Policy and Position Statement on Single Case Research and Experimental Designs

from the Division for Research Council for Exceptional Children

Wendy Rodgers, University of Nevada, Las Vegas; Timothy Lewis, University of Missouri; Rob O'Neill, University of Utah; Kimberly Vannest, Texas A&M University

Position on Single Case Research. The Division for Research (DR) of the Council for Exceptional Children comprises national and international researchers and practitioners within special education, school psychology, counseling, and allied fields. The DR issues this policy and position statement for professional audiences seeking brief summative information on Single Case Research (SCR). The DR recognizes and advocates for the legitimacy and importance of SCR in determining the impact of intervention on children and youth with disabilities and the thorough review of related SCR peer-reviewed published studies using quality indicators[1] as means to identify "evidence-based practices." Researchers in special education and other related disciplines often rely on single case design, particularly given that randomized control trials and other group designs are not feasible to conduct with the relatively small numbers of students with disabilities who are available to participate in research. This position statement is designed to be a companion document to the DR *Guide for the Use of Single Case Design Research Evidence*.

Definition and Essential Components of Single-Case Designs

SCR, also referred to as single case experimental designs or single subject designs, is a rigorous, experimental methodology to determine the functional relation(s) between independent and dependent variables for individual behaviors. Building primarily on the theoretical principles of Applied Behavior Analysis, SCR allows the examination of intra-individual change over time, a valuable asset within the field of special education, given the heterogeneity of students within disability categories, especially those considered to have low-incidence disabilities. Key characteristics of SCR are ***tight experimental control***, ***replication of the functional relation***, ***valid and reliable measures***, and ***data collected across time and/or individuals.***

Key elements of SCR can be found across a number of seminal methodology texts and professional resources (e.g., Barlow, Nock, and Hersen, 2008; Gast & Ledford, 2014; Kazdin, 2011; Kratochwill et al., 2010; O'Neill, McDonnell, Billingsley, & Jenson, 2011; Sidman, 1960). Essential features are briefly summarized below.

---

[1] Council for Exceptional Children. (2014). *Council for Exceptional Children Standards for Evidence-Based Practices in Special Education.*
What Works Clearinghouse. (2014). *Procedures and Standards Handbook.*

Repeated measurement. SCR employs repeated measures across time. Measures occur repeatedly across days or intervals of time that provide equal opportunities for behavioral responding and are primarily collected through measures of an individual's behavior (e.g., number of occurrences of problem behavior during a half hour period).

Participants serve as their own control. Data between conditions, or phases, allow intra-individual comparison over time as the independent variable is introduced, and in some cases withdrawn, over time. Baseline data typically are collected first under business-as-usual conditions (i.e., A phases) during which measures are expected to capture initial performance. Performance under such baseline conditions is compared to data from conditions in which interventions or treatments are being implemented (i.e., B phases) for noted trend, level, and variability changes. To document a functional relation between the independent and dependent variables, a minimum of two replications – as compared to baseline and any other phase changes – are necessary in which the noted changes show a clear difference as documented through visual analysis. Common single case designs include withdrawal designs (A-B-A-B), in which the independent variable is introduced, withdrawn to produce a return to baseline conditions, and then reintroduced; and multiple baseline designs (A-B or A-B-C), where the introduction of the independent variable is staggered over three or more baseline conditions representing different dependent variables among a single subject or the same independent/ dependent variable conditions across multiple subjects. (For more on design see references to additional readings at the end of this statement).

Visual Analysis

Unlike group design research that relies on statistical analyses to determine the statistical significance of differences among groups who receive the independent variable and those that do not (i.e., control groups), SCR uses visual analysis to identify possible functional relations between independent and dependent variables. When using visual analysis, a researcher examines within-phase patterns of graphed data and compares these patterns to subsequent phases to determine possible impact of the independent variable on the dependent variable. Within-phase visual analysis includes determining the trend (i.e., whether the dependent variable is increasing, decreasing, or flat) and variability (i.e., the degree to which daily data occur within a limited or wide range around a median). In addition, trend and variability are compared across phases, as is overall level change and the immediacy of that change. Reverse in trends, less variability, and an overall level increase or decrease as compared to the prior phase denotes a possible functional relation between the independent and dependent variables. Subsequent phase replication with similar patterns then confirms a functional relation.

Visual analysis remains the primary method for determining functional relations between the independent and dependent variables, but it presents a limitation when data patterns are not clear because two raters may not assess the data in the same way and draw similar conclusions about the presence of treatment effects. A comprehensive meta-analysis of studies of reliability of visual ratings produced an average agreement of .76 and found that design types, expertise of raters, features of graphs, and scale-range used to calculate reliability all have the potential to moderate the level of agreement between raters (Ninci, Vannest, Willson, & Zhang, 2015).

Ecological Validity. A wide variety of audiences and implementers use SCR to drive treatment and policy across the fields of education, psychology, counseling, medicine, dentistry, and rehabilitation. Differences in the purpose of the SCR and the high or low-stakes decisions associated with that purpose impact the use, selection, and reporting of statistics.

Visual analysis is helpful for formative decision-making, which is an important benefit of use for SCR. Additionally, visual analysis allows for researchers to determine whether behavior change consistently occurred when and only when condition changes occurred (i.e., whether an experimental effect is present). If the *size* of the effect is of interest (i.e., how big is the effect, if one was present), additional information can be gleaned via the use of an appropriate effect size measure (see below). Reporting effect sizes, confidence intervals, and *p*-values is consistent with strong science and ethical decision-making. These standards are likewise consistent with primary research organizations and funding sources. This does not mean visual analysis should be abandoned or ignored. Visual analysis and quantitative analysis can support each other as means of triangulation.

Statistical Analysis

There are a variety of proposed methods for conducting statistical analyses of SCR data, one of which is computing effect size (ES). ES provides an estimate of the level or degree of change in performance between baseline and treatment phases. ES calculations in research reporting are encouraged, expected, or even required by many professional organizations and institutions (e.g., American Psychological Association, 2010; What Works Clearinghouse within the Institute for Education Science, 2013; Centre for Evidence Based Intervention). Researchers engaged in randomized and non-randomized group design and SCR studies use ES as a universal metric to compare impact across a plethora of independent and dependent variables.

The DR supports scientifically principled, conceptually balanced, and ecologically valid guidelines for when and why to use an ES. An effect size calculation describes a standardized or non-standardized estimate of the degree or magnitude of change. An associated confidence interval estimates error, and a *p*-value identifies the likelihood that the finding was a chance occurrence. These pieces of information can be used in combination with visual analysis for the determination of a functional relation between a treatment/ intervention (independent variable) and performance (dependent variable). An ES allows comparability of results across studies (Shadish, 2014) and provides an opportunity to identify evidence based practices for populations or treatments that may be mostly (or only) evaluated to date using SCR designs. ES can help to address ambiguous treatment effects, cumulative treatment effects, adjust for undesirable baseline trend and minimize error in decision-making.

Decrease judgment error. Researchers and practitioners alike benefit from information with less judgment error. Accurate decisions to continue treatment, defensibility in classroom or clinic level intervention choice, and education policy improve when data are fully understood and room for misinterpretation is minimized. An ES cannot identify a flaw in a design or an error in measures – researchers must do their part to design valid studies. However, an ES can quantify the size of the effect once the data are determined to be of reasonable reliability and conclusion validity.

Dozens of ES's have been developed for or applied to SCR to date (e.g., *d*-statistic (Hedges, Pustejovsky, and Shadish, 2013): multi level models, (Moeyaert, Ugille, Ferro, Beretvas & Van den Noortgate, 2014), non-parametric dominance statistics e.g. Tau and Tau-U (Parker, Vannest, and Davis, 2010), and a number of non overlap techniques, the first of which remains the most widely used (PND; Scruggs and Mastropieri). Some techniques are elegant in their simplicity and hand-calculability; others offer recognition and a history of use in the large-n researcher community. Like any statistic, each has strengths and limitations related to assumptions about data independence, distribution, power, and ease of use. Many powerful choices are available to match the needs of both research and practice, and decisions for ES calculations should be consistent with design logic, data series length, variability within phases, trends within phases, intercept gap or change between phases, changes in level, and direction or stability. The increasing number of options for ES parallels the interest in ES treatments of SCR and is likely predicated by IES interpretations and standards for federal funding, although quantifying results for *n*=1 experiments is evident in the literature 50 years ago (Edgington, 1967). It should be noted that to date there is not universal consensus for the appropriateness of computing ES within SCR or the preferred manner to compute ES within and across SCR because many ESs are sensitive to procedural factors (Pustejovsky, 2017). At, present several research groups are examining the appropriateness and fit of ES across SCR.

Using SCR to Identify Evidence-Based Practices

There are accepted guidelines for conducting and analyzing SCR, and well-established guidance on how to evaluate multiple SCR studies to determine whether a strong evidence base exists for a particular instructional practice. Horner and colleagues (2005) published an article more than a decade ago that lays out their recommendations for identifying evidence-based practice using SCR. Their article has been cited over 2,000 times and used as the basis for dozens of meta-analyses spanning multiple areas within special education (e.g., Bellini & Akullian, 2007; Browder, Wakeman, Spooner, Ahlgrim-Delzell, & Algozzinexya, 2006; Odom, Collet-Klingenberg, Rogers, & Hattan, 2010; Rogers, L. A., & Graham, S., 2008; Test et al., 2009).

Horner et al. (2005) identified quality indicators along seven dimensions of SCR studies; their recommendations have been largely accepted in the field (see references below). These indicators include guidance on the specificity of descriptions of participants and setting, measurement of the dependent variables, acceptable inter-observer agreement levels, fidelity of implementation for the independent variables, and the number of clear replications of experimental effects. In addition to these technical considerations, Horner et al. included social validity as an important metric for SCR, recommending that the dependent variable and the amount of change be "socially important," that the implantation of the intervention be "practical and cost effective," and that the studies were conducted "over extended time periods, by typical intervention agents, in typical physical and social contexts" (p. 174).

Horner et al. (2005) also specified that before a practice should be considered evidence-based, there should be at least five high-quality SCR studies conducted by at least three different researchers in different geographical areas with a minimum of 20 total participants across the studies. These final criteria related to social validity and replications are unique to SCR and ensure that practices identified as evidence-based through this method are practical and effective for a diverse group of students.

Single case designs provide a socially valid and scientifically rigorous methodology for evaluating treatment effects in individuals, unique populations, and contextually specific research settings. The design and analysis requires a thorough understanding of the logic of designs and what questions can be meaningfully answered by their variations. A range of basic to complex designs and analyses are available. The Division for Research supports the use of SCR in intervention work, treatment development, design studies, and implementation science. Additional designs and expanded information on essential features can be found in CEC's quality indicators, the What Works Clearinghouse quality standards, and the user guide that accompanies this policy statement, available at these links:

http://www.cec.sped.org/~/media/Images/Standards/CEC%20EBP%20Standards%20cover/CECs%20Evidence%20Based%20Practice%20Standards.pdf

https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_scd_key_criteria_011017.pdf

References

American Psychological Association (2010). *Publication manual of the American Psychological Association* (6th Ed.) Washington, DC.

Baer, D.M., Wolf, M.M. & Risley, T.R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1* (Spring), 94-97.

Barlow, D. H., Nock, M. K., & Hersen, M. (2008). *Single case experimental designs: Strategies for studying behavior change* (3rd Ed.). New York: Pearson.

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

Browder, D. M., Wakeman, S. Y., Spooner, F., Ahlgrim-Delzell, L, & Algozzinexya, B. (2006). Research on reading instruction for individuals with significant cognitive disabilities. *Exceptional Children, 72,* 392-408. doi: 10.1177/001440290607200401

Cohen, J. (1994). The world is round (p< .05). *American Psychologist*, *49,* 997-1003.

Edgington, E.S. (1967). Statistical inference from N-1 experiments. *Journal of Psychology, 65,* 195-199.

Gast, D. L. & Ledford, J. R. (2014). *Single case research methodology: Application in special education and behavioral science*. New York: Routledge.

Hedges, L.G., Pustejovsky, J.E. & Shadish, W.R. (2012). A standardized mean difference effect size for single case designs. *Research Synthesis Methods, 3,* 224-239.

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

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

Kratochwill, T.R. & Levin, J.R. (2010). Enhancing the scientific credibility of single-case intervention research; Randomization to the rescue. *Psychological Methods, 15*, 122-144.

Kratochwill, T.R. & Levin, J.R. (Eds.) (2014). *Single-case Intervention research: Methodological and statistical advances*. Washington, DC: American Psychological Association.

Moeyaert, M, Ferron, J., Beretvas, S., Van den Noortgate, W. (2014). From a single-level analysis to a multilevel analysis of single-subject experimental data. *Journal of School Psychology, 52,* 191-211.

Ninci, J. Vannest, K. J., Willson, V., & Zhang N. (2015). Interrater agreement between visual analysts of single-case data: A meta-analysis. *Behavior Modification*, *39*, 510-541.

Odom, S. L., Collet-Klingenberg, L., Rogers, S. J., & Hattan, D. D. (2010). Evidence-based practices in interventions for children and youth with Autism Spectrum Disorders. *Preventing School Failure, 54,* 275-282. doi: 10.1080/10459881003785506

Rogers, L. A., & Graham, S. (2008). A meta-analysis of single subject design writing intervention research. *Journal of Educational Psychology, 100,* 879-906. doi: 10.1037/0022-0663.100.4.879

Test, D. W., Fowler, C. H., Richter, S. M., White, J., Mazzotti, V., Walker, A. R., . . . & Kortering, L. (2009). Evidence-based practices in secondary transition. *Career Development for Exceptional Individuals, 32*, 115-128. doi: 10.1177/0885728809336859

