Topographies of targeted behaviors for Bob included self-injury . . . Self-injury consisted of face slapping, defined as a forceful contact between an open palm and cheek. . . . Bob also displayed spitting, defined as spittle landing within 1 foot of another person. . . .
>
> The primary dependent variable was the percent of intervals in which maladaptive behaviors were observed during each 10-min session. Data

were collected using paper and pencil during consecutive 10-s intervals cued by an audio tape throughout the entire session, resulting in a total of 60 consecutive intervals. Partial interval recording was used: during each interval observers recorded whether or not any of the target maladaptive behaviors had been observed for any portion of the interval.

Observers underwent training in behavioral observation . . . and demonstrated mastery prior to participating in the study. (Treadwell & Page, 1996, pp. 65–66)

Example 2: Reliability of dependent variables when using nonstandardized measures.

Three different categories of interobserver agreement were calculated, as follows. . . . *Parent-therapist agreement:* This category involved agreement between compliance data coded by the parent and those coded live by the research therapist. In this category, interobserver agreement was obtained on 39% of sessions conducted by parents, randomly selected from each of the phases across all children. Overall agreement averaged 92% for baseline (range 82%–100%) and 98% for treatment, generalization, and follow-up sessions (range 94%–100%). (Ducharme et al., 2002, p. 588)

Explanation. A single item in the SCRIBE 2016 covers all aspects of the dependent variable/s: the what, how, and when of measuring the effect of the intervention. As with other research designs, the measurement process in SCEDs also needs to be valid and reliable. Validity is enhanced by selecting dependent variables that are (a) relevant to the behavior in question and that best match the intervention, as well as (b) accurate in their measurement, which is facilitated when the behavior that is targeted for intervention is operationally defined. Reliability is enhanced when the behavior is measured in a manner that yields consistent results.

What is measured? SCEDs commonly use a variety of dependent variables that play specific roles in the experiment. The primary outcome variable in SCED methodology is referred to as the target behavior. Target behaviors have three defining features in order to enhance quality of the study and minimize bias: they are specific, observable and replicable (Barlow et al., 2009). Other dependent variables frequently used in SCEDs may be considered akin to secondary outcome variables: Generalization measures are increasingly recognized for their important role in contributing to the external validity of the study (see also Item 24). In addition, SCEDs have a strong tradition in promoting experiments that address socially relevant behaviors and interventions. Additional measures are often incorporated into a SCED to specifically measure social and ecological validity.

Authors need to provide operational definitions of the target behavior, which should be objective, clear and complete (Kazdin, 2011), in order to convey what does and does not constitute an instance of the dependent variable. In studies where there is more than one target behavior, the report should clearly identify and describe each of the target behaviors in detail. Other dependent variables used in the study (e.g., generalization measures, social validity measures) should also be described with equal clarity and precision.

How is it measured? The "how" of measurement covers measurement procedures, including who selected and measured the target behaviors and other outcome variables, along with their training in the assessment procedures. Authors also need to provide information on the way in which the dependent variables were measured, justification for the selection of those measures, and detail regarding what constitutes a correct or incorrect response.

Because the target behaviors are highly specific to the presenting case in SCEDs, formal psychometric evaluation of the measures will generally not have been established. It is therefore recommended practice that evaluation of interobserver agreement on the target behavior is conducted and reported. When standardized instruments are used, it is essential that psychometric details regarding reliability, validity and responsiveness of the instruments are reported. Any equipment used to measure the dependent variable/s should also have established measurement properties, which, if available, should be provided in the report.

When is it measured? A distinctive feature of SCEDs, in contrast to group methodology, is that the target behavior (primary outcome variable) is measured repeatedly and frequently throughout all phases, including the baseline and intervention phases. Smith (2012, p. 519) observes that "the baseline measurement represents one of the most crucial design elements of the SCED." In spite of this, baseline data were not available for 22% of SCEDs in Smith's systematic review. Moreover, in other phases of the study, the number of data points could not be readily identified.

Target behavior/s need to be selected so that they are suitable for repeated and frequent measurement. Previously, the recommended minimum number of data points per phase was three (Barlow & Hersen, 1984; Beeson & Robey, 2006), but more recently professional guidelines recommend a minimum of five data points per phase (Horner et al., 2005; Kratochwill et al., 2010, 2013). There should be a clear description of the number of sessions in which the target behavior is measured in each phase, as well as the number of times it is measured in each session (i.e., number of trials per session). Frequency of measurement of other outcome variables depends on their role. Recommended practice is that generalization measures are probed continuously throughout all phases (Schlosser & Braun, 1994). By contrast, evaluation of social validity can only logically occur after the intervention has taken place. As with target behaviors, authors should clearly state the frequency and regularity of measurement of all other outcome measures, including phases during which such measures were taken.

Item 15—Equipment: Clearly describe any equipment and/or materials (e.g., technological aids, biofeedback, computer programs, intervention manuals or other material resources) used to measure target behavior/s and other outcome/s or deliver the interventions.

Example 1: Software.

Training in use of email interface: "Participants 1–4 used a mouse connected via USB port to activate the e-mail program. Participant 5 used a trackball instead of a mouse to accommodate his motor impairment. . . . The instructor used a number pad . . . to control presentation. . . . The program was run on . . . a laptop. . . . An altered interface was also developed to assess generalization to a slightly different platform. It included additional buttons . . . as well as rearrangement of existing buttons to novel positions." (Ehlhardt, Sohlberg, Glang & Albin, 2005, p. 571; p. 576)

Example 2: Materials.

Photo Cue Cards [*ref*] were used as instructional stimuli. Four target photographs and four control photographs were selected for each child. Target stimuli are presented by child and instructional condition in Table II. A stopwatch was used to time the length of experimental sessions. (Holcombe, Wolery, & Snyder, 1994, p. 53)

Explanation. Target behaviors may be measured using behavioral observations and standardized scales (see Item 14) or, alternatively, equipment. Similarly, many interventions used in the behavioral sciences are accompanied by materials and equipment. Complete and accurate reporting of equipment and materials is central to the issues of replicability (see Item 7) and generalizability (see Item 24).

A detailed description of the independent variable (i.e., the intervention) will include not only a description of the elements of the intervention (see Item 16), but also any specific equipment used, along with the way in which the equipment operates. Such equipment will include training manuals, computer programs, bio-

feedback techniques or any other materials required to implement the intervention.

When equipment is used to measure the dependent variable, the way in which it operates needs to be described, as well as its calibration. In addition, its measurement properties, if available, should be reported (see Item 14).

Item 16—Intervention: Describe the intervention and control condition in each phase, including how and when they were actually administered, with as much detail as possible to facilitate attempts at replication.

Example.

> Check-in sessions were scheduled once a week during baseline to collect paperwork, monitor participant functioning, and to serve as an active waitlist control condition. Following the baseline period, women entered the intervention phase, which consisted of eight weekly sessions lasting approximately 60 min. The therapist was the lead investigator of this study (SMB) who was a senior graduate student at the time of data collection. . . . Women were told they could opt to include their partners in treatment sessions if they wished . . . (Bennett et al., 2012, p. 166)

This description is accompanied by a detailed table (p. 165) describing the session, primary goal of each session, and brief description of session content and homework.

Explanation. Evidence of inadequate description of the intervention abounds in the broader health-intervention research literature using group methodology (e.g., Boutron et al., 2008; Dijkers et al., 2002; Glasziou, Meats, Heneghan, & Shepperd, 2008). The importance of describing the independent variable itself in sufficient detail to allow replication has been emphasized in the general SCED literature.

The independent variable needs to be operationally defined, similar to the descriptions of the target behavior/s and the outcome measures (see Item 14). In situations where more than one intervention is used (e.g., A-B-A-C-A-D, or alternating-treatments design) each intervention should be described. In addition to ancillary aspects of the intervention (i.e., materials, manuals, stimulus items, equipment, software programs or applications; see Item 15), the nature of the intervention per se needs to be clearly specified. This description includes information on who delivered the intervention and in which mode, whether it be individual, group, distance, carer- or educator-focused, or whether telehealth or other technologies were used. It is critical to report the exact number, duration and frequency of the intervention sessions (Baker, 2012; Warren, Fey, & Yoder, 2007). Specifically, intervention intensity or dosage should be described in terms of the following: (a) dose form (i.e., the typical task or activity being used), (b) the dose (i.e., the number of times an active ingredient or teaching episode occurs per session), (c) the dose frequency (i.e., the number of intervention sessions per unit of time), (d) the total intervention duration (i.e., the total period of time in which an intervention is provided), and (e) the cumulative intervention intensity (i.e., the product of dose by dose frequency by total intervention duration; Warren et al., 2007).

SCEDs typically compare the effect of an independent variable with either a control condition or another independent variable (i.e., another intervention). The control condition, often called the baseline condition in SCEDs, represents a period of time when the participant's target behavior is recorded repeatedly before the intervention is introduced. The same degree of specificity that is used to describe the experimental intervention should also be used for the control condition.

Item 17—Procedural fidelity: Describe how procedural fidelity was evaluated in each phase.

Example.

> Adherence to treatment procedures was accomplished by use of a manual that described the steps of the programme. . . . In addition, an independent, trained observer watched videotapes of one training session from each step of the training programme (25% of sessions), and tallied the opportunities for the experimenter behaviours and the number of times the experimenter used the expected behaviour (e.g., prompts, models, reinforcement). Adherence to the protocol was calculated using the formula (EA × 100)/ET, where EA = the experimenter behaviours. . . . Procedural reliability was 95% to 100% for each step. (Hickey, Bourgeois, & Olswang, 2004, p. 630)

Explanation. Wolery (1994) describes the collection of procedural fidelity data as having three main functions: (a) to monitor the occurrence of relevant variables, (b) to provide documentation that the experimental conditions occurred as planned, and (c) to provide information to practitioners about the use of the interventions. In spite of the seriousness of unreliable implementation of the experimental conditions, fidelity checks in SCEDs in the behavioral sciences are infrequent (27% in the series of Didden et al., 2006). This result is comparable to the data on reporting quantitative measurement of the independent variable in three reviews of the contents of *Journal of Applied Behavior Analysis* (Hagermoser Sanetti & Kratochwill, 2007–2008).

In recognition of the critical importance of the fidelity of the implementation of study protocols, an item addressing adherence to the protocol was introduced for the CONSORT Extension to Nonpharmacological Treatments (Boutron et al., 2008). Application of a prepared checklist or steps of a protocol is the best way to document procedural fidelity (see Borrelli et al., 2005; Dane & Schneider, 1998; Gast, 2010b, pp. 99–101). Steps taken to evaluate procedural fidelity should be reported by authors, including how it was measured and the results of its assessment. If procedural fidelity is evaluated during the course of the study, and found to be suboptimal, authors may consider going beyond basic reporting standards to describe whether and what steps were taken to improve procedural fidelity (Hagermoser Sanetti & Kratochwill, 2014).

Item 18—Analysis: Describe and justify all methods used to analyze data.

Example 1: Use of visual analysis.

> The split-middle technique [*ref*] was employed to detect changes in the number of successful shots within phases and resultant trend lines (Barlow & Hersen, 1984). White proposed that level, slope, and mean score of the celeration line (or trend) line be assessed as three descriptive analyses for conclusions. . . . Given that a point on the celeration line does not actually explain the performance level, for brevity we have chosen to concentrate on the slope of the celeration line. (Mesagno, Marchant, & Morris, 2009, p. 136)

Example 2: Use of statistical analysis.

> As serial dependence . . . can bias the visual inspection,[17] we checked our data in each phase for serial dependence using the lag-1 method.[12] If data were found to be significantly correlated, we transformed the data using a moving average transformation, in which the preceding and succeeding measurements were taken into account.[12,16] In addition, randomisation tests for multiple-baseline single-case designs were carried out. We expected phases B and A9 to be superior to phase A in terms of our health outcome assessment. Therefore, we tested the null hypothesis that there would be no differential effect for any of the measurement times using a randomisation test of the differences in the means between the preintervention phase and the intervention or postintervention phase.[17] A p value of <0.05 was considered statistically significant. For the premeasurements and postmeasurements, we considered change scores of 20% on validated questionnaires as clinically relevant.[32] We used Stata/IC 10.1 for Windows for the descriptive and visual analysis of the data and R version 2.14.1 for the randomisation tests.[31] (Hoogeboom et al., 2012)

Explanation. Both visual and statistical techniques can be used to analyze SCED data (see Appendix). They are considered com-

plementary rather than mutually exclusive (Maggin & Odom, 2014; Parker & Brossart, 2003; White, Rusch, Kazdin, & Hartmann, 1989) and should, arguably, be used in combination (Davis et al., 2013; Smith, 2012).

Visual analysis relies upon visual inspection of the graphed data to draw conclusions regarding the reliability and consistency of intervention effects (Lane & Gast, 2014). In the past, experts have argued that visual analysis is the most sensitive and appropriate way to detect intervention effects in SCEDs (e.g., Parsonson & Baer, 1986). It has been the traditionally preferred and most frequently used approach (Busk & Marascuilo, 1992), but has significant limitations (for discussion, see Lane & Gast, 2014; Smith, 2012). Authorities have proposed guidelines for systematizing visual analysis (e.g., Kratochwill et al., 2010; Lane & Gast, 2014), but there is not yet complete agreement about decision-making criteria to guide the process.

Statistical analyses have advantages in that they (a) use an explicit set of operational rules and replicable methods, (b) provide a direct test of the null hypothesis, (c) utilize precisely defined criteria for significance, (d) are useful when there is instability in the baseline or treatment effects are not well understood, and (e) can help to control for extraneous factors (e.g., Kazdin, 1982a, 1982b). There is no "universal gold-standard," however, and authors should choose the analytic method for SCEDs that is guided by a number of considerations: (a) design requirements and data assumptions, (b) the research question being posed, (c) features of the data being analyzed, and (d) the interpretability, ease of computation and proven validity of the technique for analysis of SCED data. For detailed discussion of these issues, see Manolov, Gast, Perdices, and Evans (2014) and Shadish (2014). If statistical analyses are used, authors need to report whether underlying assumptions and other requirements pertinent to the technique were evaluated for the data set being analyzed. This information will allow the reader to determine the suitability of the analytic methods used.

It is critically important that authors fully and clearly describe the method/s of analysis used, regardless of whether they select visual, statistical or both techniques. Authors should state if the method/s of analysis was prespecified before commencement of data collection, and those changes (if any) that were subsequently made as a consequence of limiting/problematic features of the data. The rationale for selecting the analytic technique in terms of its appropriateness to the study design and the research question should be clearly stated, and the source reference describing the technique should be cited. In the case of visual analysis, it is important to clearly report the features of the data that were selected for analysis (along with reasons) and whether a systematic protocol was used and, if so, which one.

Section 4: Results—Sequence Completed, Outcomes and Estimation, Adverse Events (Items 19–21)

Item 19—Sequence completed: For each participant, report the sequence actually completed, including the number of trials for each session for each case. For participant/s who did not complete, state when they stopped and the reasons.

Example: Deviation from protocol: Interruption of treatment.

> Ms. O had 24 sessions of psychotherapy over a one hundred and 47 day period, which included two periods of treatment interruption; once in the middle of the BCT [Behavior Change Treatment] phase, and once at the end of the BCT phase. Ms. O was randomly assigned to receive AFT [Alliance Focused Treatment] for the first 4-week therapy phase, BCT for the second 4-week therapy phase, and AFT for the last 4-week therapy phase. The two treatment interruptions coincided with the Thanksgiving and Christmas holidays, and occurred during BCT only. (Satir et al., 2011, p. 406)

Explanation. A major benefit of SCEDs lies in their flexibility. As described in Item 6, it is possible during the conduct of the investigation to change and fine-tune procedures and interventions that do not appear to be working. Any such deviation from the original plan of the study, however, needs to be clearly reported (see also Item 25). The present item pertains specifically to the report of procedural variations that reflect changes in any of phase sequence, phase order, and/or number of sessions actually completed by each participant.

Changes to phase sequence may occur in response to clinical or ethical concerns, such as subsequently deciding to commence the investigation with a treatment phase rather than the predetermined randomized sequence. For similar reasons, a phase might be prematurely terminated, interrupted, extended, or substituted for reasons of lack of efficacy, harms, periods of absence, and so forth. Post hoc changes to the randomization schedule or the intended duration or structure of phases can significantly weaken the internal validity of the study and therefore need to be reported.

Missing data due to attrition may bias the results, especially if this reflects intentional or systematic noncompliance (Smith, 2012). Moreover, if participants do not complete all phases as planned, there may also be insufficient phase repetitions to demonstrate adequate experimental control. Attrition of participants within units (such as classrooms) over time may confound intervention effects especially if attrition is nonrandom and associated with the intervention itself.

In order to minimize such bias, Horner et al. (2005) argue that, regardless of attrition or missing data, results of any participant for whom there is data for both a baseline and an intervention phase should be reported. If techniques for dealing with missing data are used (e.g., retrospective data completion, such as completing diaries), this also needs to be reported, given that such techniques can introduce significant bias in the data due to incorrect recollection/recall (Bolger, Davis, & Rafaeli, 2003). Accordingly, any changes to the sequencing of phases or missing data should be reported, along with their reasons so that the reader can evaluate the integrity of the results and their interpretation.

Item 20—Outcomes and estimation: For each participant, report results, including raw data, for each target behavior and other outcome/s.

Example 1: Provide raw data (Figure 2).

Example 2: Statistical analysis.

> The three measures of treatment gains are shown in Table 2. Four patients . . . demonstrated significant treatment gains as determined by all three measures (*C* statistic, effect size, modified CDC [conservative dual criteria]). One patient . . . did not demonstrate significant gains on any measure. The patients who improved made variable gains in other areas as indicated by a significant increase in their WAB [Western Aphasia Battery] AQs [Aphasia Quotient] (mean increase = 5.58, $SD = 2.32$, $t = 4.81$, $df = 3$, $p < .05$). (Crosson et al., 2009)

Explanation. Traditionally, SCED results are reported in graphical form which is, arguably, the clearest and most unambiguous way of depicting the major features of the raw data. At minimum, SCED data for each session should be reported in graphic form. This does not preclude the additional reporting of raw data for each session (or for each trial within a session) for each phase of the study in a tabulated format. Although tabular presentation is acceptable (and even desirable for verification in meta-analysis), relevant features (e.g., consistency across similar phases) may not be directly obvious in this format. Alternatively, authors may provide information about where the raw numerical

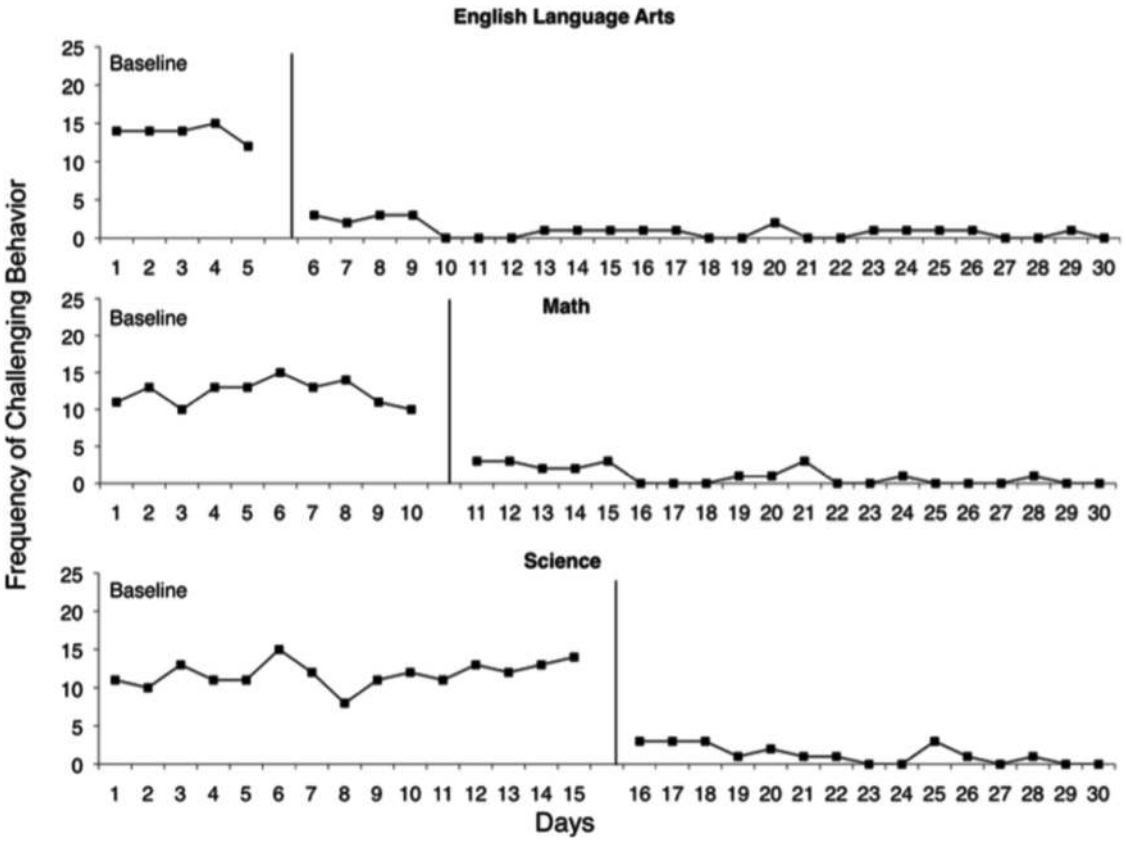

Figure 2. "Jason's frequency of challenging behaviors, across settings." Reprinted from "Structured Flexibility: The Use of Context-Sensitive Self-Regulatory Scripts to Support Young Persons With Acquired Brain Injury and Behavioral Difficulties," by T. J. Feeney, 2010, *Journal of Head Trauma Rehabilitation, 25*, p. 419. Copyright 2015 by Wolters Kluwer Health. Reprinted with permission.

data set can be accessed. Raw data should be reported in the results section for each measurement point/session in each phase of the study for each participant, setting and target behavior/s. Even though aggregation of data (e.g., averaging results over several sessions), may provide a clearer view of the apparent intervention effect, it may also mask or misrepresent various important features. Such features may include (a) stability of the initial baseline phase, (b) variability and trends within a phase, (c) degree of consistency between similar phases (e.g., intervention phases), (d) the degree of overlap between baseline and intervention phases, (e) magnitude of effect latency following intervention phase onset. Readers need to be able to critically evaluate all these aspects of the data. They also need to be able to draw their own conclusions about how adequately the investigators have taken any anomalies into account when appraising the clinical value of the intervention (Barlow et al., 2009).

The metric used on the horizontal axis of graphed data should be in units of real time (i.e., days, weeks, etc.) rather than session number. This information allows the reader to more clearly interpret and critically evaluate the results of the study. Providing an exact chronology of the time interval between sessions allows the reader to accurately evaluate patterns of consistency between similar phases and effect latency following intervention onset. Carr (2005) argues that using a real-time metric on the horizontal axis is particularly relevant in multiple-baseline designs. The reason is so that the reader can determine the order in which sessions across participants, settings or behaviors were conducted relative to each other. More importantly, it also allows the reader to see how many sessions occurred in the initial A-phase of the second baseline (or third, fourth, etc.) of the design after the intervention was introduced in the first baseline.

Results of any statistical analyses for the target behavior/s and other relevant outcome variables also need to be reported. This report should be done for each participant and setting in the study. Irrespective of the analysis used, authors need to clearly state which phases and features of the data were compared. Any changes or deviations in a preplanned analysis strategy should also be reported, as well as the reasons that it was necessary to make such changes. For statistical analyses, the value of the calculated statistic/s, standard errors, and any associated probability level should also be reported.

Item 21—Adverse events: State whether or not any adverse events occurred for any participant and the phase in which they occurred. The current reporting of adverse events is rare in SCEDs where behavioral interventions have been applied. Only a single example (Hoogeboom et al., 2012) was identified where authors mentioned that adverse events were monitored, but no information was provided on the way in which adverse effects were measured, nor were results provided.

Explanation. The reporting of adverse events or harms is rare in SCEDs of behavioral interventions. Their report is a more common feature in medical interventions in randomized and non-

randomized trials and observational studies (Golder, Loke, & Bland, 2011), even though it is often inadequate (Papanikolaou, Christidi, & Ioannidis, 2006; Vandenbroucke, 2006). This problem is despite the reporting of harms being a specific checklist item in the CONSORT 2010 Statement (Schulz, Altman, & Moher, 2010).

Although there are no clear definitions of *harms*, the CONSORT Extension for Better Reporting of Harms in RCTs (Ioannidis et al., 2004) has recommended *harms* as the preferred term to describe an adverse event (as opposed to describing the "safety" of a treatment). Particular recommendations include that (a) authors make specific mention of harm in the title or abstract, as well as the introduction where harms are a primary outcome measure; and (b) there is specification of the harms in reporting of results, with special attention paid to discontinuations and withdrawals due to adverse events. The guideline also recommends that authors should (a) present the absolute risk of each adverse event (specifying type, grade, and seriousness per arm of the trial), (b) describe any subgroup analysis and exploratory analysis for harms, and (c) provide a balanced discussion of benefits and harms with emphasis on study limitations, generalizability and other sources of information on harms (Ioannidis et al., 2004). The guide can be adapted to accommodate SCED methodology.

The SCRIBE 2016 recommends that authors make full and explicit disclosure of any harms or adverse events that occurred to any participant during the course of the SCED trial, including the absence of these events. Loke and colleagues (2007) suggest a framework to enable a systematic, manageable and clinically useful way to define adverse effects. It includes a predefined classification of adverse effects as diagnosed by the clinician, by test results, or by participant-reported symptoms (e.g., pain).

Section 5: Discussion—Interpretation, Limitations, Applicability (Items 22–24)

Item 22—Interpretation: Summarize findings and interpret the results in the context of current evidence.

Example.

> Results showed that the repeated reading program combining several research-based components . . . improved fluency on second-grade transfer passages for the three participants lending support to the existing literature on repeated reading [*ref*]. . . . With research on repeated reading spanning decades and numerous studies demonstrating successful outcomes . . . this practice holds great promise as a strategy for improving reading fluency. However, as suggested by [the metaanalysis of] Chard and colleagues (2009), the current research literature on repeated reading is not sufficient for it to be designated as an evidence-based practice. (Lo, Cooke, & Starling, 2011, pp. 133, 136)

Explanation. An early section of the discussion needs to provide a clear and concise summary of the findings of the study, including the strength of the intervention effect and the clinical importance of the findings. The results should be interpreted in terms that are specific to the study, as well as more generally with reference to the current literature. Interpretation specific to the study will benefit from taking into account the aims of the study, along with the robustness of methodology and procedures. Item 23 on limitations of the study is a separate item in the SCRIBE 2016 Statement, but has obvious relevance to the item on interpretation.

Vandenbroucke and colleagues (2007) note that overinterpretation is a common problem in observational studies. For the STROBE Statement, they advocate that caution is exercised in interpreting the findings of a study. A cautious approach to interpretation is particularly pertinent to SCEDs in instances where findings are unclear or adequate replication has not occurred (see Item 7).

In terms of interpreting the study findings in the context of the current evidence from the literature, the CONSORT Statement (Schulz et al., 2010) suggests that results of clinical trials are interpreted with respect to the knowledge base, as synthesized in systematic reviews. In some research areas using SCEDs, meta-analyses are available and it is recommended that information from these be incorporated when available (as in the above example of Lo et al., 2011, referring to the meta-analysis of Chard, Ketterlin-Geller, Baker, Doabler, & Apichatabutra, 2009).

Item 23—Limitations: Discuss limitations, addressing sources of potential bias and imprecision.

Example.

> The validity of comparisons between AFT [Alliance Focused Treatment] and BCT [Behavioral Change Treatment] was compromised by several factors, including the significant intervention interruptions during BCT, the administration of psychotropic medication during the study. . . . Furthermore, there may have been significant carry-over effects from one phase to another, as skills learned during one phase could not be "unlearned" . . . Limitations to measurement include possible limitation to the accuracy of the self-report nature of the kilocalorie intake, which was not verified by other sources of data collection (e.g., independent observation). (Satir et al., 2011, p. 417)

Explanation. It is often tempting for authors to focus on the positive findings of their results and to glide over the flaws in their study. However, all studies have limitations that can either bias or confound results. Important limitations reflect threats to the validity of the study that introduce potential bias in the findings.

There are many potential limitations in SCED studies that compromise the extent to which results are unbiased, reliable and likely to generalize (e.g., see Horner et al., 2005; Tate et al., 2013b; Wolery, Dunlap & Ledford, 2011). These limitations include the following: (a) poor matching of the design to the type of intervention (see Item 5); (b) inadequate replication (see Items 7 and 24); (c) absence of blinding (see Item 9); (d) lack of randomization (see Item 8); (e) imprecision with respect to the description of the participant's functional abilities (and, where applicable, diagnosis), making it difficult for readers to generalize to other cases (see Items 10 and 11); (f) problems with the operational definition of the target behavior or assessment of its reliability, insufficient number of data points in some or all of the phases to meet minimum standards (see Item 14); (g) absence of information regarding procedural fidelity (see Item 17); and (h) reliance on visual analysis for ambiguous cases, insufficient number of data points required for specific statistical analyses (e.g., the *C* statistic requires at least eight data points per phase), use of statistical procedures that do not deal adequately with extant features of the data (e.g., trend, variability or auto-correlation), or whose underlying assumptions are not met (see Item 18).

In terms of adequate reporting, authors need to provide a systematic discussion of specific limitations associated with their findings (as described above) which is also contextualized within the relevant literature. In this way the reader is provided with a realistic, critical appraisal of the contribution that the study makes to the field and the extent to which the results reflect a strong finding that is likely to be repeated.

Item 24—Applicability: Discuss applicability and implications of the study findings.

Example.

> The results of the present study replicate the findings of [ref] with regard to the effect of using DRA [differential reinforcement of alternative

behavior], nonremoval of the fork, and stimulus fading to increase variety of food intake. The study extends previous findings by showing that the intervention package was effective independent of who fed the child. The effect of our treatment on John's consumption of nonpreferred foods did not generalize across settings in the absence of intervention in the home, but multisetting training led to transfer across settings and caregivers. Our study also shows that the treatment package described by [ref] was effective during typically occurring mealtimes with regularly scheduled food types, and that the treatment was effective for increasing the number and variety of originally nonpreferred foods. (Valdimarsdottir, Halldorsdottir, & SigurÐardóttir, 2010, p. 105)

Explanation. The concept of applicability or generality is based on the assumption that inferences can be drawn from the condition in which an intervention effect was demonstrated, to other conditions based on known similarities and differences between these conditions (Gast, 2010a).

The reader should be provided with a discussion of implications of (a) conceptual/theoretical considerations, (b) clinical/practical considerations, and (c) methodological considerations, which impact on the generalizability of the findings. Conclusions could be drawn, for example, from information provided about replication (Item 7), inclusion and exclusion criteria (Item 10), participant characteristics (Item 11), and generalization measures (Item 14).

The replication process (see also Item 7) involves an increasing number of variations to the different dimensions (most importantly participants, setting and practitioner) that can be changed with every replication (Barlow et al., 2009; Sidman, 1960). Each replication will add information regarding the generalizability of the findings. The stage of the replication process (direct or systematic; see, e.g., Gast, 2010a) should be made clear. Similarities and differences between the current and previous studies need to be explicated.

Reference should be made to factors that determined baseline performance, such as described in Items 10 and 11, because these factors are hypothesized to influence relations between the independent variable and the dependent variable in a very specific way (Horner et al., 2005; Wolery, et al. 2011). If outcome measures additional to the target behavior (Item 14) are used for the purpose of generalization, authors should discuss the evidence to support (or not) a functional relationship between the independent variable and the generalization variable in the context of a theoretical framework, if available. If responses to the intervention differed across participants (or between studies), reasons and proposed causal relationships of this finding should be discussed and new propositions made to explain the finding.

Section 6: Documentation—Protocol, Funding (Items 25–26)

Item 25—Protocol: If available, state where a study protocol can be accessed. No published SCED studies were identified that contained information on how to access a protocol. A number of published reports indicated that the research protocol of a SCED was reviewed by an institutional research committee, but this pertains to Item 13. Several SCED protocols were identified in trial registries (Lloyd at www.anzctr.org.au, Trial ID ACTRN12611000812998; Pool at www.anzctr.org.au, Trial ID ACTRN12611000531910; Wambaugh, Mauszycki, Cameron, Wright, & Nessler, 2013; Wambaugh, Nessler, C& Wright, 2013, at www.clinicaltrials.gov) but personal communication with the authors indicated that the studies were not yet published (Lloyd & Sherrington, 2011) or the published report did not make reference to the protocol (Pool, Blackmore, Bear, & Valentine, 2014; Wambaugh, Mauszycki, et al., 2013; Wambaugh, Nessler, et al., 2013).

Explanation. Trial protocol availability was a new item introduced for the CONSORT 2010 Statement (Schulz et al., 2010) and is also included in the CENT 2015 Statement (Shamseer et al., 2015; Vohra et al., 2015). Moher et al. (2010a, p. 21) indicate that the protocol refers to the planned methods of the "complete trial (rather than a protocol of a specific procedure within a trial)." The rationale for including the protocol item in the CONSORT 2010 Statement was based on published evidence of the discrepancy between the trial protocol and the subsequent published report (e.g., selective reporting of results, post hoc change in the main outcome measure). Such discrepancies continue to be documented (Dwan et al., 2011).

Having a protocol readily accessible makes authors accountable for any changes that are made to the planned research design and analysis. As noted, however, a special feature of SCED methodology is its flexibility in terms of modifying the design and/or intervention after the trial has commenced (e.g., if the participant does not respond to intervention or specific research problems arise). These modifications do not necessarily compromise the experiment (Connell & Thompson, 1986; Gravetter & Forzano, 2009). As a consequence, in a SCED the design and/or intervention actually received may depart from an a priori protocol. Such departure is considered acceptable within SCED methodology, as long as the authors declare such departure and provide justification for it (see Item 6). Nonetheless, the flexibility to modify the design and intervention in SCEDs does not imply that an a priori protocol is not relevant.

The SPIRIT 2013 Statement (Standard Protocol Items: Recommendations for Interventional Trials; Chan et al., 2013) provides guidance for the report of protocols of clinical trials and can be used as a guide for preparing a protocol on a SCED study. As noted in the CONSORT 2010 Explanation and Elaboration document (Moher et al., 2010a), there are many ways in which a protocol can be made available, including trial registries, journals that publish protocols, website of the journal publishing the main results of the study, author's institutional website, contact with the author. Reference to the study protocol can be made either in the text or as a footnote.

Item 26—Funding: Identify source/s of funding and other support; describe the role of funders.

Example.

Funding.

The study was financed by the Sint Maartenskliniek Nijmegen and Woerden, the Netherlands.

Competing interests.

All authors declare: no support from any organisation for the submitted work, no financial relationships with any organisations that might have an interest in the submitted work in the previous 3 years and no other relationships or activities that could appear to have influenced the submitted work. (Hoogeboom et al., 2012, p. 8)

Explanation. Journals often request that authors disclose funding sources. It is important that readers of the manuscript can make a judgment as to whether the funders have control over knowledge dissemination or whether they provided funds but the researchers worked independently and autonomously. A consistently and overwhelmingly strong association between industry support and proindustry findings in group clinical trials has been demonstrated (Sismondo, 2008a, 2008b). The likelihood of finding a positive result when backed by industry funding has been estimated in the range of 3.6 to 4.05 times greater than when not (Bekelman, Li, & Gross, 2003; Lexchin, Bero, Djulbegovic, & Clark, 2003). Although methodological quality is not necessarily compromised in industry-funded research (Sismondo, 2008a), bias may infiltrate in other ways such as the decision to use less active controls (Bekelman et al., 2003).

Provision of a grant number can be helpful because it enables readers to retrieve the details of the grant. Otherwise, one needs to know the following: Did the funder provide equipment or money? Did the funder have control over where the study was published or what was published? Were the funders the authors, or did they edit the manuscript? Any conflicts of interest should be stated explicitly. This information can be provided in the body of the text or in a footnote.

Conclusion

We developed the SCRIBE 2016 to assist investigators in the behavioral sciences to report SCEDs with transparency, accuracy, clarity and completeness. This article provides rationale for and explanation of the 26 SCRIBE 2016 items. It also includes examples of adequate reporting of specific SCRIBE 2016 items, drawing on articles in the published literature. Authors reporting on one specific type of single-case methodology, the medical *N*-of-1 trial with multiple cross-overs, will find it helpful to consult the CENT 2015 Statement (Shamseer et al., 2015; Vohra et al., 2015), which was developed for that particular methodology. We welcome feedback from users of the SCRIBE 2016, which can be made through the SCRIBE website (www.sydney.edu.au/medicine/research/scribe).

Since the first CONSORT guideline appeared in 1996 (Begg et al., 1996), medical journals have continued advocating the use of prescriptive reporting guidelines in the CONSORT tradition. The EQUATOR network (www.equator-network.org) is a useful resource to keep up-to-date with new developments in this field. This development has not occurred in the behavioral sciences to the same degree, although many authors of intervention studies in the behavioral sciences consult relevant CONSORT Statements (e.g., CONSORT Extension to Nonpharmacological Interventions; Boutron et al., 2008). The influence of CONSORT is such that the peak medical journals require that authors address all of the criteria of the relevant guideline in their report. The benefit has resulted in improved reporting (Turner et al., 2012). It will be advantageous to the behavioral sciences if journals publishing single-case methodology also endorse use of the SCRIBE 2016 in this way.

References

Allison, D. B., & Gorman, B. S. (1993). Calculating effect sizes for meta-analysis: The case of the single case. *Behaviour Research and Therapy, 31,* 621–631. http://dx.doi.org/10.1016/0005-7967(93)90115-B

Alresheed, F., Hott, B. L., & Bano, C. (2013). Single subject research: A synthesis of analytic methods. *Journal of Special Education Apprenticeship, 2,* 1–18. http://josea.info/archives/vol2no1/vol2no1-1-FT.pdf

Arntz, A., Sofi, D., & van Breukelen, G. (2013). Imagery Rescripting as treatment for complicated PTSD in refugees: A multiple baseline case series study. *Behaviour Research and Therapy, 51,* 274–283. http://dx.doi.org/10.1016/j.brat.2013.02.009

Baker, E. (2012). Optimal intervention intensity. *International Journal of Speech-Language Pathology, 14,* 401–409. http://dx.doi.org/10.3109/17549507.2012.700323

Bang, H., Ni, L., & Davis, C. E. (2004). Assessment of blinding in clinical trials. *Controlled Clinical Trials, 25,* 143–156. http://dx.doi.org/10.1016/j.cct.2003.10.016

Barlow, D. H., & Hersen, M. (1984). *Single case experimental designs. Strategies for studying behaviour change* (2nd ed.). Boston, MA: Allyn & Bacon.

Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). Boston, MA: Pearson.

Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject treatment research: Lessons learned from the aphasia literature. *Neuropsychology Review, 16,* 161–169. http://dx.doi.org/10.1007/s11065-006-9013-7

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., Olkin, I., . . . Stroup, D. F. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT Statement. *Journal of the American Medical Association, 276,* 637–639. http://dx.doi.org/10.1001/jama.1996.03540080059030

Bekelman, J. E., Li, Y., & Gross, C. P. (2003). Scope and impact of financial conflicts of interest in biomedical research: A systematic review. *Journal of the American Medical Association, 289,* 454–465. http://dx.doi.org/10.1001/jama.289.4.454

Bennett, S. M., Ehrenreich-May, J., Litz, B. T., Boisseau, C. L., & Barlow, D. H. (2012). Development and preliminary evaluation of a cognitive-behavioral intervention for perinatal grief. *Cognitive and Behavioral Practice, 19,* 161–173. http://dx.doi.org/10.1016/j.cbpra.2011.01.002

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,* 129–141. http://dx.doi.org/10.1080/17489530802446302

Bernard, R. S., Cohen, L. L., & Moffett, K. (2009). A token economy for exercise adherence in pediatric cystic fibrosis: A single-subject analysis. *Journal of Pediatric Psychology, 34,* 354–365. http://dx.doi.org/10.1093/jpepsy/jsn101

Bloom, M., & Fischer, J. (1982). *Evaluating practice: Guidelines for the accountable professional* (1st ed.). Englewood Cliffs, NJ: Prentice Hall.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54,* 579–616. http://dx.doi.org/10.1146/annurev.psych.54.101601.145030

Borrelli, B., Sepinwall, D., Ernst, D., Bellg, A. J., Czajkowski, S., Breger, R., . . . Orwig, D. (2005). A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *Journal of Consulting and Clinical Psychology, 73,* 852–860. http://dx.doi.org/10.1037/0022-006X.73.5.852

Boutron, I., Guittet, L., Estellat, C., Moher, D., Hróbjartsson, A., & Ravaud, P. (2007). Reporting methods of blinding in randomized trials assessing nonpharmacological treatments. *PLoS Medicine, 4,* e61. http://dx.doi.org/10.1371/journal.pmed.0040061

Boutron, I., Moher, D., Altman, D. G., Schulz, K. F., & Ravaud, P., & the CONSORT Group. (2008). Extending the CONSORT Statement to randomized trials of nonpharmacologic treatment: Explanation and elaboration. *Annals of Internal Medicine, 148,* 295–309. http://dx.doi.org/10.7326/0003-4819-148-4-200802190-00008

Boutron, I., Tubach, F., Giraudeau, B., & Ravaud, P. (2004). Blinding was judged more difficult to achieve and maintain in nonpharmacologic than pharmacologic trials. *Journal of Clinical Epidemiology, 57,* 543–550. http://dx.doi.org/10.1016/j.jclinepi.2003.12.010

Box, G. E., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control* (1st ed.). San Francisco, CA: Holden-Day.

Bulté, I., & Onghena, P. (2012). When the truth hits you between the eyes: A software tool for the visual analysis of single-case experimental data. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 8,* 104–114. http://dx.doi.org/10.1027/1614-2241/a000042

Busk, P., & Marascuilo, L. A. (1992). Statistical analysis in single-case research: Issues, procedures, and applications to multiple behaviors. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case research design and analysis: New directions for psychology and education* (pp. 159–186). Hillsdale, NJ: Erlbaum.

Carr, J. E. (2005). Recommendations for reporting multiple-baseline designs across participants. *Behavioral Interventions, 20,* 219–224. http://dx.doi.org/10.1002/bin.191

Center, B. A., Skiba, R. J., & Casey, A. (1985). A methodology for the quantitative synthesis of intra-subject design research. *The Journal of Special Education, 19,* 387–400. http://dx.doi.org/10.1177/002246698501900404

Chan, A-W., Tetzlaff, J. M., Altman, D. G., Laupacis, A., Gøtzsche, P. C., Krleža-Jerić, K., . . . Moher, D. (2013). SPIRIT 2013 Statement: Defining standard protocol items for clinical trials. *Annals of Internal Medicine, 158,* 200–207. http://dx.doi.org/10.7326/0003-4819-158-3-201302050-00583

Chard, D. J., Ketterlin-Geller, L. R., Baker, S. K., Doabler, C., & Apichatabutra, C. (2009). Repeated reading intervention for students with learning disabilities: Status of the evidence. *Exceptional Children, 75,* 263–281. http://ecx.sagepub.com/content/75/3/263.short

Chorpita, B. F., Albano, A. M., Heimberg, R. G., & Barlow, D. H. (1996). A systematic replication of the prescriptive treatment of school refusal behavior in a single subject. *Journal of Behavior Therapy and Experimental Psychiatry, 27,* 281–290. http://dx.doi.org/10.1016/S0005-7916(96)00023-7

Connell, P. J., & Thompson, C. K. (1986). Flexibility of single-subject experimental designs. Part III: Using flexibility to design or modify experiments. *The Journal of Speech and Hearing Disorders, 51,* 214–225. http://dx.doi.org/10.1044/jshd.5103.214

Crosson, B., Moore, A. B., McGregor, K. M., Chang, Y.-L., Benjamin, M., Gopinath, K., . . . White, K. D. (2009). Regional changes in word-production laterality after a naming treatment designed to produce a rightward shift in frontal activity. *Brain and Language, 111,* 73–85. http://dx.doi.org/10.1016/j.bandl.2009.08.001

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18,* 23–45. http://dx.doi.org/10.1016/S0272-7358(97)00043-3

Davis, D. H., Gagné, P., Fredrick, L. D., Alberto, P. A., Waugh, R. E., & Haardörfer, R. (2013). Augmenting visual analysis in single-case research with hierarchical linear modeling. *Behavior Modification, 37,* 62–89. http://dx.doi.org/10.1177/0145445512453734

DeCarlo, L. T., & Tryon, W. W. (1993). Estimating and testing autocorrelation with small samples: A comparison of the C-statistic to a modified estimator. *Behaviour Research and Therapy, 31,* 781–788. http://dx.doi.org/10.1016/0005-7967(93)90009-J

Didden, R., Korzilius, H., van Oorsouw, W., & Sturmey, P. (2006). Behavioral treatment of challenging behaviors in individuals with mild mental retardation: Meta-analysis of single-subject research. *American Journal on Mental Retardation, 111,* 290–298. http://dx.doi.org/10.1352/0895-8017(2006)111[290:BTOCBI]2.0.CO;2

Dijkers, M., Kropp, G. C., Esper, R. M., Yavuzer, G., Cullen, N., & Bakdalieh, Y. (2002). Quality of intervention research reporting in medical rehabilitation journals. *American Journal of Physical Medicine & Rehabilitation, 81,* 21–33. http://dx.doi.org/10.1097/00002060-200201000-00005

Ducharme, J. M., Davidson, A., & Rushford, N. (2002). Treatment of oppositional behavior in children of parents with brain injury and chronic pain. *Journal of Emotional and Behavioral Disorders, 10,* 241–248. http://dx.doi.org/10.1177/10634266020100040601

Dwan, K., Altman, D. G., Cresswell, L., Blundell, M., Gamble, C. L., & Williamson, P. R. (2011). Comparison of protocols and registry entries to published reports for randomised controlled trials. *Cochrane Database of Systematic Reviews, 1,* MR000031. http://dx.doi.org/10.1002/14651858.mr000031.pub2

Edgington, E. S. (1980). Validity of randomization tests for one-subject experiments. *Journal of Educational Statistics, 5,* 235–251. http://dx.doi.org/10.2307/1164966

Edgington, E. S. (1996). Randomized single-subject experimental designs. *Behaviour Research and Therapy, 34,* 567–574. http://dx.doi.org/10.1016/0005-7967(96)00012-5

Edinger, J. D., Wohlgemuth, W. K., Radtke, R. A., Marsh, G. R., & Quillian, R. E. (2001). Cognitive behavioral therapy for treatment of chronic primary insomnia: A randomized controlled trial. *Journal of the American Medical Association, 285,* 1856–1864. http://dx.doi.org/10.1001/jama.285.14.1856

Ehlhardt, L. A., Sohlberg, M. M., Glang, A., & Albin, R. (2005). TEACH-M: A pilot study evaluating an instructional sequence for persons with impaired memory and executive functions. *Brain Injury, 19,* 569–583. http://dx.doi.org/10.1080/02699050400013550

Feeney, T. J. (2010). Structured flexibility: The use of context-sensitive self-regulatory scripts to support young persons with acquired brain injury and behavioral difficulties. *The Journal of Head Trauma Rehabilitation, 25,* 416–425. http://dx.doi.org/10.1097/HTR.0b013e3181fbc0a2

Ferron, J. M., & Levin, J. R. (2014). Single-case permutation and randomization statistical tests: Present status, promising new developments. In T. R. Kratochwill and J. R. Levin (Eds.), *Single-case intervention research: Methodological and statistical advances* (pp. 153–183). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/14376-006

Ferron, J., & Onghena, P. (1996). The power of randomization tests for single-case phase designs. *Journal of Experimental Education, 64,* 231–239. http://dx.doi.org/10.1080/00220973.1996.9943805

Ferron, J., & Ware, W. (1994). Using randomization tests with responsive single-case designs. *Behaviour Research and Therapy, 32,* 787–791. http://dx.doi.org/10.1016/0005-7967(94)90037-X

Fisher, R. A. (1920). A mathematical examination of the methods of determining the accuracy of an observation by the mean error, and by the mean square error. *Monthly Notices of the Royal Astronomical Society, 80,* 758–770. http://dx.doi.org/10.1093/mnras/80.8.758

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association, 32,* 675–701. http://dx.doi.org/10.1080/01621459.1937.10503522

Gabler, N. B., Duan, N., Vohra, S., & Kravitz, R. L. (2011). *N*-of-1 trials in the medical literature: A systematic review. *Medical Care, 49,* 761–768. http://dx.doi.org/10.1097/MLR.0b013e318215d90d

Gagnier, J. J., Kienle, G., Altman, D. G., Moher, D., Sox, H., & Riley, D., & the CARE Group. (2014). The CARE guidelines: Consensus-based clinical case report guideline development. *Journal of Clinical Epidemiology, 67,* 46–51. http://dx.doi.org/10.1016/j.jclinepi.2013.08.003

Gast, D. L. (2010a). Replication. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 110–128). New York, NY: Routledge.

Gast, D. L. (Ed.). (2010b). *Single subject research methodology in behavioral sciences* (1st ed.). New York, NY: Routledge.

Glasziou, P., Meats, E., Heneghan, C., & Shepperd, S. (2008). What is missing from descriptions of treatment in trials and reviews? *British Medical Journal, 336,* 1472–1474. http://dx.doi.org/10.1136/bmj.39590.732037.47

Golder, S., Loke, Y. K., & Bland, M. (2011). Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: Methodological overview. *PLoS Medicine, 8,* e1001026. http://dx.doi.org/10.1371/journal.pmed.1001026

Gorsuch, R. L. (1983). Three methods for analyzing limited time-series (N of 1) data. *Behavioral Assessment, 5,* 141–154.

Gottman, J. M., & Glass, G. V. (1978). Analysis of interrupted time-series experiments. In T. R. Kratochwill (Ed.), *Single-subject research: Strategies for evaluating change* (pp. 197–235). New York, NY: Academic Press. http://dx.doi.org/10.1016/B978-0-12-425850-1.50011-9

Gravetter, F. J., & Forzano, L. A. B. (2009). *Research methods for the behavioral sciences* (3rd ed.). Belmont, CA: Wadsworth.

Guyatt, G., Jaeschke, R., & McGinn, T. (2002). N-of-1 randomized controlled trials. In G. Guyatt & D. Rennie (Eds.), *Users' guides to the medical literature* (pp. 275–290). Chicago, IL: American Medical Association.

Hagermoser Sanetti, L. M., & Kratochwill, T. R. (2007–2008). Treatment integrity in behavioral consultation: Measurement, promotion, and outcomes. *International Journal of Behavioral and Consultation Therapy, 4,* 95–114.

Hagermoser Sanetti, L. M., & Kratochwill, T. R. (Eds.). (2014). *Treatment integrity: A foundation for evidence-based practice in applied psychology*. Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/14275-000

Hammond, D., & Gast, D. L. (2010). Descriptive analysis of single subject research designs: 1983–2007. *Education and Training in Autism and Developmental Disabilities, 45,* 187–202. http://www.jstor.org/stable/23879806

Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The scientist practitioner: Research and accountability in the age of managed care* (2nd ed.). Boston, MA: Pearson.

Hedges, L. V., Pustejovsky, J. E., & Shadish, W. R. (2013). A standardized mean difference effect size for multiple baseline designs across individuals. *Research Synthesis Methods, 4,* 324–341. http://dx.doi.org/10.1002/jrsm.1086

Hickey, E., Bourgeois, M., & Olswang, L. (2004). Effects of training volunteers to converse with nursing home residents with aphasia. *Aphasiology, 18,* 625–637. http://dx.doi.org/10.1080/02687030444000093

Higginbotham, D. J., & Bedrosian, J. (1995). Subject selection in AAC research: Decision points. *Augmentative and Alternative Communication, 11,* 11–13. http://dx.doi.org/10.1080/07434619512331277099

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., . . . Michie, S. (2014). Better reporting of interventions: Template for Intervention Description and Replication (TIDieR) checklist and guide. *British Medical Journal, 348,* g1687. http://dx.doi.org/10.1136/bmj.g1687

Holcombe, A., Wolery, M., & Snyder, E. (1994). Effects of two levels of procedural fidelity with constant time delay on children's learning. *Journal of Behavioral Education, 4,* 49–73. http://dx.doi.org/10.1007/BF01560509

Hoogeboom, T. J., Kwakkenbos, L., Rietveld, L., den Broeder, A. A., de Bie, R. A., & van den Ende, C. H. M. (2012). Feasibility and potential effectiveness of a non-pharmacological multidisciplinary care programme for persons with generalised osteoarthritis: A randomised, multiple-baseline single-case study. *British Medical Journal Open, 2,* e001161. http://dx.doi.org/10.1136/bmjopen-2012-001161

Hopewell, S., Clarke, M., Moher, D., Wager, E., Middleton, P., Altman, D. G., . . . the CONSORT Group. (2008). CONSORT for reporting randomized controlled trials in journal and conference abstracts: Explanation and elaboration. *PLoS Medicine, 5,* e20. http://dx.doi.org/10.1371/journal.pmed.0050020

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. http://dx.doi.org/10.1177/001440290507100203

Ioannidis, J. P., Evans, S. J., Gøtzsche, P. C., O'Neill, R. T., Altman, D. G., Schulz, K., . . . the CONSORT Group. (2004). Better reporting of harms in randomized trials: An extension of the CONSORT Statement. *Annals of Internal Medicine, 141,* 781–788. http://dx.doi.org/10.7326/0003-4819-141-10-200411160-00009

Kazdin, A. E. (1982a). Single-case experimental designs in clinical research and practice. *New Directions for Methodology of Social & Behavioral Science, 13,* 33–47.

Kazdin, A. E. (1982b). *Single-case research designs: Methods for clinical and applied settings*. New York, NY: Oxford University Press.

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings*. New York, NY: Oxford University Press.

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://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34,* 26–38. http://dx.doi.org/10.1177/0741932512452794

Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 15,* 124–144. http://dx.doi.org/10.1037/a0017736

Kratochwill, T. R., & Levin, J. R. (2014). *Single-case intervention research: Methodological and statistical advances*. Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/14376-000

Kravitz, R., Duan, N., eds, & the DEcIDE Methods Center *N*-of-1 Guidance Panel. (2014). *Design and implementation of N-of-1 trials: A user's guide*. AHRQ Publication No. 13(14)-EHC122-EF. Retrieved from www.effectivehealthcare.ahrq.gov/N-1-Trials.cfm

Krishef, C. H. (1991). *Fundamental approaches to single subject design and analysis*. Malabar, FL: Krieger Publishing Company.

Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation, 24,* 445–463. http://dx.doi.org/10.1080/09602011.2013.815636

Lane, K., Wolery, M., Reichow, B., & Rogers, L. (2007). Describing baseline conditions: Suggestions for study reports. *Journal of Behavioral Education, 16,* 224–234. http://dx.doi.org/10.1007/s10864-006-9036-4

Leon, S. A., Rosenbek, J. C., Crucian, G. P., Hieber, B., Holiway, B., Rodriguez, A. D., . . . Gonzalez-Rothi, L. (2005). Active treatments for aprosodia secondary to right hemisphere stroke. *Journal of Rehabilitation Research and Development, 42,* 93–101. http://dx.doi.org/10.1682/jrrd.2003.12.0182

Lexchin, J., Bero, L. A., Djulbegovic, B., & Clark, O. (2003). Pharmaceutical industry sponsorship and research outcome and quality: Systematic review. *British Medical Journal, 326,* 1167–1170. http://dx.doi.org/10.1136/bmj.326.7400.1167

Lloyd, B., & Sherrington, C. (2011). *Exercise-based videogames for stroke rehabilitation at home: A single subject randomised trial*. Retrieved from https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=343162&isClinicalTrial=False

Lo, Y.-y., Cooke, N. L., & Starling, A. L. P. (2011). Using a repeated reading program to improve generalization of oral reading fluency. *Education & Treatment of Children, 34,* 115–140. http://dx.doi.org/10.1353/etc.2011.0007

Loke, Y. K., Price, D., & Herxheimer, A., & the Cochrane Adverse Effects Methods Group. (2007). Systematic reviews of adverse effects: Framework for a structured approach. *BMC Medical Research Methodology, 7,* 32. http://dx.doi.org/10.1186/1471-2288-7-32

Maggin, D. M., Briesch, A. M., Chafouleas, S. M., Ferguson, T. D., & Clark, C. (2014). A comparison of rubrics for identifying empirically supported practices with single-case research. *Journal of Behavioral Education, 23,* 287–311. http://dx.doi.org/10.1007/s10864-013-9187-z

Maggin, D. M., Chafouleas, S. M., Goddard, K. M., & Johnson, A. H. (2011). A systematic evaluation of token economies as a classroom management tool for students with challenging behavior. *Journal of School Psychology, 49,* 529–554. http://dx.doi.org/10.1016/j.jsp.2011.05.001

Maggin, D. M., & Odom, S. L. (2014). Evaluating single-case research data for systematic review: A commentary for the special issue. *Journal of School Psychology, 52,* 237–241. http://dx.doi.org/10.1016/j.jsp.2014.01.002

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, 19,* 109–135. http://dx.doi.org/10.1080/09362835.2011.565725

Manolov, R., Gast, D. L., Perdices, M., & Evans, J. J. (2014). Single-case experimental designs: Reflections on conduct and analysis. *Neuropsychological Rehabilitation,* 24(3–4):*634*–660. http://dx.doi.org/10.1080/09602011.2014.903199

Mechling, L. C. (2006). Comparison of the effects of three approaches on the frequency of stimulus activations, via a single switch, by students with profound intellectual disabilities. *The Journal of Special Education, 40,* 94–102. http://dx.doi.org/10.1177/00224669060400020501

Mechling, L. C., & Gast, D. L. (2010). Ethical principles and practices. In D. L. Gast (Ed.), *Single subject research methodology in behavioral sciences* (pp. 32–56). New York, NY: Routledge Publishers.

Mesagno, C., Marchant, D., & Morris, T. (2009). Alleviating choking: The sounds of distraction. *Journal of Applied Sport Psychology, 21,* 131–147. http://dx.doi.org/10.1080/10413200902795091

Methe, S. A., & Hintze, J. M. (2003). Evaluating teacher modeling as a strategy to increase student reading behavior. *School Psychology Review, 32,* 617–623.

Miltenberger, R. G., Flessner, C., Gatheridge, B., Johnson, B., Satterlund, M., & Egemo, K. (2004). Evaluation of behavioral skills training to prevent gun play in children. *Journal of Applied Behavior Analysis, 37,* 513–516. http://dx.doi.org/10.1901/jaba.2004.37-513

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., . . . Altman, D. G. (2010a). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *British Medical Journal, 340,* c869. http://dx.doi.org/10.1136/bmj.c869

Moher, D., Schulz, K. F., Simera, I., & Altman, D. G. (2010b). Guidance for developers of health research reporting guidelines. *PLoS Medicine, 7,* e1000217. http://dx.doi.org/10.1371/journal.pmed.1000217

Montgomery, P., Grant, S., Hopewell, S., Macdonald, G., Moher, D., Michie, S., & Mayo-Wilson, E. (2013). Protocol for CONSORT-SPI: An extension for social and psychological interventions. *Implementation Science; IS, 8,* 99. http://dx.doi.org/10.1186/1748-5908-8-99

National Health and Medical Research Council. (2009). *NHMRC Australian Health Ethics Committee (AHEC) position statement: Monitoring and reporting of safety for clinical trials involving therapeutic products*. Canberra, Australia: Author.

OCEBM. (2011). *The Oxford 2011 levels of evidence*. Retrieved from http://www.cebm.net/index.aspx?o=5653

Onghena, P., & Edgington, E. S. (2005). Customization of pain treatments: Single-case design and analysis. *The Clinical Journal of Pain, 21,* 56–68. http://dx.doi.org/10.1097/00002508-200501000-00007

Papanikolaou, P. N., Christidi, G. D., & Ioannidis, J. P. A. (2006). Comparison of evidence on harms of medical interventions in randomized and nonrandomized studies. *CMAJ Canadian Medical Association Journal, 174,* 635–641. http://dx.doi.org/10.1503/cmaj.050873

Parker, R. I., & Brossart, D. F. (2003). Evaluating single-case research data: A comparison of seven statistical methods. *Behavior Therapy, 34,* 189–211. http://dx.doi.org/10.1016/S0005-7894(03)80013-8

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 40,* 357–367. http://dx.doi.org/10.1016/j.beth.2008.10.006

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The "improvement rate difference" for single-case research. *Exceptional Children, 75,* 135–150. http://ecx.sagepub.com/content/75/2/135.abstract

Parker, R. I., Vannest, K. J., & Davis, J. L. (2014). Non-overlap analysis for single-case research. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-case intervention research: Methodological and statistical advances* (pp. 127–151). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/14376-005

Parker, R. I., Vannest, K. J., Davis, J. L., & Sauber, S. B. (2011). Combining nonoverlap and trend for single-case research: Tau-U. *Behavior Therapy, 42,* 284–299. http://dx.doi.org/10.1016/j.beth.2010.08.006

Parsonson, B. S., & Baer, D. M. (1986). The graphic analysis of data. In A. Poling & R. W. Fuqua (Eds.), *Research methods in applied behavior analysis: Issues and advances* (pp. 157–186). New York, NY: Plenum Press. http://dx.doi.org/10.1007/978-1-4684-8786-2_8

Perdices, M., Schultz, R., Tate, R., McDonald, S., Togher, L., Savage, S., . . . Smith, K. (2006). The evidence base of neuropsychological rehabilitation in acquired brain impairment (ABI): How good is the research? *Brain Impairment, 7,* 119–132. http://dx.doi.org/10.1375/brim.7.2.119

Perdices, M., & Tate, R. L. (2009). Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognised and undervalued? *Neuropsychological Rehabilitation, 19,* 904–927. http://dx.doi.org/10.1080/09602010903040691

Pool, D., Blackmore, A. M., Bear, N., & Valentine, J. (2014). Effects of short-term daily community walk aide use on children with unilateral spastic cerebral palsy. *Pediatric Physical Therapy, 26,* 308–317. http://dx.doi.org/10.1097/PEP.0000000000000057

Punja, S., Eslick, I., Duan, N., Vohra, S., & the DEcIDE Methods Center *N*-of-1 Guidance Panel. (2014). An ethical framework for N-of-1 trials: Clinical care, quality improvement, or human subjects research? In R. L. Kravitz, N. Duan, & the DEcIDE Methods Center *N*-of-1 Guidance Panel (Eds.), *Design and implementation of N-of-1 trials: A user's guide*. AHRQ Publication No. 13(14)-EHC122-EF (pp. 13–22). Rockville, MD: Agency for Healthcare Research and Quality. Retrieved from http://www.effectivehealthcare.ahrq.gov/N-1-Trials.cfm

Rider, J. D., Wright, H. H., Marshall, R. C., & Page, J. L. (2008). Using semantic feature analysis to improve contextual discourse in adults with aphasia. *American Journal of Speech-Language Pathology, 17,* 161–172. http://dx.doi.org/10.1044/1058-0360(2008/016)

Robey, R., Shultz, M., Crawford, A., & Sinner, C. (1999). Single-subject clinical-outcome research: Designs, data, effect sizes, and analyses. *Aphasiology, 13,* 445–473. http://dx.doi.org/10.1080/026870399402028

Rose, M., & Sussmilch, G. (2008). The effects of semantic and gesture treatments on verb retrieval and verb use in aphasia. *Aphasiology, 22,* 691–706. http://dx.doi.org/10.1080/02687030701800800

Satir, D. A., Goodman, D. M., Shingleton, R. M., Porcerelli, J. H., Gorman, B. S., Pratt, E. M., . . . Thompson-Brenner, H. (2011). Alliance-focused therapy for anorexia nervosa: Integrative relational and behavioral change treatments in a single-case experimental design. *Psychotherapy: Theory, Research, & Practice, 48,* 401–420. http://dx.doi.org/10.1037/a0026216

Schlosser, R. W., & Braun, U. (1994). Efficacy of AAC interventions: Methodologic issues in evaluating behavior change, generalization, and effects. *AAC Augmentative and Alternative Communication, 10,* 207–223. http://dx.doi.org/10.1080/07434619412331276920

Schulz, K. F., Altman, D. G., Moher, D., & the CONSORT Group. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *BMC Medicine, 8,* 18. http://dx.doi.org/10.1186/1741-7015-8-18

Schulz, K. F., & Grimes, D. A. (2002). Blinding in randomised trials: Hiding who got what. *Lancet, 359,* 696–700. http://dx.doi.org/10.1016/S0140-6736(02)07816-9

Scott, D., Scott, L. M., & Goldwater, B. (1997). A performance improvement program for an international-level track and field athlete. *Journal of Applied Behavior Analysis, 30,* 573–575. http://dx.doi.org/10.1901/jaba.1997.30-573

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,* 9–19. http://dx.doi.org/10.1177/0741932512440730

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education, 8,* 24–33. http://dx.doi.org/10.1177/074193258700800206

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

Shadish, W. R., Hedges, L. V., Pustejovsky, J. E., Boyajian, J. G., Sullivan, K. J., Andrade, A., & Barrientos, J. L. (2014). A d-statistic for single-case designs that is equivalent to the usual between-groups d-statistic. *Neuropsychological Rehabilitation, 24,* 528–553. http://dx.doi.org/10.1080/09602011.2013.819021

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. http://dx.doi.org/10.1080/17489530802581603

Shadish, W. R., & Sullivan, K. J. (2011). Characteristics of single-case designs used to assess intervention effects in 2008. *Behavior Research Methods, 43,* 971–980. http://dx.doi.org/10.3758/s13428-011-0111-y

Shamseer, L., Sampson, M., Bukutu, C., Schmid, C. H., Nikles, J., Tate, R. . . . and the CENT group. (2015). CONSORT extension for reporting *N*-of-1 trials (CENT) 2015: Explanation and elaboration. *British Medical Journal, 350,* h1793. http://dx.doi.org/10.1136/bmj/h1793

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York, NY: Basic Books, Inc.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York, NY: McGraw-Hill.

Simera, I., Moher, D., Hoey, J., Schulz, K. F., & Altman, D. G. (2010). A catalogue of reporting guidelines for health research. *European Journal of Clinical Investigation, 40,* 35–53. http://dx.doi.org/10.1111/j.1365-2362.2009.02234.x

Sismondo, S. (2008a). How pharmaceutical industry funding affects trial outcomes: Causal structures and responses. *Social Science & Medicine, 66,* 1909–1914. http://dx.doi.org/10.1016/j.socscimed.2008.01.010

Sismondo, S. (2008b). Pharmaceutical company funding and its consequences: A qualitative systematic review. *Contemporary Clinical Trials, 29,* 109–113. http://dx.doi.org/10.1016/j.cct.2007.08.001

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

Student. (1908). The probable error of a mean. *Biometrika, 6,* 1–25. http://www.jstor.org/stable/2331554

Swaminathan, H., Rogers, H. J., Horner, R. H., Sugai, G., & Smolkowski, K. (2014). Regression models and effect size measures for single case designs. *Neuropsychological Rehabilitation, 24,* 554–571. http://www.tandfonline.com/doi/abs/10.1080/09602011.2014.887586

Tate, R. L., Perdices, M., McDonald, S., Togher, L., & Rosenkoetter, U. (2014). The design, conduct and report of single-case research: Resources to improve the quality of the neurorehabilitation literature. *Neuropsychological Rehabilitation, 24,* 315–331. http://dx.doi.org/10.1080/09602011.2013.875043

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., . . . Wilson, B. (2016). The Single-Case Reporting guideline In

BEhavioural interventions (SCRIBE) 2016 statement. *Archives of Scientific Psychology, 4,* 1–9.

Tate, R. L., Perdices, M., Rosenkoetter, U., Wakim, D., Godbee, K., Togher, L., & McDonald, S. (2013a). *Manual for the critical appraisal of single-case reports using the Risk of Bias in N-of-1 Trials (RoBiNT) Scale*. Unpublished manuscript, Rehabilitation Studies Unit, Sydney Medical School - Northern, The University of Sydney, Sydney, Australia.

Tate, R. L., Perdices, M., Rosenkoetter, U., Wakim, D., Godbee, K., Togher, L., & McDonald, S. (2013b). Revision of a method quality rating scale for single-case experimental designs and *N*-of-1 trials: The 15-item Risk of Bias in *N*-of-1 Trials (RoBiNT) Scale. *Neuropsychological Rehabilitation, 23,* 619–638. http://dx.doi.org/10.1080/09602011.2013.824383

Tate, R. L., Rosenkoetter, U., Sigmundsdottir, L., Wakim, D., Doubleday, J., Togher, L., . . .Perdices, M. (2015). *The Risk of Bias in N-of-1 Trials (RoBiNT) Scale: An expanded manual for the critical appraisal of single-case reports*. Sydney, Australia: Author.

ter Kuile, M. M., Bulté, I., Weijenborg, P. T., Beekman, A., Melles, R., & Onghena, P. (2009). Therapist-aided exposure for women with lifelong vaginismus: A replicated single-case design. *Journal of Consulting and Clinical Psychology, 77,* 149–159. http://dx.doi.org/10.1037/a0014273

Treadwell, K., & Page, T. J. (1996). Functional analysis: Identifying the environmental determinants of severe behavior disorders. *The Journal of Head Trauma Rehabilitation, 11,* 62–74. http://dx.doi.org/10.1097/00001199-199602000-00008

Tryon, W. W. (1982). A simplified time-series analysis for evaluating treatment interventions. *Journal of Applied Behavior Analysis, 15,* 423–429. http://dx.doi.org/10.1901/jaba.1982.15-423

Turner, L., Shamseer, L., Altman, D. G., Weeks, L., Peters, J., Kober, T., . . . Moher, D. (2012). Consolidated standards of reporting trials (CONSORT) and the completeness of reporting of randomised controlled trials (RCTs) published in medical journals. *Cochrane Database of Systematic Reviews, 11,* MR000030. http://systematicreviewsjournal.biomedcentral.com/articles/10.1186/2046-4053-1-60

Valdimarsdóttir, H., Halldórsdóttir, L. Y., & SigurÐardóttir, Z. G. (2010). Increasing the variety of foods consumed by a picky eater: Generalization of effects across caregivers and settings. *Journal of Applied Behavior Analysis, 43,* 101–105. http://dx.doi.org/10.1901/jaba.2010.43-101

Vandenbroucke, J. P. (2006). What is the best evidence for determining harms of medical treatment? *CMAJ Canadian Medical Association Journal, 174,* 645–646. http://dx.doi.org/10.1503/cmaj.051484

Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., . . . the STROBE Initiative. (2007). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *PLoS Medicine, 4,* e297.

Vlaeyen, J. W., de Jong, J., Geilen, M., Heuts, P. H., & van Breukelen, G. (2001). Graded exposure in vivo in the treatment of pain-related fear: A replicated single-case experimental design in four patients with chronic low back pain. *Behaviour Research and Therapy, 39,* 151–166. http://dx.doi.org/10.1016/S0005-7967(99)00174-6

Vohra, S., Shamseer, L., Sampson, M., Bukutu, C., Schmid, C. H., Tate, R., . . . for the CENT group. (2015). CONSORT extension for reporting *N*-of-1 trials (CENT) 2015 Statement. *British Medical Journal, 350,* h1738. http://dx.doi.org/10.1136/bmj/h1738

Wambaugh, J. L., Mauszycki, S., Cameron, R., Wright, S., & Nessler, C. (2013). Semantic feature analysis: Incorporating typicality treatment and mediating strategy training to promote generalization. *American Journal of Speech-Language Pathology, 22,* S334–S369. http://dx.doi.org/10.1044/1058-0360(2013/12-0070)

Wambaugh, J. L., Nessler, C., & Wright, S. (2013). Modified response elaboration training: Application to procedural discourse and personal recounts. *American Journal of Speech-Language Pathology, 22,* S409–S425. http://dx.doi.org/10.1044/1058-0360(2013/12-0063)

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities Research Reviews, 13,* 70–77. http://dx.doi.org/10.1002/mrdd.20139

Wenman, R., Bowen, A., Tallis, R. C., Gardener, E., Cross, S., & Niven, D. (2003). Use of a randomised single case experimental design to evaluate therapy for unilateral neglect. *Neuropsychological Rehabilitation, 13,* 441–459. http://dx.doi.org/10.1080/09602010343000048

White, D., Rusch, F., Kazdin, A., & Hartmann, D. (1989). Applications of meta-analysis in individual-subject research. *Behavioral Assessment, 11,* 281–296.

White, O. R., & Haring, N. G. (1980). *Exceptional teaching*. Columbus, OH: Merrill.

Wilcoxin, F. (1947). Probability tables for individual comparisons by ranking methods. *Biometrics, 3,* 119–122. http://dx.doi.org/10.2307/3001946

Wolery, M. (1994). Procedural fidelity: A reminder of its functions. *Journal of Behavioral Education, 4,* 381–386. http://dx.doi.org/10.1007/BF01539539

Wolery, M., Dunlap, G., & Ledford, J. R. (2011). Single-case experimental methods: Suggestions for reporting. *Journal of Early Intervention, 33,* 103–109. http://dx.doi.org/10.1177/1053815111418235

Wolery, M., & Ezell, H. K. (1993). Subject descriptions and single-subject research. *Journal of Learning Disabilities, 26,* 642–647. http://dx.doi.org/10.1177/002221949302601001

Wood, L., Egger, M., Gluud, L. L., Schulz, K. F., Jüni, P., Altman, D. G., . . . Sterne, J. A. C. (2008). Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: Meta-epidemiological study. *British Medical Journal, 336,* 601–605. http://dx.doi.org/10.1136/bmj.39465.451748.AD

Wright, H. H., Marshall, R. C., Wilson, K. B., & Page, J. L. (2008). Using a written cueing hierarchy to improve verbal naming in aphasia. *Aphasiology, 22,* 522–536. http://dx.doi.org/10.1080/02687030701487905

(*Appendix follows*)

Appendix

Selection of Data Analysis Methods Applicable to SCEDs

For a comprehensive discussion of statistical techniques in SCEDs, see Kazdin (2011).

Visual Analysis

- Guidelines for systematic analysis (Kratochwill et al., 2010)
- Software for visual analysis (Bulté & Onghena, 2012)

Quasistatistical Techniques

- Split-middle trend line (Kazdin, 1982a; White & Haring, 1980)
- Binomial distribution test (Siegel & Castellan, 1988)
- Standard deviation band (Bloom & Fischer, 1982; Krishef, 1991)

Time Series Analysis

- *C* statistic (DeCarlo & Tryon, 1993; Tryon, 1982)
- Auto-regressive integrated moving average (Box & Jenkins, 1970; Gottman & Glass, 1978)

"Traditional" Inferential Statistics Tests

Parametric.

- *t* Test (Student, 1908)
- *F* Test (Fisher, 1920)

Nonparametric.

- Wilcoxon matched-pairs signed-ranks test (Wilcoxon, 1947)
- Friedman two-way analysis of variance (Friedman, 1937)

Effect Sizes

For reviews, see Alresheed, Hott, and Bano (2013); Beretvas and Chung (2008); Shadish, Rindskopf, and Hedges (2008):

- Trend analysis effect size (Gorsuch, 1983)
- Improvement rate difference (Parker, Vannest, & Brown, 2009)
- Nonoverlap of all pairs* (Parker & Vannest, 2009)
- Percentage of nonoverlapping data* (Scruggs & Mastropieri, 2013; Scruggs, Mastropieri, & Casto, 1987)
- Tau-U (Parker, Vannest, David & Sauber, 2011)
- Bayesian probability model
- Mean-shift and mean-plus trend models (Allison & Gorman, 1993; Center, Skiba, & Casey, 1985)
- *d* Statistic (Hedges, Pustejovsky, & Shadish, 2013; Shadish, 2014)
- Multilevel linear modelling (Swaminathan, Rogers, Horner, Sugai, & Smolkowski, 2014)

Randomization Tests (Edgington, 1980, 1996; Ferron & Onghena, 1996; Ferron & Ware, 1994; Onghena & Edgington, 2005)

* For more information on nonoverlap methods, see Parker, Vannest, and Davis (2014).

Received August 21, 2015
Revision received November 15, 2015
Accepted November 16, 2015 ■