



CEC Division for Research Position Statement:

Negative Effects of Minimum Requirements for Data Points in Multiple Baseline Designs and Multiple Probe Designs in the What Works Clearinghouse Standards Handbook, Version 4.0.

Karen R. Harris (Arizona State University), Nathan A. Stevenson (Kent State University), and James M. Kauffman (University of Virginia)

Adopted October, 2019

The mission of the What Works Clearinghouse (WWC) is "to be a central and trusted source of scientific evidence for what works in education" (WWC, 2017, p. 1). Established standards for research methodology are a useful mechanism to help researchers, and other interested parties, assess the scientific value of a singular study. Such standards enable individuals and institutions to evaluate the quality and extent of evidence for a given strategy, practice, or intervention. However, as the saying goes, any medicine that is strong enough to help is also strong enough to hurt. Misuse and/or misapplication of methodological standards, although usually helpful, can also have harmful, unintended consequences.

We first identify two single-case design (SCD) pilot standards in the WWC Standards Handbook, Version 4.0, (WWC, 2017) that are having unintended ill-effects on scientists' ability to produce meaningful high-quality research, and thus the implementation of evidence-based practices. For readers who do not have a deep background in SCD experimental research, commonly used in special education research (Harris, 2015), a brief review of the roles and benefits of this method is presented next. Each of the two pilot standards of concern are then discussed in depth. Concerns regarding the impact of these two WWC Pilot SCD Standards in the review processes of scientific journals follow. Finally, changes to the WWC Pilot Single Case Design Standards, Version 4.0, that alleviate the issues raised here are proposed.

The Two SCD Pilot Standards of Concern

The WWC Standards Handbook, Version 4.0 (2017), states that multiple baseline designs "Must have a minimum of six phases with at least five data points per phase to be rated *Meets WWC Pilot SCD Standards Without Reservations*" (p. A-5). Issues and concerns regarding this standard

for research on both instructional and behavioral interventions are addressed, including: lack of recognition of established standards and logic in SCD research for the minimum number of data points in multiple baseline designs, the ethics of potential harm to participants, threats to validity, and other potential negative consequences of the requirement of a minimum of five data points in each phase to sufficiently establish data trend, level, and variability.

The second WWC standard addressed involves use of multiple probes in multiple baseline designs. The current guidelines state, "Within the first three sessions, the design must include three consecutive probe points for each case to *Meet Pilot SCD Standards Without Reservations...*" (p. A-5). Failure to meet this requirement results in a study rating of *Does Not Meet WWC Pilot SCD Standards*. We establish that the same issues and potential negative consequences regarding the requirement for five data points per phase in a multiple baseline design to be rated *Meets WWC Pilot SCD Standards Without Reservations* apply to the requirement of "three consecutive probe points for each case within the first three sessions in a multiple probes design" to be rated *Meets Pilot SCD Standards Without Reservations*.

Role and Benefits of SCD Research

For decades, single-case experimental design has provided researchers a mechanism to explore questions not easily explored through other methods of research (Ledford, Zimmerman, Schwartz, & Odom, 2018; Rodgers, Lewis, O'Neill, & Vannest, 2018; Sidman, 1960). Numerous fields of study within the social sciences rely on SCD as an important mechanism for evaluating treatment effects, causal relations, and exploring various phenomena of interest. As Rodgers et al. (2018) stated, "Single case designs provide a socially valid and scientifically rigorous methodology for evaluating treatment effects in individuals, unique populations, and contextually specific research settings" (p. 5). Special education research in particular is known for use of single-case design as a means for evaluating the effects of academic and behavioral interventions among students with disabilities (Council for Exceptional Children, 2014; Kennedy, 2015; Ledford et al., 2018). SCD studies, however, regularly appear in highly respected research journals across multiple fields. The rigor and scientific validity of SCD in education research on instruction is on par with group design and qualitative methods (Plavnick & Ferreri, 2013).

Current federal legislation requires scientific evidence for educational practices (Plavnick & Ferreri, 2013). Rigorous, experimental SCD studies have high internal validity and a clear path to external validity (Council for Exceptional Children, 2014; Kennedy, 2005; Ledford & Gast, 2018; Rodgers et al., 2018). The role of SCD in evaluation and evolution of both academic/instructional and behavioral interventions, and the advantages SCD holds for causal analyses in teaching and learning, are clear (Barlow, Nock, & Hersen, 2009; Harris, 2015; Ledford et al., 2018; Lloyd, Saltzman, & Kauffman, 1981; Plavnik & Ferreri, 2013). The scientific record is replete with examples of important findings, strategies, programs, and interventions directly attributable to research based on single-case research design.

Further, SCD studies clearly and strongly allow researchers to address complex learning ecologies; examine with whom, how, when, and why a behavioral intervention or approach to instruction or development works; and closely investigate and respond to individual differences associated with participants (Ledford et al., 2018; Kennedy, 2005; Ledford & Gast, 2018;

Plavnick & Ferreri, 2013). As Plavnick and Ferreri noted, “An important benefit of SCED [Single-Case Experimental Design] for the application of practices or interventions . . . is that the design allows for individual differences associated with participants: a feature noticeably missing from comparisons of outcomes between groups of participants” (p. 550).

Thus, this experimental research design allows researchers to fine-tune an instructional or behavioral intervention in important ways to meet student needs while maintaining the focus on cause and effect. Researchers may use single-case studies to: (a) provide initial validation of an intervention, (b) parse the active ingredients of a given treatment, (c) substantiate cause-and-effect before engaging in group experimental studies, or (d) add to the research base required to establish an intervention as evidence-based (Council for Exceptional Children, 2014; Ledford et al., 2018; Plavnick & Ferreri, 2013; Rodgers et al., 2018). Some researchers combine SCD and experimental or qualitative methods within a single study or in a line of studies to create a robust and meaningful body of evidence around a particular question or instructional intervention (cf. Graham, Harris, & Zito, 2005).

SCD is also an indispensable tool for instructional research among populations where group design is not a viable option. For instance, for students belonging to low-incidence populations (e.g. students with intellectual disabilities, autism, deafness, etc.) logistical, geographic, and financial constraints can make group design a practical impossibility. SCD is often the only research design that can be reasonably and faithfully executed with small populations. According to the most recent data from the National Center for Education Statistics (NCES), 9 of the 13 federally recognized categories of special education represent less than 1% of all school aged children (NCES, 2017). There are very few financially and logistically feasible options for conducting experimental research with such small populations. SCD also provides a useful mechanism to explore the direct effects of educational practices and gather nuanced information specific to these participants.

In sum, SCD research (a) allows initial validation of an intervention and sets up replication studies; (b) addresses individual student needs and characteristics and allows responses to individual differences; allows researchers to address complex learning ecologies and individual differences; (c) allows examination of with whom, how, when, and why an intervention or approach to instruction or development works; (d) enables researchers to develop specific causal inferences; and (e) can add to the high quality research base needed to determine whether an approach is evidence-based.

The creation, therefore, of standards, even pilot standards, for conduct and evaluation of this empirical research approach must be done with consideration of the potential negative impacts. Inappropriate standards could lead to making SCD unpalatable in many cases and meaningfully reduce use of this approach by many researchers.

Concerns with and Consequences of the Five Data Points Per Phase Requirement in Instructional and Behavioral Research

The requirement of five data points per phase to meet WWC pilot SCD standards without reservations (WWC, 2017, p. A-5) is negatively affecting the study of instructional and

behavioral interventions in important ways. In terms of instructional interventions, many researchers are deeply concerned about the substantive, potentially damaging consequences of this WWC pilot SCD standard. Two strong WWC author groups developed the standard of five data points per phase. These groups are comprised of individuals widely regarded as experts in SCD research. There is no intention here to dispute the groups' collective expertise or experience. These groups, however, represented primarily behavioral/social areas of research (where five data points can often be collected quickly and easily for observable target behaviors such as making eye contact, contributing to a conversation, or yelling). Attention to the consequences of this standard for research on instructional interventions (e.g., instruction in fractions or in writing and argumentative essay), in which academic performance measures (rather than observational measures) are typical, is warranted. Attention to the consequences of this standard for research on behavioral interventions, in some cases, is also warranted.

A priori evidence. First, a critical standard in SCD research related to determining the minimum number of data points per phase/condition has long been that the greater the variability within the target variable, the greater the number of data points needed to establish stability in each phase (Barlow et al., 2009; Council for Exceptional Children, 2014; Kennedy, 2005; Ledford et al., 2018; Ledford & Gast, 2018). A long-standing SCD principle is that shorter baselines are logical and appropriate when there is *a priori* evidence that an individual is already performing at a stable, undesirable level (Barlow et al., 2009; Council for Exceptional Children, 2014; Harris et al., 2019; Horner & Baer, 1978; Ledford et al., 2018; Ledford & Gast, 2018; McKeown et al., 2015).

For decades, SCD studies have applied the standard of a minimum of three data points during baseline and following phases. If stability is not established within three data points, then additional data points are gathered until a stable pattern is evident (Barlow et al., 2009; Kennedy, 2005; Ledford & Gast, 2018). The current SCD standards of the Council for Exceptional Children (2014) and the Guide for the Use of Single Case Design Research Evidence (CEC-Division for Research, Ledford et al., 2018) stated that three data points can be sufficient to establish stability and level of performance, particularly where performance is low (such as in adding fractions or writing an argumentative essay) or high (such as a behavior dangerous to the participant or others) and unlikely to improve without intervention.

In SCD studies of instructional interventions (frequently with students at-risk and students with learning disabilities, emotional/behavioral disorders, autism, and so on), students are selected for an intervention by first meeting a set of criteria, such as some combination of the following: (a) a probe administered to assess the student's current performance on the academic task; (b) scores below the 33rd or 25th percentile on a standardized measure of the academic ability to be addressed; (c) teacher interviews and student work indicating the student is having difficulty with this area in class; and (d) an IEP addressing the targeted area. The *a priori* evidence is thus strong that the academic behavior is stable and fewer baselines probes are needed. For example, students who are unable to do, or perform poorly on, academic tasks such as arithmetic operations with fractions or writing an argumentative essay typically show little variability in baseline. In such cases, *three data points can be sufficient to establish a stable pattern* (Harris et al., 2019; Kennedy, 2005; Ledford et al., 2018; Ledford & Gast, 2018; McKeown et al., 2015). *A priori* data substantiates confidence in the slope and level of performance that is stable within three data points.

Similarly, a priori data are often readily available in SCD studies of behavioral interventions. Students who exhibit problematic behaviors may have an established history of problematic behaviors that can be verified with school records, including disciplinary referrals and data from functional behavioral assessments (FBA). In behavioral studies, such data are not only available, but often a key factor in the inclusion criteria for study participants. Failure to consider such data as a part of baseline evaluations risks ignoring critical information and unnecessarily extending the period of baseline data collection.

Ethical issues. Next, ethical considerations must be considered when determining the number of data points needed to establish stability in both instructional and behavioral interventions. (Harris et al., 2019; Ledford & Gast, 2018; McKeown et al., 2015). It is important to consider potential harm to participants when conducting extended baselines, including baselines that use multiple probes (Horner & Baer, 1978). The ethical issue of potential harm to participants in instructional research is straightforward: requiring students to fail repeatedly on an academic task they are unable to do well, or at all, potentially allows all of the significant harms to students well established to accompany repeated failure. For example, collecting a minimum of five data points per participant during baseline when participants cannot do the targeted tasks well or at all can create effects such as reduced motivation, boredom, increased dislike of or unwillingness to engage in the task, fatigue, behavioral issues, and other aspects of reactivity to assessment.

The ethical issue here in behavioral research is also straightforward. When a behavior is occurring that is dangerous to the participant (such as self-harm) or others (such as biting), or disruptive enough to jeopardize instructional time and opportunity for non-participants, exposing individuals to an extended baseline is unethical (Kennedy, 2005). We concur with Kennedy, who stated, "baseline needs to be as long as necessary but no longer" (p. 38). When a stable baseline cannot be established for behaviors that cannot ethically be allowed to continue without intervention, researchers focus on changing the pattern of the behavior (Kennedy, 2005).

Threats to internal validity. In addition, a related significant risk occurs when data collection during baseline or other phases is unnecessarily extended: threats to internal validity. As well established in the SCD research literature (cf. Horner & Baer, 1978; Barlow et al., 2009; Harris et al., 2019; Ledford & Gast, 2018), prolonged baselines in instructional interventions when students cannot do an academic task well or at all also can result in a decrease in academic performance across baseline data collection. In cases where this occurs due to reactivity to assessment, the effects of the instructional intervention are inflated based on comparison of baseline data to intervention data. Thus, unnecessarily prolonged data collection in baseline can result in meaningful threats to the internal validity of a study and should be avoided when performance is low and stable on academic tasks (Barlow et al., 2009; Harris et al., 2019; Horner & Baer, 1978).

Reactivity to assessment. Similarly, assessing beyond stability in phases after baseline can also result in reactivity to assessment (e.g., writing fatigue has been observed in the collection of five data points after intervention in several studies; cf. Harris et al., 2019; McKeown et al., 2015) which creates a threat to internal validity. In this case, a decrease in performance that was stable after intervention due to fatigue or other reactions to extended data

collection would lead to under appraising intervention effects. It is important to note here that these threats to internal validity occur in not only the visual analysis of SCD data, but also when applying statistical approaches to data collected in SCD studies. When baseline data or data in following phases are artificially depressed or inflated due to confounds such as those noted here, threats to internal validity exist in multiple approaches to statistical analyses. Further, statistical analyses are not required in the WWC SCD pilot standards (Council for Exceptional Children, 2014; Ledford et al., 2018; Rodgers et al., 2018; WWC, 2017). While arguments against the need for a minimum of five data points per phase have also been raised when statistical analysis is used (cf. Hwang, Levin, & Johnson, 2018), statistical analysis issues are not reviewed here.

Multiple Probes Design Requirement of Three Consecutive Probe Points for Each Case within the First Three Sessions

Each of the issues detailed in the previous section apply to the WWC Pilot SCD Standard for multiple baseline studies that use multiple probes in baseline for both instructional and behavioral interventions. The standard states, "Within the first three sessions, the design must include three consecutive probe points for each case to *Meet Pilot SCD Standards Without Reservations*." (WWC, 2017, p. A-5-6). Requiring the second, third, and any following cases to have three probes within the first three sessions as well as "a probe point in a session where another case either (a) first receives the intervention or (b) reaches the prespecified intervention criterion" and "three consecutive probe points" just prior to introducing the independent variable (WWC, 2017, p. A-5-6) means that baseline data collection will be unnecessarily extended for each case from the second on.

Under the current multiple probes pilot SCD standard, the first case requires a minimum of three consecutive probe points. The second case requires a minimum of six or seven probe points (three within the first three sessions and three just prior to introducing the independent variable, and if needed one probe point when another case first receives the intervention or meets prespecified intervention criterion). The third case requires a minimum of seven, eight or more probe points, and so on. Thus, the previously discussed issues of: (a) shorter baselines being logical and appropriate when there is a priori evidence that an individual is already performing at a stable, undesirable level; (b) ethical issues of potential harm to participants, (c) threats to internal validity; and (d) reactivity to assessment apply to all cases after the first case (Barlow et al., 2009; Council for Exceptional Children, 2014; Harris et al., 2019; Horner & Baer, 1978; Ledford et al., 2018; Ledford & Gast, 2018; McKeown et al., 2015).

The number of multiple probes required to meet Pilot SCD Standards Without Reservations can be meaningfully reduced by maintaining the requirement for a probe point in a session "where another case either (a) first receives the intervention or (b) reaches the prespecified intervention criterion" and eliminating the requirement for three consecutive probe points for each case within the first three sessions for the second and succeeding cases. As previously mentioned, this additionally requires for the second and succeeding cases that: (a) stability is achieved in a minimum of three data points just prior to introduction of the independent variable; (b) there is a priori evidence that an individual is already performing at a stable, undesirable level; (c) ethical issues indicate potential harm to participants; and (d) unnecessarily prolonged data collection in baseline can result in reactivity to assessment and meaningful threats to the internal validity of a study.

Footnote in the WWC SCD Pilot Standards Regarding Potential Exceptions

The current WWC SCD Pilot Standards (2017) include a footnote regarding potential exceptions to the standard of five data points per phase, "If the topic area team leadership determines that there are exceptions to this standard, these will be specified in the topic area or practice guide protocol (e.g., extreme self-injurious behavior might warrant a lower threshold of only one or two data points)" (footnote 29, p. A-5). Similarly, the current pilot standards include a footnote regarding potential exceptions to the criteria for multiple probe designs, "If the topic area team leadership determines that there are exceptions to these standards, they will be specified in the topic area or practice guide protocol (e.g., conditions when stable data patterns necessitate collecting fewer than three consecutive probe points just prior to introducing the intervention or when collecting overlapping initial pre-intervention points is not possible)" (footnote 30, p. A-5).

Leaving such a decision to team leadership or a panel evaluating specific studies creates additional concerns. In the design and conduct of research, and in seeking research funding, it is critical that researchers and others know the standards before the conduct of research. Waiting to see what a panel or team leadership might determine to be, or not to be, justified exceptions to the two standards targeted here after the research has been conducted, published, and then reviewed by a WWC panel undermines research planning and leaves researchers and funders uncertain about the ability of studies to meet standards. Criteria for meeting standards without reservations, with reservations, and not meeting standards need to be operationalized clearly and transparently.

Impacts on Editors, Reviewers, Research Publication, and the Field

The WWC standards were developed for evaluating the extent and quality of research in education. Such standards aid rigorous, transparent review and provide meaningful information to professionals, parents, funding organizations, and others. Presumably, WWC standards were not intended for use as a part of the scientific gatekeeping mechanism in peer-review research journals. Nevertheless, WWC standards have gained attention and prestige in and out of the research community, and thus influence the decisions of reviewers and journal editors. Further, the importance of these standards is stressed among educational professionals and organizations. The distinction between "meets WWC pilot SCD standards without reservations" and "meets WWC pilot SCD standards with reservations" is, therefore, critical. Considering high quality SCD studies to meet WWC pilot SCD standards with reservations based on the requirement of a minimum of five data points per phase when this is neither appropriate nor necessary misrepresents both single studies and important bodies of research.

Editors of some journals, as well as reviewers, have indicated that they are not willing to publish SCD studies that do not meet the five data points per phase minimum or the requirement of three consecutive probe points for each case within the first three sessions in a multiple probes design, when in all other ways the study is sound and the results are meaningful. This position is held despite the fact that these standards are still "pilot" standards. There are study characteristics that may critically compromise the scientific integrity of any given SCD study and warrant a decision of meeting the WWC Pilot SCD Standards with reservations or not meeting the pilot standards. Inclusion of three to four data points per phase instead of five, however, should not be a standard

for meeting pilot standards with reservations, for the reasons articulated previously here. Similarly, requiring three consecutive probe points for each case within the first three sessions in a multiple probes design is not necessary when baseline performance is low and stable across multiple probes for each case succeeding the first and three probes prior to initiating intervention.

Finally, journal editors, reviewers, and authors should consider the issues and concerns raised here. Editors, as many already do, should establish and post transparent standards for their journals for SCD research, communicate clearly with authors and their reviewers what their standards are for SCD research submitted to their journal, and follow up with reviewers who apply inappropriate standards. It is important that researchers, including journal editors and reviewers, join in a robust discussion of whether the current pilot WWC SCD standards need to be revised or whether aspects of these pilot standards should be deemed inappropriate for manuscript review based on decades of SCD research in special education (cf. Huitema, 1986; Kennedy, 2005; Kratochwill & Levin, 2010; Ledford, 2017; Wolery, 2013). The purpose of such discussion is to create deep thinking, discuss possible solutions, and fine-tune existing standards, rather than to polarize positions.

Proposed Changes to the WWC Pilot Single Case Design Standards

Changes that alleviate the issues raised here are proposed. First, rather than an arbitrary standard of five data points per phase in multiple baseline research focused on instructional interventions, WWC Pilot SCD Standards should explicitly recognize and define the conditions in which fewer data points can be used without compromising the scientific integrity of the study. As discussed here, standards and safeguards that can be used to increase confidence in the level, trend, and stability of data when less than five points have been collected include: (a) stability is achieved in three to four data points in baseline and following phases; (b) there is a priori evidence that an individual is already performing at a stable, undesirable level; (c) ethical issues indicate potential harm to participants; and (d) unnecessarily prolonged data collection in baseline can result in reactivity to assessment and meaningful threats to the internal validity of a study. When researchers have demonstrated stability in three or four data points, and one or more of the additional standards or safeguards listed here have been met, the number of data points per phase in multiple baseline design studies should meet WWC pilot standards without reservations.

In terms of the multiple probes requirement for three consecutive probe points for each case within the first three sessions, the same issues, standards and safeguards noted for the requirement for five data points per phase can be applied. The WWC Pilot SCD Standards should state that a minimum of three consecutive probe points are required for the first case (if stability is demonstrated). For each case following the first case, the design must include a minimum of three consecutive probe points before intervention is initiated as well as one probe just prior to the introduction of the intervention for preceding cases. Thus, with three baselines, the first case would have a minimum of three data points before intervention. The second case would have a minimum of one data point taken concurrently with the three baseline probes taken for case one and a minimum of three data points taken just before intervention (resulting in a minimum of four data points during baseline for the second case). The third case would have a minimum of one additional probe taken concurrently with the three baseline probes taken for case two and a minimum of three probes taken just prior to introduction of the independent

variable (resulting in a minimum of 5 baseline probes taken for case 3; and so on for succeeding cases). Although a total of five or more probes may be necessary for some participants, the concerns noted regarding extended probing are frequently alleviated due to the probes being spaced further apart in time (cf. Harris et al., 2019). An example is provided in Figure 1. If issues with extended data collection are not alleviated in these cases, researchers must consider study design further and, as always, carefully monitor participants during data collection.

The WWC has striven hard to be "a central and trusted source of scientific evidence for what works in education" (WWC, 2017, p. 1). The SCD standards developed by the WWC are highly beneficial to our field, but these pilot standards require further evaluation and considerations. By addressing the issues and concerns discussed here regarding the WWC SCD Pilot Standards regarding SCD, and making the suggested changes, the current WWC SCD standards will be improved, potential negative effects on research and participants in SCD research will be assuaged, and confidence in the SCD standards among the research community will be enhanced.

References

- Barlow, D.H., Nock, M.K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change* (3rd ed.). London, England: Pearson.
- Council for Exceptional Children Standards for Evidence-Based Practices in Special Education (2014). Retrieved from:
<https://www.cec.sped.org/~media/Files/Standards/Evidence%20based%20Practices%20and%20Practice/EBP%20FINAL.pdf>
- Graham, S., & Harris, K.R., & Zito, J. (2005). Promoting internal and external validity: A synergism of laboratory experiments and classroom based research. In G. Phye, D. H. Robinson, J. Levin (Eds.), *Experimental methods for educational intervention* (pp. 235-265). San Diego, CA: Elvieser.
- Harris, K.R. (2015). Message from the President - The single-case experiment: A design study on steroids? *NEP/15 (Newsletter, APA Division 15: Educational Psychology)*, pps. 1-2
DOI: <https://apadiv15.org/wp-content/uploads/2015/04/Message-From-the-President-Spring-NEP.pdf>
- Harris, K.R., Ray, A., Graham, S., & Houston, J. (2019). Answering the challenge: SRS D instruction for close reading of text to write to persuade with 4th and 5th grade students experiencing writing difficulties. *Reading and Writing: An Interdisciplinary Journal*, 32(6), pps. 1459-1482.
- Horner, R.D., & Baer, D.M. (1978). Multiple-probe technique: A variation on the multiple baseline. *Journal of Applied Behavior Analysis*, 11, 189-196.
- Huitema, B.E. (1986). Statistical analysis and single-subject designs. In A. Poling & R.W. Fuqua (Eds.) *Research Methods in Applied Behavior Analysis* (pp. 209-232). Springer, Boston, MA. Doi: https://doi.org/10.1007/978-1-4684-8786-2_10
- Hwang, Y., Levin, J.R., & Johnson, E.W. (2018). Pictorial mnemonic-strategy interventions for children with special needs: Illustration of a multiply randomized single-case crossover design, *Developmental Neurorehabilitation*, 21(4), 223-237. doi:
<https://doi.org/10.3109/17518423.2015.1100689>
- Kennedy, C.H. (2005). *Single-case designs for educational research*. Boston: Allyn and Bacon.
- Kratochwill, T.R., & Levin, J. (2010). Enhancing the scientific credibility of single-case design intervention research: Randomization to the rescue. *Psychological Methods*, 15, 124-144.
- Ledford, J.R. (2017). No randomization? No problem. Experimental control and random assignment in single case research. *American Journal of Evaluation*. DOI:
<https://journals.sagepub.com/doi/abs/10.1177/1098214017723110?journalCode=ajec#>
- Ledford, J.R., & Gast, D.L. (Eds.) (2018). *Single case research methodology: Applications in special education and behavioral sciences* (3rd ed.). New York: Routledge.

- Ledford, J.R., Zimmerman, K.N., Schwartz, I., & Odom, S. (2018). CEC – Division for Research Guide for the Use of Single Case Design Research Evidence. Retrieved from: https://higherlogicdownload.s3.amazonaws.com/SPED/b7acd4b4-bc4d-4c1f-a7d4-efab3d52da44/UploadedImages/Evidence%20Based%20Practice/CEC-DR_SCD_Guide.pdf
- Lloyd, J. W. Saltzman, N. J., & Kauffman, J. M. (1981). Predictable generalization in academic learning as a result of preskills and strategy training. *Learning Disability Quarterly*, 4, 203-216.
- McKeown, D., Kimball, K., & Ledford, J. (2015). Effects of asynchronous audio feedback on the story revision practices of students with emotional/behavioral disorders. *Education and Treatment of Children*, 38(4), 541–564.
- National Center for Education Statistics, Statistics of Public Elementary and Secondary School Systems (2017), 1977-78 and 1980-81; and Common Core of Data (CCD), "State Nonfiscal Survey of Public Elementary/Secondary Education," 1990-91 through 2015-16. Retrieved https://nces.ed.gov/programs/digest/d17/tables/dt17_204.30.asp
- Plavnick, J. B., & Ferreri, S. F. (2013). Single-case experimental designs in educational research: A methodology for causal analyses in teaching and learning. *Educational Psychology Review*, 25, 549-569.
- Rodgers, W., Lewis, T., O'Neill, R., & Vannest, K. (2018). Policy and Position Statement on Single Case Research and Experimental Designs from the Division for Research Council for Exceptional Children. Retrieved from: https://higherlogicdownload.s3.amazonaws.com/SPED/b7acd4b4-bc4d-4c1f-a7d4-efab3d52da44/UploadedImages/Evidence%20Based%20Practice/CEC-DR_SCD_Policy.pdf
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York: Basic Books.
- Wolery, M. (2013). A commentary: Single-case design technical document of the What Works Clearinghouse. *Remedial and Special Education*, 43 (1), 39-43.
- What Works Clearinghouse (2017). What works clearinghouse standards handbook (Version 4.0). Retrieved from: https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Figure 1 (from Harris et al., 2019)

