Abstract Title Page
Not included in page count.

Title: What Works Clearinghouse (WWC) Standards for Evaluating Single Case Designs (SCDs)

Author(s): Shannon Monahan, Thomas Kratochwill, and Stephen Lipscomb

Abstract Body

Limit 5 pages single spaced.

Background / Context:

The What Works Clearinghouse (WWC) seeks to provide educators, policymakers, researchers, and the public with a central and trusted source of scientific evidence for what works in education. The WWC was established in 2002 by the U.S. Department of Education's Institute of Education Sciences (IES). It serves as a decision-making resource by helping the education community locate and recognize credible and reliable evidence. Reviewers working for the WWC assess the quality of studies that evaluate the effectiveness of specific educational products, interventions, practices or approaches. To ensure fairness and consistency in the review of these studies, the WWC uses rigorous and transparent standards for systematically reviewing and synthesizing existing research. From its inception, the WWC reviewed experimental and quasi-experimental group comparison studies. In an effort to expand the pool of rigorous scientific evidence available for review, in 2009, the WWC assembled a panel of national experts in single-case design (SCD) and analysis to develop WWC SCD Standards.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

This presentation delineates the criteria the WWC uses to identify SCDs and the standards that the WWC uses to evaluate the rigor of SCD studies. The WWC SCD standards are bifurcated into Design and Evidence standards (see Appendix B, Figure 1). The Design standards evaluate the internal validity of the design. Reviewers trained in visual analysis then apply the Evidence standards to evaluate the effect of interventions and practices on relevant outcomes.

Significance / Novelty of study:

SCDs are adaptations of interrupted time-series designs and can provide a rigorous experimental evaluation of intervention effects (Horner & Spaulding, in press; Kazdin, 1982, in press; Kratochwill, 1978; Kratochwill & Levin, 1992; Shadish, Cook, & Campbell, 2002). As experimental designs, a central goal of SCDs is to determine whether a causal relation (i.e., functional relation) exists between the introduction of a researcher-manipulated independent variable (i.e., an intervention) and change in a dependent (i.e., outcome) variable (Horner & Spaulding, in press; Levin, O'Donnell, & Kratochwill, 2003). SCDs can provide a strong basis for establishing causal inference, and these designs are widely used in applied and clinical disciplines in psychology and education, such as school psychology and the field of special education. The WWC has recognized that single case research designs provide an alternative design approach that is especially valuable in situations that are not amenable to study through randomized trials (e.g., small numbers of participants available) and also provide the scientific rigor needed to document experimental control. The development of WWC SCD standards not only expands the pool of studies available for WWC review, it also offers guidance to the research community on the highest quality attributes to consider when designing these studies.

Statistical, Measurement, or Econometric Model:

Although the basic SCD has many variations, these designs often involve repeated, systematic measurement of a dependent variable before, during, and after the active manipulation of an independent variable (e.g., applying an intervention). Experimental control involves replication of the intervention in the experiment and this replication is addressed with one of the following methods (Horner, et al., 2005):

- Introduction and withdrawal (i.e., reversal) of the independent variable (e.g., ABAB design)
- Iterative manipulation of the independent variable across different observational phases (e.g., alternating treatments design)
- Staggered introduction of the independent variable across different points in time (e.g., multiple baseline design)

The WWC SCD standards are intended to guide WWC reviewers in identifying and evaluating the rigor of SCDs as well as characterizing effects on outcomes. The first section of the WWC SCD standards assists with identifying whether a study is a SCD. As depicted in Appendix B Figure 1, a SCD should be reviewed using the 'Criteria for Designs that Meet Evidence Standards', to determine those that *Meet Evidence Standards,* those that *Meet Evidence Standards with Reservations,* and those that *Do Not Meet Evidence Standards.*

Studies that meet evidence standards (with or without reservations) should then be reviewed using the 'Criteria for Demonstrating Evidence of a Relation between an Independent Variable and a Dependent Variable' (see Appendix B, Figure 1). This process results in a categorization scheme that is similar to that used for evaluating evidence credibility by inferential statistical techniques (hypothesis testing, effect-size estimation, and confidence-interval construction) in traditional group designs. This review results in a sorting of SCD studies into three groups: those that have *Strong Evidence of a Causal Relation,* those that have *Moderate Evidence of a Causal Relation,* and those that have *No Evidence of a Causal Relation.*

CRITERIA FOR DESIGNS THAT MEET EVIDENCE STANDARDS

In order to *Meet Evidence Standards,* the following design criteria must be present:

- The independent variable (i.e., the intervention) must be systematically manipulated, with the researcher determining when and how the independent variable conditions change.
- Each outcome variable must be measured systematically over time by more than one assessor, and the study needs to collect inter-assessor agreement in each phase and on

at least twenty percent of the data points in each condition (e.g., baseline, intervention) and the inter-assessor agreement must meet minimal thresholds.

- The study must include at least three attempts to demonstrate an intervention effect at three different points in time or with three different phase repetitions.[*]
- For a phase to qualify as an attempt to demonstrate an effect, the phase must have a minimum of three data points and preference is given to 5 data points per phase.

CRITERIA FOR DEMONSTRATING EVIDENCE OF A RELATION BETWEEN AN INDEPENDENT VARIABLE AND AN OUTCOME VARIABLE

Single-case researchers traditionally have relied on visual analysis of the data to determine (a) whether evidence of a relation between an independent variable and an outcome variable exists; and (b) the strength or magnitude of that relation (Hersen & Barlow, 1976; Kazdin, 1982; Kennedy, 2005; Kratochwill, 1978; Kratochwill & Levin, 1992; McReynolds & Kearns, 1983; Richards, Taylor, Ramasamy, & Richards, 1999; Tawney & Gast, 1984; White & Haring, 1980). An inferred causal relation requires that changes in the outcome measure resulted from manipulation of the independent variable. A causal relation is demonstrated if the data across all phases of the study document at least three demonstrations of an effect at a minimum of three different points in time. An effect is documented when the data pattern in one phase (e.g., an intervention phase) differs more than would be expected from the data pattern observed or extrapolated from the previous phase (e.g., a baseline phase; Horner et al., 2005).

WWC rules for conducting visual analysis involve four steps (Kratochwill et al., 2010; Parsonson & Baer, 1978). The first step is documentation of a predictable baseline pattern of data. If a convincing baseline pattern is documented, then the second step consists of examining the data within each phase of the study to assess the within-phase pattern(s). The key question is to assess whether there are sufficient data with sufficient consistency to demonstrate a predictable pattern of responding. The third step in the visual analysis process is to compare the data from each phase with the data in the adjacent (or similar) phase to assess whether manipulation of the independent variable was associated with an "effect." An effect is demonstrated if manipulation of the independent variable is associated with predicted change in the pattern of the dependent variable. The fourth step in visual analysis is to integrate all the information from all phases of the study to determine whether there are at least three demonstrations of an effect at different points in time (i.e., documentation of a causal or functional relation) (Horner et al., in press).

For studies that meet WWC evidence standards (with and without reservations), the following rules are used to determine whether the study provides evidence of a causal relation. In order to provide *Strong Evidence*, at least two WWC reviewers certified in visual (or graphical) analysis must verify that a causal relation was documented. Specifically this is operationalized as at least three demonstrations of the intervention effect along with no non-effects by:

[*] The three demonstrations criterion is based on professional convention (Horner, Swaminathan, Sugai, & Smolkowski, under review). More demonstrations further increase confidence in experimental control (Kratochwill & Levin, in press).

- Documenting the consistency of level, trend, and variability within each phase
- Documenting the immediacy of the effect, the proportion of overlap, the consistency of the data across phases in order to demonstrate an intervention effect, and comparing the observed and projected patterns of the outcome variable
- Examining external factors and anomalies (e.g., a sudden change of level within a phase)

Overall, the WWC SCD Standards provide guidance to reviewers for evaluating the internal validity of a study and for rating the strength of the evidence presented.

Usefulness / Applicability of Method:

The goal of a SCD is usually to answer "Is this intervention more effective than the current "baseline" or "business-as-usual" condition?" SCDs are implemented when pursuing the following research objectives (Horner et al., 2005):

- Determining whether a causal relation exists between the introduction of an independent variable and a change in the dependent variable. For example, a research question might be "Does Intervention B reduce a problem behavior for this case (or these cases)?"
- Evaluating the effect of altering a component of a multi-component independent variable on a dependent variable. For example, a research question might be "Does adding Intervention C to Intervention B further reduce a problem behavior for this case (or these cases)?"
- Evaluating the relative effects of two or more independent variables (e.g., alternating treatments) on a dependent variable. For example, a research question might be "Is Intervention B or Intervention C more effective in reducing a problem behavior for this case (or these cases)?"

SCDs are especially appropriate for pursuing research questions in applied and clinical fields. This application is largely because disorders with low prevalence may be difficult to study with traditional group designs that require a large number of participants for adequate statistical power (Odom, et al., 2005). Currently, as the WWC has expanded into special education topic areas (e.g., Children Classified with an Emotional or Behavioral Disorder), the number of SCDs identified in literature reviews has increased. In some instances, the evidence base for a single intervention may consist of over 50 SCDs. The inclusion of SCDs for review gives the WWC access to this wide body of relevant research.

Conclusions:

The development of WWC SCD Standards serves two primary purposes. First, it expands the pool of rigorous research that can be reviewed by the WWC and that will ultimately be evaluated to inform educational decision making. Second, it offers guidelines to the research community at large of best practice in single case design.

A noted limitation of the current WWC SCD standards is the singular focus on visual analysis and the inability to calculate effect-size (ES). Most researchers using SCDs base their inferences on visual analysis, and there are no agreed-upon methods or standards for effect size estimation though several parametric and non-parametric quantitative methods have been proposed. Each quantitative method has flaws, and most are not comparable with those used in group-comparison studies. Shadish et al. (2008) have developed an estimator for continuous outcomes that is promising, though the distribution theory is still being derived and tested. As the field reaches greater consensus about appropriate statistical analyses and quantitative effect-size measures, new WWC standards for effect demonstration will need to be developed.

Appendices

Not included in page count.

Appendix A. References

Hersen, M., & Barlow, D. H. (1976). *Single-case experimental designs: Strategies for studying behavior change*. New York: Pergamon.

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(2), 165–179.

Horner, R., & Spaulding, S. (in press). Single-Case Research Designs. Encyclopedia. Springer.

Horner, R., Swaminathan, H., Sugai, G., & Smolkowski, K. (in press). Expanding analysis of single case research. Washington, DC: Institute of Education Science, U.S. Department of Education.

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

Kennedy, C. H. (2005). *Single-case designs for educational research*. Boston: Allyn and Bacon.

Kratochwill, T. R. (Ed.). (1978). *Single subject research: Strategies for evaluating change*. New York: Academic Press.

Kratochwill, T. R., & Levin, J. R. (Eds.). (1992). *Single-case research design and analysis: New directions for psychology and education*. Hillsdale, NJ: Erlbaum.

Kratochwill, T. R., & Levin, J. R. (in press). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods*.

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 What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Levin, J. R., O'Donnell, A. M., & Kratochwill, T. R. (2003). Educational/psychological intervention research. In I. B. Weiner (Series Ed.) and W. M. Reynolds & G. E. Miller (Vol. Eds.). *Handbook of psychology: Vol. 7. Educational psychology* (pp. 557–581). New York: Wiley.

McReynolds, L. & Kearns, K. (1983). *Single-subject experimental designs in communicative disorders*. Baltimore: University Park Press.

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children* 71(2), 137–148.

Parsonson, B., & Baer, D. (1978). The analysis and presentation of graphic data. In T. Kratchowill (Ed.) *Single Subject Research* (pp. 101–166). New York: Academic Press.

Richards, S. B., Taylor, R., Ramasamy, R., & Richards, R. Y. (1999). *Single subject research: Applications in educational and clinical settings*. Belmont, CA: Wadsworth.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin

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*, 3, 188–196.

Tawney, J. W., & Gast, D. L. (1984*). Single subject research in special education*. Columbus, OH: Merrill.

White, O. R., & Haring, N. G. (1980). *Exceptional teaching* (2nd ed.). Columbus, Ohio: Charles E. Merrill.

Appendix B. Tables and Figures

FIGURE 1

PROCEDURE FOR APPLYING SCD STANDARDS: FIRST EVALUATE DESIGN, THEN IF APPLICABLE, EVALUATE EVIDENCE