Research in Developmental Disabilities xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Research in Developmental Disabilities

journal homepage: www.elsevier.com/locate/redevdis

Between-case standardized effect size analysis of single case designs: Examination of the two methods

Samuel L. Odom[a,1], Erin E. Barton[b,*,1], Brian Reichow[c,1], Hariharan Swaminathan[d,1], James E. Pustejovsky[e,1]

[a] University of North Carolina at Chapel Hill, United States
[b] Vanderbilt University, United States
[c] University of Florida, United States
[d] University of Connecticut, United States
[e] University of Texas at Austin, United States

ARTICLE INFO

Number of reviews completed is 2

Keywords:
Single case research
meta-analysis
single case effect sizes

ABSTRACT

An increasing movement in single case research is to employ statistical analyses as one form of data analysis. Researchers have proposed different statistical approaches. The purpose of this paper is to examine the utility and discriminant validity of two novel types of between-case standardized effect size analyses with two existing systematic reviews. The between-case analyses found greater effect sizes for the studies in the object play review and smaller effect sizes for studies of sensory intervention, which were consistent with the overall conclusions reached in the original systematic reviews. These findings provide evidence of discriminant validity, although concerns remain around the methods' utility across different single case research designs. Future directions for research and development also are provided.

What this paper adds

Our paper provides initial support for the between-case effect size (ES_{BC}) approach, although more research is needed. The availability of a feasible approach for estimating effect sizes that is similar in logic to the effect sizes used group design efficacy studies might facilitate the use of single case design research in meta-analytic reviews, from which they are now often excluded.

1. Introduction

Single case design (SCD) has been the foundational experimental methodology in applied behavior analysis since the inception of the field (Baer, Wolf, & Risley, 1968). SCD researchers focus on examining change over time in individual participants. This contrasts with the typical focus of experimental group design research, which focuses on *average* effects *across* participants. Because of this conceptual difference, SCD researchers have struggled with establishing effect size measures for SCD research that are conceptually consistent with experimental group design research. Given that bodies of research on specific interventions often include SCD and group design studies, a between-participants approach may allow for better integration of evidence from the two methodological

* Corresponding author.
E-mail address: erin.e.barton@vanderbilt.edu (E.E. Barton).

[1] All authors conceptualized the paper and reviewed and approved the final manuscript. Swaminathan and Pustejovsky conducted all statistical analyses. Odom, Barton, and Reichow drafted the introduction and discussion sections. Odom, Barton, Swaminathan, and Pustejovsky drafted the method section. Odom, Barton, Reichow, Swaminathan, and Pustejovsky drafted the results section.

https://doi.org/10.1016/j.ridd.2018.05.009
Received 22 October 2017; Received in revised form 11 May 2018; Accepted 15 May 2018
0891-4222/

Please cite this article as: Odom, S.L., Research in Developmental Disabilities (2018), https://doi.org/10.1016/j.ridd.2018.05.009

approaches. The purpose of this paper was to examine the utility and validity of two new analytic approaches that generate standardized between case effect sizes (ES_{BC}) for SCD studies.

1.1. Critical features of single case design methodology, calculation of effect size, and meta-analysis

In SCD, researchers establish plausible causal inferences through replication of the demonstration of functional relations between the independent and dependent variable. Sidman (1960) proposed that research be guided by baseline logic, which involves the repeated measurement of behavior(s) under at least two conditions. When the data patterns in one condition differs from what is predicted based on the preceding condition, behavior change is demonstrated. Current standards require at least three consistent demonstrations of behavior change at three different points in time to demonstrate a functional relation (e.g., Horner et al., 2005; Kratochwill et al., 2013).

Most previous approaches to computing effect sizes for SCD studies have based the metrics on *differences within a single participant* (i.e., changes across conditions of an individual participant within a study or experimental manipulation). Statistical analysis of SCD data has unique features not encountered in the analysis of experimental group design data. Because SCD data are collected on individual participants or cases across time, autocorrelation and serial dependence may occur in the data, making traditional inferential statistics used in group design research not directly applicable (Tawney & Gast, 1984). Further, SCDs typically involve only a small number of participants and a small or moderate number of repeated measurements on each individual (Shadish & Sullivan, 2011); this precludes the use of time series methods that require scores of measurement occasions. For decades, researchers have proposed different approaches for conducting statistical analysis of SCD (e.g., Kazdin, 1976; Shadish, 2014). In a recent review commissioned by the U.S. Department of Education Institute of Education Sciences to provide guidance on statistical analysis of SCD studies, Shadish, Hedges, Horner, and Odom, 2015 proposed several advantages of calculating ES_{BC} rather than within case effect sizes (ES_{WC}). The major advantage of ES_{BC} is that it places the SCD effect size on a metric that is comparable to between subjects effect size (such as Cohen's *d*) which permits aggregation across studies. The ES_{BC}, however is relatively new and additional studies are needed to assess its capacity to accurately quantify and aggregate magnitude of effects across studies.

1.2. Visual analysis standards and limitations

SCD research has a long history of being used to evaluate behavioral interventions. Three primary design options are the withdrawal or A-B - A-B design, multiple baseline or probe design, and the alternating treatment design (Gast & Ledford, 2014). Multiple baseline and multiple probe designs can involve replication across behaviors or settings for a single participant or replication across participants. Across these designs, independent variable conditions are ordered in three different ways, respectively: (a) introduction and withdrawal, (b) time lagged, (c) rapid implementation across observations. SCD researchers present the repeated measurement of the dependent variable data graphically (i.e., rather than grouped as averages) to allow for independent judgments about data patterns and the presence of functional relations. Traditionally, SCD researchers have used visual analysis of graphed data as the primary mode of data analysis (Parsonson & Baer, 1978). Current visual analysis guidelines describe four characteristics of data patterns to examine within and across adjacent conditions: level, trend, stability, and variability (Horner et al., 2005; Kazdin, 2011; Kratochwill et al., 2013; Ledford & Gast, 2018). Three additional features are immediacy of the behavior change, overlap in data points, and consistency of the magnitude of behavior change or data patterns (Barton, Lloyd, Spriggs, & Gast, 2018).

In experimental group design research, researchers have established standardized effect sizes for group design studies (e.g., Cohen's *d*, Hedges' *g*), primarily basing this analysis on *average performance differences between participants* in contrasting groups (e.g., treatment and control). Visual analysis confirms that a functional relation exists and allows for judgments related to the magnitude of behavior change for individual participants. However, visual analysis does not provide a metric for quantifying treatment effects, which is important for aggregating findings across studies to inform the identification of evidence-based practices. This is critical in special education. The *Individuals with Disabilities Education Improvement Act* (IDEIA, 2004) and the *Every Student Succeeds Act* (ESSA, U. S. Department of Education, 2015) calls for the use of "scientifically based research" as the basis for programs for students with disabilities. Meta-analytic methods for group research are well established and address this need. However, meta-analytic procedures for SCD are still evolving (Shadish, 2014).

1.3. Rationale for calculating effect sizes

As noted previously, summary judgments about the presence or absence of functional relations do not speak to the magnitude of the effect. Standardized effect size calculations are designed to reflect quantitatively the magnitude (or robustness) of effects. In experimental group designs, the effect size represents relative differences in mean performances of individuals in treatment and control groups (Glass, McGaw, & Smith, 1981), with the classic calculation of an effect size being Cohen's *d* statistic (Cohen, 1977). The *d* statistic is computed by subtracting posttest mean of the control group from the posttest mean of the treatment group, divided by a standard deviation, which can vary but is often a pooled standard deviation of the mean standard deviation of the posttest treatment and control groups. Because the *d* statistic overestimates effect size for small samples, Hedges (1981) established a correction for small samples (Hedges' *g*). Both Cohen's *d* and Hedges' *g* are standardized mean effect sizes and interpretations are identical.

For SCDs, the analogous approach would seem to be to treat data points in an A and B condition as if they were individuals, compute the mean of each condition, subtract Condition A from Condition B and divided by the pooled averaged standard deviation.