International Journal of Research & Method in Education

ISSN: 1743-727X (Print) 1743-7288 (Online) Journal homepage: http://www.tandfonline.com/loi/cwse20

The educational validity and utility of single-case design research in building evidence-based practices in education

Lefki Kourea & Ya-yu Lo

INTERNATIONAL JOURNAL OF RESEARCH & METHOD IN EDUCATION, 2016
http://dx.doi.org/10.1080/1743727X.2016.1160278

The educational validity and utility of single-case design research in building evidence-based practices in education

Lefki Kourea and Ya-yu Lo

Department of Special Education and Child Development, University of North Carolina at Charlotte, Charlotte, NC, USA

ABSTRACT

Improving academic, behavioural, and social outcomes of students through empirical research has been a firm commitment among researchers, policy-makers, and other professionals in education across Europe and the United States (U.S.). To assist in building scientific evidences, executive bodies such as the European Commission and the Institute for Education Sciences at the U.S. Department of Education have established systematic dialogues and expert panels to determine the rigour, relevance, and impact of educational practices and supports as investigated through a range of research methodologies. In this paper, we give an overview of the single-case design (SCD) research, which is part of the quantitative experimental research and is used widely in education and applied behaviour analysis. First, we describe the philosophical assumptions and defining features of SCD methodology and we explain how SCD researchers exert control for data believability. Second, we review criteria suggested by task-force panels (e.g. European Platform for Investing in Children, What Works Clearinghouse, Council for Exceptional Children) to determine evidence-based practices (EBPs). Finally, we conclude with strengths and criticisms addressed for SCD methodology and we discuss implications on how SCD research would contribute on improving learning outcomes of all students and advancing the knowledge base of EBPs.

ARTICLE HISTORY
Received 29 June 2015
Accepted 9 February 2016

KEYWORDS
Evidence-based practices; experimental research; single-case research

Introduction

Improving academic, behavioural, and social outcomes of students, especially those with identified risk markers (e.g. low academic performance, low socio-economic family background, ethnic minority status, disability status) has been a strong commitment among researchers, policy-makers, and other professionals in education across Europe and the United States (U.S.). Current U.S. legislations for students with disabilities (i.e. Individuals with Disabilities Education Improvement Act of 2004) and for students in Elementary and Secondary Education (i.e. No Child Left Behind Act 2001 [NCLB]) mandate the identification and use of instructional strategies, which are supported by credible research and thus have been shown to improve student outcomes positively. Likewise, the European Commission (EC 2013) in its latest child policy recommendation entitled 'Investing in children – breaking the cycle of disadvantage' asks Member States to follow an evidence-based decision approach in their policy-making for improving the lives of students and their families (Kilburn and Mattox 2014). This is the first European recommendation in the history of the European Union (EU) that guides Member States to establish an evidence-based approach for supporting students with risk markers effectively (van Stolk and Kilburn n.d.).

CONTACT Lefki Kourea lkourea@uncc.edu Department of Special Education and Child Development, University of North Carolina at Charlotte, 9201 University City Boulevard, Charlotte, NC 28223, USA

Policy efforts for a data-based movement in education have been influenced by the pioneering work of professionals in the medical field, who have followed an evidence-based approach to improve their own practices (Cutspec 2004; Dunst, Trivette, and Cutspec 2002). By designing and implementing randomized controlled trials (RCTs), medical researchers want to ensure that any new medical product (e.g. screening test, surgery technique, or drug) is likely to produce a positive behavioural change on research participants. The RCTs allow researchers to maximize the impact of their product by using randomization that would minimize the interference of any extraneous variables under controlled clinical conditions (Morris 2004). The telos of medical scientists has been to provide an evidence-based product that would maintain its positive impact over time and across people and settings.

Policy stakeholders have embraced the use of RCTs as the 'gold standard' for the identification and dissemination of evidence-based practices (EBPs) in education. Both the American legislation (i.e. NCLB) and the European Platform for Investing in Children (EPIC) – the evidence-based online platform developed by the RAND Europe research group on behalf of the EC (van Stolk and Kilburn n.d.) – favour the use of randomized group designs as these designs would provide credible evidence for identifying effective educational practices. However, since its enactment the NCLB has been criticized for the strong emphasis on RCTs and an ongoing debate has sparked among educational researchers and professionals on topics related to 'who sets the guidelines for deciding on EBPs' 'how should actually EBPs be determined' 'who is influenced by' and 'what the long-term effects might be' (McDonnell and O'Neill 2003). Cutspec (2004) cautions about a potential limitation of using RCTs as the sole criterion for determining EBPs in education. He contends that unlike most of the medical research, which is conducted under highly controlled conditions, educational scientists conduct their research in applied settings, where school conditions do not easily lend themselves to randomization in a similar manner as lab conditions do. Therefore, focusing solely on RCTs may underestimate the validity and utility of other experimental research approaches that do not incorporate randomization per se. In fact, Shavelson et al. (2003) argue that despite efforts from policymakers to promote RCTs, design studies (i.e. experimental research) are suitable in education based on the questions they address.

Developing a mutual understanding at policy and scientific levels on how EBPs should be defined and reaching to a consensus on 'what works' in education is a slow plodding process. A beginning and critical step to this process is to focus on the type of research questions asked to determine the level of scientific inquiry and verification. Research questions are important because they set the theoretical framework in which research approaches, designs and data analyses are employed. As Table 1 shows, there are three major types of research questions: descriptive, causative, and mechanism/process (NRC 2002).

Descriptive questions aim at exploring and/or describing a set of variables for a particular population, problem, or theory. To this end, research approaches that could be followed involve surveys, developmental studies (e.g. cohort, trend, or panel studies), or qualitative ones (e.g. ethnography,

Table 1. Demonstrating EBPs based on the level of scientific inquiry.

	Level of scientific inquiry		
	Descriptive	Causative	Mechanism
Focus Research approach/ research design	'What?' • Survey • Correlational • Qualitative	'What are the effects of X on Y?' • Randomized experimental group • Quasi-experimental group • Single-subject (or single-case, or $n = 1$)	'How?' 'Why?' • Qualitative • Single-subject (or single-case, or $n = 1$)
Outcome	Identify descriptive characteristics to a problem/theory	Identify causal relationships	Identify important elements contributing to theory/programme

case studies). Descriptive questions also focus on explaining simple associations among variables (e.g. concurrent, predictive or ex post facto studies).

Causative questions inquire about cause-and-effect relationships between variables of interest. The investigation of causality is often times decided upon strong theoretical background as well as descriptive information provided by correlational and/or descriptive studies. In education, when randomization is feasible participants are randomly assigned to one of the two levels of the independent variable (treatment vs. control group). With randomization, groups become equal to begin with and any extraneous variables that might interfere in explaining causal changes in the outcomes are minimized or ruled out (Campbell and Stanley 1963). When randomization is not permissible, educational researchers design other types of experimental designs at the group level (quasi-experimental studies) or at the individual level (single-subject or single-case design [SCD] research). In such experimental studies, the sample is purposive.

Mechanism or process questions focus on determining reasons underlying causal relationships. In other words, after a cause-and-effect relationship has been established, scientists may inquire 'how' or 'why' one group performed better than the other. However, searching for underlying mechanisms does not always require pre-established causal effects. Often times, scientists may also employ studies for explaining and/or predicting those causal links related to variables under study or they may want to determine intervention components that could produce the maximum impact on people. 'Curriculum design studies' may well be a good fit in this category because the end goal for researchers is to explain how the new curriculum can improve student academic achievement.

All three categories are interrelated and contribute to the understanding of science and the advancement of scientific knowledge in educational research. However, determining 'what works' in education is highly important and relevant as teachers are continuously asked to do more in increasingly diverse classrooms. The 'what works' question denotes causality as it prompts researchers and policy-makers to examine the impact and efficacy of educational programmes (i.e. 'Is this instructional programme effective for these students?'). Based on Table 1, causative questions are answered by investigating cause–effect relationships with the manipulation of variables in a carefully controlled design. Experimental manipulation is included in the experimental group designs (i.e. true and quasi-experimental) as well as the experimental SCD research. So far, research and policy communities have been focusing on randomized group designs, thus disregarding other rigorous experimental research such as the SCD (Fletcher and Francis 2004; Horner 2006; Morris 2004; NRC 2002; Reyna 2004). The purpose of this paper is to provide a theoretical overview of the SCD methodology by describing its main philosophical assumptions and features and present SCD quality indicators for determining EBPs.

SCD research

In the early stages of the American debate on what constitutes EBPs in social sciences, SCD research was omitted from professional discussions (McDonnell and O'Neill 2003). Interestingly enough, during the last decade or so SCDs have been featured as an important experimental methodology in the EBPs movement (Horner et al. 2005; Kazdin 2011; Shadish et al. 2015). One plausible reason for SCD not having been recognized in this debate is that SCD research has not been widely used in general education compared to the experimental group design approach and thus it is less understood (Layng, Stikeleather, and Twyman 2004). Another consideration is that research and policy-making communities have been concentrating efforts on defining EBPs in social sciences following the hypothetico-deductive scientific tradition (Davies 1999) while omitting instructional practices deriving from an inductive approach (Plavnick and Ferreri 2013).

The SCD methodology derives from a scientific positivist tradition, the experimental analysis of operant behaviour (Skinner 1938), and since then SCD has been utilized to study human behaviour in applied settings (i.e. Science of Applied Behaviour Analysis; Cooper, Heron, and Heward 2007). SCD research has been widely used in areas such as special education, communication science and disorders (Byiers, Reichle, and Symonsa 2012) and school psychology (Kratochwill and Levin 2014; Kratochwill and Stoiber 2002). The purpose of SCD is to identify possible functional relationships between practices (i.e. independent variables) and human behaviour (dependent variables). Hence, SCD researchers are tasked to manipulate changes in the enviroment and examine the impact of programmes on socially significant behaviours. For further information about SCD methodology, readers may refer to these comprehensive resources: (Barlow, Nock, and Hersen 2008; Gast 2009; Kazdin 2011; Tawney and Gast 1984).

Philosophical assumptions of SCD research

To understand SCD research as a positivist inductive approach and how experimental control is exerted, one should first examine its philosophical underpinnings. First, behaviour exists only at the individual level. This means that groups do not behave but only individuals do (Cooper, Heron, and Heward 2007; Johnson and Pennypacker 1993). Environmental variables influence the individuals' behaviour and each person's behaviour varies differently. That is why researchers utilize SCD to examine individual behaviour. Second, behaviour is a continuous natural phenomenon that occurs in a constantly changing environment. It is clear that behaviour exists as long as the person lives (Johnson and Pennypacker 1993). Thus, continuous environmental changes influence behaviour-environment relations. The only way to observe and record the behavioural continuity is by measuring it repeatedly. That's why SCD researchers measure behaviour continuously over time. Note, however, that observation and recording of target behaviour takes place in selected times and locations. For instance, when attempting to observe student social interactions in school, researchers would select those time blocks and locations where target students would present problem social behaviours.

Third, behaviour is assumed to be determined. A central assumption in SCD is that the universe is a lawful and orderly place, in which natural phenomena are functionally related with other events. Similarly, behaviour – as a natural phenomenon – is related in an orderly way with other environmental variables. SCD allows researchers through specifically designed experiments to identify those functional relations between one variable (e.g. instructional reading programme) and another (e.g. reading fluency) while holding constant all other possible extraneous variables.

Finally, behavioural variability is assumed to be extrinsic to people (Johnson and Pennypacker 1993; Sidman 1960). As such, SCD researchers are able to demonstrate variability through experimental manipulation of an independent variable. When behavioural variability is documented under different conditions, then experimental control is exerted. Having set the philosophical assumptions, in which SCD researchers operate in, below is an overview of the SCD main characteristics. For a detailed description of the SCD assumptions as well as SCD features, interested readers could visit Johnson and Pennypacker (1993).

Features of SCD research

The first feature of SCD research is that the unit of analysis is the individual. Although SCD researchers may use more than one individual in their experimental study, the subject's behaviour is measured repeatedly and compared continuously across experimental conditions. SCD researchers avoid making group comparisons but treat each subject as an intact experiment. Using analogous terms from the group experimental research, the same individual serves both as the comparison as well as the treatment group in SCD research.

The next SCD characteristic is that the behaviour of interest (i.e. dependent variable) is defined operationally so that other researchers could study similar behaviours. An operational definition consists of clarity, comprehensiveness, and objectivity. Clarity refers to a clear topographical description of behaviour that would allow a second observer to measure it; comprehensiveness focus on the inclusion and exclusion criteria of behavioural responses; objectivity refers to a measurable index of target behaviour so that both a second observer and a SCD researcher would record independently and consistently. When referring to behaviours of interest, these may include socially significant behaviours that the participant's social milieu has identified as important for change. For instance, in educational settings, significant behaviours may include academic (e.g. reading fluency, math additions, vocabulary definitions) as well as social ones (e.g. on-task behaviour, hand raising, turn taking, listening).

A third distinguishing feature of SCD research is the presentation of two or more experimental conditions with the same individual. Experimental conditions include the sequential introduction of baseline and intervention conditions. The first condition refers to the current state of environmental variables (e.g. physical setting, traditional instruction, materials) under which the behaviour of interest is present and is measured. Baseline condition serves as control phase, where response patterns are compared to those responses under intervention condition. The intervention condition entails the presentation and manipulation of a new variable (e.g. educational programme) in the current environment.

The schematic arrangement of the above experimental conditions is known as the experimental design, the fourth main SCD feature. Based on the selected design, data are graphically displayed showing the number of baseline and intervention sessions an individual participated along with his behavioural responsiveness in each experimental condition. After SCD researchers have selected a specific experimental design, they arrange the order in which experimental conditions will be introduced so that meaningful and valid comparisons can be made across conditions (Johnson and Pennypacker 1993). SCD research includes a variety of experimental designs. Some of the most common ones are shown in Figure 1. Essentially, the basic premise of each experimental design is first establishing a baseline (or control) condition, where target behaviour is measured repeatedly. For example, if selected behaviour for observation is talk-outs in classroom, then measuring baseline performance could either reveal a stable level of talk-outs or an upward trend of talk-outs over time in classroom. Then, the independent variable (e.g. social skills instruction) is applied and SCD researchers continue to measure target behaviour (e.g. talk-outs) until another stable level or steady positive behavioural trend is established. Multiple replications of those behavioural effects within subject or across subjects in the same design allow researchers to reliably predict future responding, verify and replicate past behaviour in baseline and intervention conditions separately (Cooper, Heron, and Heward 2007).

As Figure 1 shows, SCD researchers arrange their experimentation in a manner in which independent variable can be introduced and then withdrawn (see Type A – reversal design); it can be applied across subjects, settings, or behaviours in a staggered chronological order (see Type B – multiple-baseline design across subjects/settings/behaviours); two different independent variables can be implemented in an alternating manner across conditions based on predetermined schedule (see Type C – alternating treatments design); and it can be implemented in stepwise increases or decreases allowing for replication of the effects of the independent variable at each criterion change (see Type D – changing criterion design). Since the scope here is not to analyse in detail each experimental design, readers may visit Cooper, Heron, and Heward (2007) for a detailed description of these designs as well as variations of those.

Nonetheless, a couple of important considerations should be noted here about the aforementioned designs. First, certain designs (e.g. reversal and multiple baseline) are vulnerable to possible multiple-treatment interference, where interactive effects other than intervention alone might affect participant's behaviour. Such interactive effects include carryover effects (i.e. effects of a previous experimental condition on student's behaviour under another condition) as well as sequence effects (ordering of the experimental conditions). These interactive effects are threats to the internal validity of the study as they affect the relation between the behaviour of interest and intervention (Gast 2009). Sequencing interference can be minimized by introducing experimental conditions to participants in a counterbalanced manner. Carryover effects are less easy to control. They should be identified and documented when evaluating treatment outcomes. Second, ethical and

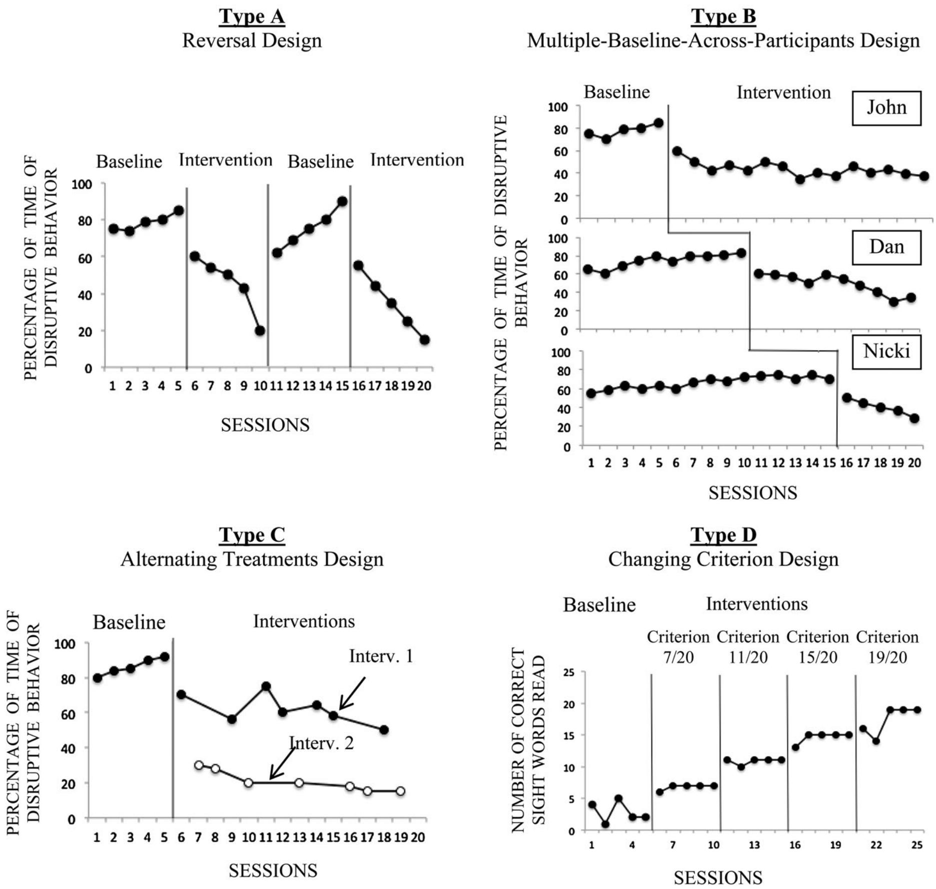

Figure 1. Depicting hypothetical data using four types of SCR designs.

educational concerns might be raised in reversal designs, in which a treatment is withdrawn only to be re-introduced at a later time of the experiment. As Cooper et al. (2007) note, even if the reversal design offers the best scientific approach and does not impose any ethical concerns for the participants, the full support of the people involved must be obtained prior to intervention withdrawal. If support is not provided by the persons in charge of withdrawing intervention, then the procedural integrity will suffer. Additionally, in the case where intervention is conducted for dangerous behaviours, it is not ethical to select the reversal tactic as it will pose serious health concerns to participants.

Finally, SCD researchers follow a unique approach in analysing their data, which is visual inspection. This is one of the oldest but most powerful dynamic data analysis techniques in experimental research (Parsonson and Baer 1992). By collecting repeated measurements of target behaviour, researchers plot data on a graph and inspect for within- and between-phase response patterns. During *within-phase analysis*, SCD researchers look for changes at: (a) data level (i.e. changes in central tendency measures); (b) data trend (slope and magnitude); and (c) data variability (i.e. degree of data dispersion from the best-fit line). When data variability is high (i.e. data are dispersed on the graph), then more data points are required to document a clear pattern. *Between-phase analysis* allows researchers to look for patterns across experimental conditions. Analysis is based on two dimensions: (a) changes in the level, trend and variability of data, and (b) percentage of overlapping data from baseline to intervention (Kennedy 2005). In addition to using visual inspection as a data analytic strategy, other researchers suggested parametric and nonparametric statistical analyses in SCD research (e.g. Edgington 1992; Gliner, Morgan, and Harmon 2000; Gorman and Allison 1997; Jones, Vaught, and Weinrott 1977; Kratochwill and Levin 2014).

Demonstrating experimental control and data believability

At the heart of any experimental study lies the experimental control. In SCD research, experimental control is demonstrated by manipulating with fidelity the independent variable *and* controlling all other extraneous variables, whose interference might contaminate the effects under study. Experimental control means that a functional (causal) relation exists between dependent and independent variable, where the former variable is dependent upon the presence or absence of the latter. In essence, experimental control is evident when researchers are able to say that the independent variable is the *only* possible explanation that could justify reliably the observed changes in the dependent variable after the intervention had been implemented. When such conclusion is drawn, then researchers should be confident that their study presents high degree of internal validity.

Demonstrating functional relations is a complex task that requires good evidence of experimental control of independent and extraneous variables. In SCD methodology, experimental control is enhanced by using a key technique called *steady state strategy* (Johnson and Pennypacker 1993). This strategy allows researchers to measure same behaviour repeatedly across different experimental conditions. Steady state strategy is critical for demonstrating functional relation because it allows researchers to: (a) control or eliminate confounding variables and (b) obtain a stable pattern of responding that provides a complete and reliable picture of intervention effects. Stable data denote steady trend and low data variability in each phase. It is important to obtain stable data prior to moving to the next phase as this allows researchers to determine more clearly the presence of functional relationship between independent variable and target behaviour.

In SCD research, stable data are described experimentally following the baseline logic (Cooper, Heron, and Heward 2007). Baseline logic requires three elements to demonstrate functional relation between intervention and behaviour of interest: prediction, verification, and replication. Regardless of which research design is applied, researchers must ensure that all three elements are examined in order to provide evidence of experimental control. Figure 2 provides a schematic representation of baseline logic using one of the most powerful SCDs, the reversal design. The graph shows hypothetical data of disruptive behaviour patterns of an individual student across baseline (e.g. classroom) and intervention conditions (e.g. self-monitoring strategy).

After a steady baseline responding is established under Baseline 1, SCD researchers decide to implement intervention. Concurrently, researchers *predict* that if baseline were to continue in a non-changing environment, then predicted data would have fallen within the range of current baseline data. This prediction is represented with open data circles in the shaded box under Intervention 1. As researchers implement intervention, they continue to measure behaviour repeatedly (see Intervention 1). A change in the behaviour is recorded and SCD researchers may speculate that independent variable may have influenced target behaviour. However, such assertion about a possible causal link between intervention and behaviour is invalid because it is not clear yet whether the observed change in behaviour was produced by the implementation of independent variable (e.g. self-monitoring strategy) alone or in combination with other extraneous uncontrolled variables (e.g. seat change) that could have coincided with the onset of intervention.

Therefore, SCD researchers determine the extent to which behaviour change was a function of independent variable and nothing else by proceeding to the next experimental condition (Baseline 2). After stable intervention responding has been demonstrated previously, intervention is withdrawn and baseline is re-introduced. If target behaviour changes in the absence of intervention, then Cooper et al. (2007) state that such evidence denotes two important affirmations: First, response change strengthens the presence of intervention as a controlling factor for behaviour change. SCD researchers predict that if intervention were to continue then behavioural responding would have

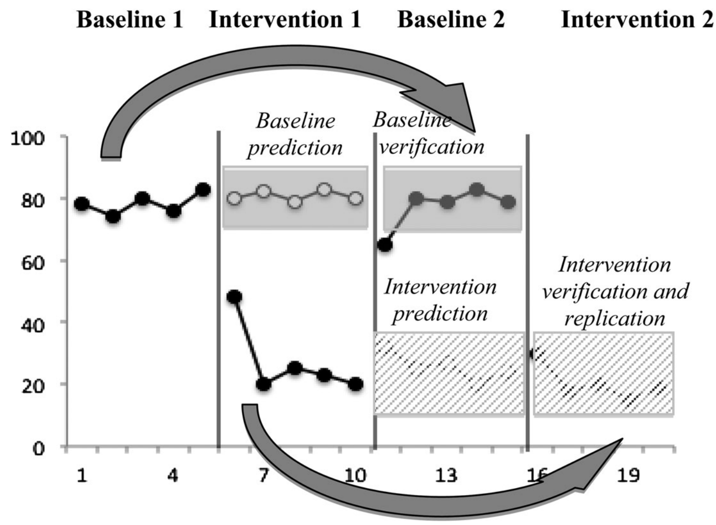

Figure 2. Demonstration of experimental control.

maintained similar patterns as in Intervention 1. This response prediction is depicted with open circles in the shaded box with upward diagonal lines in Baseline 2. Second, if new behaviour patterns vary in a similar way as the ones under Baseline 1, then the conclusion of a functional relation becomes more convincing. Of course, such finding *verifies* SCD researchers' initial baseline prediction. The verification is demonstrated in the shaded box under Baseline 2.

Although baseline control over target behaviour is verified, the effects of the independent variable still do not provide strong evidence for a causal relationship. For SCD researchers, the speculation of the covariance of independent variable and uncontrolled variables is still possible. As in all experimental research designs, replication of intervention effects is a strong indication of data reliability as well as evidence of minimizing or eliminating the effects of uncontrollable variables. SCD researchers are tasked to replicate the intervention effects by repeating intervention manipulations. Within the reversal design, after steady baseline responding is documented in Baseline 2, SCD researchers proceed to Intervention 2. Upon the re-introduction of intervention, experimental effects are replicated again as shown in the shaded box with upward diagonal lines. Replicating intervention effects again and again demonstrates data believability and reliability of functional relationship between intervention and behaviour. If additional manipulations are conducted and consistent replications of response patterns would occur accordingly, then research outcomes become even more convincing. However, how many more experimental manipulations are necessary in the study is also a question that concerns the external validity and believability of SCD research.

Data believability refers to the 'extent to which single-subject researchers can convince others that the data represent what they are intended to represent and thus warrant interpretation' (Johnson and Pennypacker 1993, 364). In other words, data believability is strengthened by replicating behavioural changes. Therefore, by demonstrating effectively intra- and inter-subject replication of their findings, SCD researchers are more confident to argue to consumers of research about the effectiveness and impact of their intervention. Failure to replicate findings within the same experiment is an issue of internal validity threat that could be due to interference of external variables, or treatment integrity issues, or both. Likewise, if researchers fail to replicate their findings with other subjects across studies, then this relates to external validity concerns.

In SCD research external validity is about generalizing a valid and reliable functional relation with other subjects in different settings and over time (Cooper, Heron, and Heward 2007). Following this inductive reasoning, SCD researchers are able to determine over time how functional relations work and what impact intervention programmes present (Kennedy 2005). Generality is demonstrated by direct as well as systematic replication (Sidman 1960).

Direct replication refers to repeating same experimental procedures of an earlier experiment either within the same subject or across subjects with similar characteristics (e.g. same age, gender, ethnicity, similar skill repertoire) to the ones of the previous experiment. Conversely, systematic replication strengthens the generality of SCD outcomes by varying aspects of experimental components (subjects, setting, target behaviour, intervention, etc.) either in the same experiment (intra-experiment replication) or in more than one experiments (inter-experiment replication). In the latter case, by varying experimental procedures from one study to another, more empirical evidence is collected regarding the changes in functional relations. For instance, in one study researchers may examine the effects of teacher praise and feedback on students' on-task behaviour. In a subsequent study researchers may extend the previous findings to another population and systematically investigate the effects of praise and feedback separately.

There are no absolute rules with respect to the timing of replication, the type of replication, or the degree of variation in experimental procedures to follow. Therefore, SCD researchers are asked to make informed judgements (Kennedy 2005). Their decisions, though, are opened to public scrutiny. In essence, successful systematic replications establish a body of scientific knowledge of valid and reliable interventions and set the stage for further empirical investigation of functional relations.

Establishing EBPs using SCD research

Many authorities have discussed the importance of using research designs that can infer causality when determining EBPs and considering only experimental group comparison designs (such as RCTs and quasi-experiments) and SCDs (CEC 2014; Cook et al. 2008; Gersten et al. 2005; Horner et al. 2005). These research designs are unique because they allow researchers to determine whether the independent variable (a programme or a practice) causes changes in the targeted dependent variables (e.g. student outcomes). In addition to the selection of research designs, Cook and Cook (2011) also discussed three additional crucial factors to consider when determining EBPs: quality of research, quantity of research, and magnitude of effect. Quality of research refers to a set of standards or criteria (or called *quality indicators*) that a research study must meet in order to show its methodological soundness and rigour. Quantity of research is determined based on multiple sources of high quality studies with positive effects. Magnitude of effect refers to the degree to which a practice must result in robustly positive and socially valid student outcomes for it to be designated as evidence based.

Several leading professional organizations in social sciences and education across U.S. and Europe have put forth efforts in establishing review criteria or evidence standards for identifying EBPs. In the following sections, we briefly describe these efforts with a focus on SCD research.

Council for Exceptional Children

The Council for Exceptional Children (CEC) is the largest international professional organization in special education located in the U.S.; it is dedicated to improving the educational outcomes of individuals with exceptionalities (with disabilities or of talents and giftedness). In 2003, the Division for Research of CEC established a task force to develop the quality indicators for conducting experimental group research, correlational research, SCD, and qualitative design research, in response to the limited view of many research communities considering RCTs being the only acceptable design practice (Odom et al. 2005). Subsequently, the *Exceptional Children*, CEC's flagship journal and one of the most respected scholarly journals in special education, published a special issue (2005 [issue 71 (2)]) to communicate the roles of different methodologies in informing practices in special education. In this special issue, Horner et al. (2005) established 21 quality indicators within SCDs that attend to seven categories (i.e. description of participants and settings, dependent variable, independent variable, baseline, experimental control/internal validity, external validity, and social validity). Horner et al. proposed that within SCDs, a practice can be labelled as evidence based if there is: (a) an operationally defined practice; (b) a clearly identified context in which the reviewed practice was conducted; (c) an adequate level of implementation fidelity; (d) documented effect of the practice that is functionally related to the changes in dependent variable; and (e) a series of replications of the treatment effects across at least five high quality SCD studies published in peer-reviewed journals, across at least three different researchers and three different geographical locations, and with at least 20 participants (i.e. 5-3-20 rule). Although some may deem the 5-3-20 rule to be somewhat arbitrary, Horner and Kratochwill (2012) state that this standard is proposed to centre on replication of causal effects with replication across research contexts and participants that reflect realistic generalization within SCD research and may serve as a beginning point to bring the field forward with determining a professional standard for EBPs using SCD studies. Since its publication, Horner et al.'s proposed quality indicators have been adopted and applied in comprehensive literature reviews, research syntheses, or meta-analyses to determine if a practice or programme is considered evidence based. For example, Browder et al. (2009) reviewed 30 SCD research studies that used time delay to teach literacy skills to students with significant development disabilities and found 22 met the Horner et al.'s quality indicators with sufficient groups of researchers and participants; consequently, the authors identified time delay as an EBP for teaching picture and sight word recognition to individuals with significant developmental disabilities. In another study, Chard et al. (2009) reviewed studies that involved using repeated reading intervention with students with or at risk for learning disabilities. Researchers found six SCD studies but none of these studies met the minimum requirements for high quality research in all seven categories as specified by Horner et al.; as a result, repeated reading intervention is not yet an EBP for students with learning disabilities based on the reviewed SCD studies.

Most recently, CEC (2014) released a new set of standards that define qualify indicators and criteria for EBPs in special education using both experimental group comparison designs and SCDs by building on the work of Gersten et al. (2005) and Horner et al. (2005) in the *Exceptional Children* special issue mentioned above. The workgroup, consisting of seven special education researchers and leaders, identified a set of quality indicators that are essential for methodological soundness for both group designs and SCDs. The quality indicators are reflected in eight areas, including (a) context and setting, (b) participants, (c) intervention agent, (d) description of practice, (e) implementation fidelity, (f) internal validity, (g) outcome measures/dependent variables, and (h) data analysis. In order to classify the evidence base of practices, a practice or a programme must first be determined to be methodologically sound, followed by being categorized as having positive (i.e. at least 75% of cases resulted in meaningful changes with at least three cases), neutral or mixed, or negative effects (i.e. at least 75% of cases resulted in nontherapeutic changes). When quantifying the evidence base, a practice or a programme under review will receive a designation in one of the five categories: EBPs, potentially EBPs, mixed effects, insufficient evidence, or negative effects. Based on the criteria delineated by CEC, SCD studies of methodological rigour play essential roles in building EBPs in that SCD studies alone or in combination with group design studies can result in EBP classification.

What Works Clearinghouse

What Works Clearinghouse (WWC) is an initiative of the Institute for Education Sciences (IES) housed within U.S. Department of Education; the mission of IES is to provide rigorous and relevant evidence to guide education practice and policy and, in turns, to improve educational outcomes of all students (IES 2015). Since its inception in 2002, WWC has aimed to provide 'a central and trusted source of scientific evidence for *what works* in education' by establishing criteria for determining evidence, conducting thorough reviews of research studies, and disseminating reports to inform researchers, educators, and policy-makers to improve student outcomes (WWC 2014, 1). Although introduced later, WWC has accepted SCDs as adequate methodology that contributes to the identification of scientific evidence and provided standards for evaluation (Kratochwill et al. 2010; WWC 2014). Specifically, the SCD expert panel led by Kratochwill et al. (2010) delineated a three-step process to determine if a practice is considered evidence based. The first step involves evaluating the adequacy and rigour of the experimental design, including (a) level of active and systematic manipulation of the independent variable(s), (b) level of systematic and repeated measures of outcome variables with adequate inter-rater agreement, (c) the number of phase repetitions and effect replications, and (d) the number of data points within each phase. Results of this evaluation is to determine if the design *meets evidence standards*, *meets evidence standards with reservations*, or *does not meet evidence standards*. The second step involves conducting visual analyses of graphic results of the SCD studies that meet evidence standards with or without reservations to determine if the findings are of *strong evidence*, *moderate evidence*, or *no evidence* of a causal relation between the independent variable and the dependent variable. The third step is to assess the overall evidence by applying the 5-3-20 rule to determine if there are at least five SCD studies that meet evidence standards with or without reservations, that involve at least three different research teams at three different geographical locations, and the combined experiments include at least 20 participants (Horner et al. 2005). Additionally, Kratochwill et al. proposed the potential of estimating effect sizes (e.g. regression-based estimates, standardized mean difference, or nonparametric methods) within SCD studies that go beyond visual analysis, but cautioned against using sole method. The concept of estimating effect sizes has received increased attention, but it remains complex and debatable due to the difficulty in finding the most suitable effect size metrics for SCD data (Horner and Kratochwill 2012). Discussion on different statistical analyses for SCD data is beyond the scope of this paper, but readers may refer to other work for this topic (e.g. Jenson et al. 2007; Shadish and Rindskopf 2007).

Recently, researchers have used the WWC guidelines for SCD studies to evaluate if a practice is evidence based. For example, Carr, Moore, and Anderson (2014) reviewed 23 SCD studies that used self-management intervention with students with autism spectrum disorders and used the WWC guidelines to determine if each study met evidence standards, met evidence standards with reservation, or did not meet evidence standards. The researchers also used the percentage of non-overlapping data (PND) metric to calculate effect sizes. Results of the synthesis show that 12 studies, conducted by eight research teams with 34 participants, met WWC standards and 11 studies, conducted by 11 research teams with 36 participants, met standards with reservations. Findings support self-management intervention being an EBP for individuals with autism spectrum disorders. Fallon et al. (2015) also used the WWC guidelines to review 47 SCD studies that examined the effects of performance feedback to increase teachers' implementation fidelity. The quantitative analysis included visual analysis and three quantitative metrics (i.e. improvement rate difference, PND, and standardized mean difference). Fallon et al. found that 29 SCD studies including 102 cases met standards or met standards with reservations, with 54 cases showing strong evidence and 48 cases showing moderate cases. The studies were conducted across 21 research terms and a range of geographical locations, supporting performance feedback to be an EBP based on the WWC criteria.

European Platform for Investing in Children

The EPIC serves as the only central platform for EU Member States to acquire information about EBPs for children and their families. EPIC was established to meet EC's latest child policy (see EC 2013) and it is operated by the EC's Directorate-General for Employment, Social Affairs and Inclusion. EPIC's evaluation framework is based on an American framework, developed by the RAND institution for the online EBPs platform *Promising Practices Network* (PPN) (van Stolk and Kilburn n.d.). A similar framework to the PPN's platform was adopted in EPIC to determine EBPs in Europe and it is defined by three criteria: *evidence of effectiveness*, *transferability*, and *enduring impact* (EPIC 2013). The evidence of effectiveness is based on the quality of research evidence. Research quality is characterized by these features: sample size ($n \geq 20$), minimum alpha level ($p < .1$), comparison group, participant attrition (<25%), and effect size (≥10% of standard deviation). The transferability refers to the degree to which a practice has been evaluated in at least two EU Member States and the practice has sufficient specificity to allow for future replications. Finally, the enduring impact exists when a practice is evaluated with a follow-up study of at least 2 years. For all of these categories, the expert panel may assign a score ranging from 0 (does not meet standards), to + (passing level), to ++ (highest level). Criteria for designating a practice to be a 'best practice' require at least a + in each of the three evaluation categories; 'promising practice' requires at least a + in evidence of effectiveness and a + in at least one of the other two categories. A practice is considered emergent if it receives at least a + in evidence of effectiveness.

After a considerable review of the EPIC's practice user registry and the evaluation criteria for EBPs listed at the EPIC, there is no reference to studies utilizing the SCD methodology. Programme reviews listed on the EPIC (2013) platform focus on group research designs. Additionally, the EPIC's policy thematic agenda is broader than those of CEC and WWW. The latter two organizations concentrate exclusive efforts on improving school outcomes while EPIC includes broader social topics such as parent support with childcare, assisting students at risk, making workplaces family friendly, and developing successful transition outcomes for secondary school students.

Strengths and limitations of SCD research

According to Kratochwill et al. (2010), SCDs allow researchers to address major threats to internal validity in experiment through (a) the structure of the design (called 'methodological soundness') and (b) systematic replication of the intervention effect within the experiment (called 'evidence credibility, 4). Specifically, phase repetition, effect replications (within or across participants), and repeated measurement inherited within SCDs can control for threats to internal validity, including history, ambiguous temporal precedence (through active manipulation of the independent variable), maturation, statistical regression, and testing (Kratochwill et al. 2010). As discussed previously, when studies are conducted with high quality, SCDs provide a rigorous experimental methodology that can indicate causal effects between the dependent variable and the independent variable (Cook and Cook 2011; Horner et al. 2005; Kazdin 2011). SCDs are particularly beneficial and suitable in education for a number of reasons. First, SCDs are appropriate for systematically testing an intervention and its associated variables using a small number of participants before it is tested in a larger scale in a group comparison design format. By nature, SCDs involve a small number of participants whose behavioural performance is investigated frequently and continuously (Matson et al. 2012). Second, the focus on treating each unit (e.g. one participant, one group of individuals, a school) as its own control within SCD studies allows researchers to investigate the dynamic relationship between the unit's responses and the intervention, therefore variability within individual unit is not masked (Cooper, Heron, and Heward 2007; Kennedy 2005). SCDs also offer a great degree of micro vs. macro analysis of an intervention (Matson et al. 2012). Third, in practice, researchers often have difficulties finding a large group of participants, particularly those with special characteristics (e.g. with certain types of disabilities or at-risk status), with similar profiles for a RCT or a quasi-experimental design. As a result, SCDs offer a suitable alternative.

There are also limitations in SCD research. First, in SCD studies, a convenience sample is commonly used and, therefore, inferences to the population on statistical ground is not possible (Schlosser 2009). In other words, findings from a SCD study are specific to the given participants within the specific experiment, and the results cannot be generalized to settings, behaviours, or participants beyond those defined in the same experiment. Evidence of a practice using SCDs is enhanced through multiple replications of effects across different studies, research teams, and participants or unit of analysis (Horner et al. 2005). The previous discussion about establishing EBPs using SCDs is based on this notion of multiple replications and accumulation of research across participants, research teams, and geographical locations. Schlosser (2009) called this *logical generality*, in that researchers may anticipate similar results for participants with similar characteristics beyond current studies when the practice has been replicated a sufficient number of times.

Second, unlike RCTs research, randomization is not a part of SCDs to demonstrate the highest level of methodological soundness as viewed by many research communities. To address this issue, Kratochwill and Levin (2010) suggest *randomized trials stage* as a way to enhance the scientific credibility of traditional SCDs by 'including a randomized trial component in the evaluation of experimental interventions or programs' (130). Specifically, the application of randomization in SCDs may involve randomizing the within-series sequencing of baseline and intervention phases (as in reversal design and alternating treatments design) or randomizing the specific time points each baseline and intervention phrase will start (Kratochwill and Levin 2010). Readers are encouraged to refer to Kratochwill and Levin's work on the application of randomization in SCDs and cautions in such an application to ensure that the design will remain its rigour and methodological soundness that will result in a causal relationship and sound internal validity.

Implications for practice

The purpose of this paper was to give an overview of the SCD experimental research and present SCD quality indicators for evaluating educational programmes as EBPs. Based on our own literature search and review, we have not found any professional organizations in Europe that would consider SCD methodology as part of their evaluation process for identifying EBPs. In America, SCD research has gained significant attention in social sciences over the last decade with a number of major organizations (e.g. American Psychology Association, CEC, US IES) recommending SCD alongside RCTs for identifying EBPs (Kratochwill and Levin 2014).

We contend that SCD is an important methodology educational researchers and policy stakeholders should consider in the EBPs debate in U.S. and Europe for a couple of reasons. First, when working in applied settings such as schools, establishing functional relations is the most important step educational professionals should take into account for understanding the effectiveness of any type of academic and/or behavioural intervention being implemented (Sheridan 2014). As Shavelson et al. (2003) contend, any type of design studies should be interventionist, collaborative, theory driven, utility orientated as well as process focused. This is true for SCD research as it is based on social enterprise and its interest is improving student outcomes using theory-driven procedures and incorporating practitioners in the implementation process. SCD research offers such rigorous framework, in which small-scale experimental studies could be designed and implemented to determine what is and is not working and for whom. The large-scale randomized group experimental studies that have traditionally been considered the 'gold standard' are not cost-effective and time-efficient when trying to determine intervention types and components that would produce robust positive student outcomes. Conversely SCD research, since its inception, has been addressing this educational gap by providing a variety of design options, requiring systematic repeated measurements of student performance, utilizing graphic data displays, and conducting intensive visual inspection to determine the effects of an independent variable (e.g. educational programme) on socially significant behaviours. As a result, SCD research outcomes have demonstrated that instructional strategies such as time delay (Browder et al. 2009), self-management (Carr, Moore, and Anderson 2014), performance feedback (Fallon et al. 2015) have now been identified as EBPs.

Second, SCD research is suitable for curriculum development projects because it lends itself to the guidelines of formative evaluation process (Layng, Stikeleather, and Twyman 2004). In SCD, individuals are exposed to specific instructional components while their performance is measured repeatedly and analysed intensively over a period of time. When participant non-responsiveness is evident under certain instructional condition then researchers change programme components by exerting experimental control. Procedural changes in SCD approach permits researchers to identify what programme elements may produce the highest impact on the learner outcomes. Working with children with diverse needs is important to consider those educational practices that are rigorously and thoroughly evaluated by research methodologies where individual variance is studied carefully rather than being masked into a group format (Sheridan 2014).

Conclusion

Policy-makers and professionals call for improving student outcomes, especially those with identified risk markers. In this EBPs debate as to what works, we contend that both RCTs and SCDs offer highly quality outcomes. However, they differ vastly in the questions they ask: the first centres upon the behaviour of groups while the second targets the behaviour of individuals. In efforts to find effective strategies for at-risk population, it is important to focus on the behaviour of individuals. Thus, SCD methodology is a valuable methodological approach in the EBPs movement.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the European Commission under a Marie Skłodowska-Curie International Outgoing Fellowship [grant number PIOF-GA-2012-331065].

References

Barlow, D. H., M. K. Nock, and M. Hersen. 2008. *Single Case Experimental Designs: Strategies for Studying Behavior Change*. 3rd ed. Boston, MA: Allyn & Bacon.

Browder, D., L. Ahlgrim-Delzell, F. Spooner, P. J. Mims, and J. N. Baker. 2009. "Reviewing the Evidence Base for Using Time Delay to Teach Picture and Word Recognition to Students with Severe Developmental Disabilities." *Exceptional Children* 75: 343–364.

Byiers, B. J., J. Reichle, and F. J. Symonsa. 2012. "Single-Subject Experimental Design for Evidence-Based Practice." *American Journal of Speech-Language Pathology* 21: 397–414.

Campbell, D. T., and J. C. Stanley. 1963. *Experimental and Quasi-Experimental Designs for Research*. Chicago, IL: Rand McNally.

Carr, M. E., D. W. Moore, and A. Anderson. 2014. "Self-Management Interventions on Students with Autism: A Meta-Analysis of Single-Subject Research." *Exceptional Children* 81: 28–44. doi:10.1177/0014402914532235.

CEC (Council for Exceptional Children). 2014. *Council for Exceptional Children Standards for Evidence-Based Practices in Special Education*. Arlington, VA: Council for Exceptional Children. http://www.cec.sped.org/~/media/Files/Standards/Evidence%20based%20Practices%20and%20Practice/EBP%20FINAL.pdf.

Chard, D. J., L. R. Ketterlin-Geller, S. K. Baker, C. Doabler, and C. Apichatabutra. 2009. "Repeated Reading Interventions for Students with Learning Disabilities: Status of the Evidence." *Exceptional Children* 75: 263–281.

Cook, B. G., and S. C. Cook. 2011. "Unraveling Evidence-Based Practices in Special Education." *The Journal of Special Education* 47: 71–82. doi:10.1177/0022466911420877.

Cook, B. G., M. Tankersley, L. Cook, and T. J. Landrum. 2008. "Evidence-Based Practices in Special Education: Some Practical Considerations." *Intervention in School and Clinic* 44: 69–75.

Cooper, J. O., T. E. Heron, and W. L. Heward. 2007. *Applied Behavior Analysis*. 2nd ed. Upper Saddle River, NJ: Person.

Cutspec, P. A. 2004. "Bridging the Research-to-Practice Gap: Evidence-Based Education." *Centerscope* 2 (2): 1–8.

Davies, P. 1999. "What Is Evidence-Based Education?" *British Journal of Educational Studies* 47 (2): 108–121.

Dunst, C. J., C. M. Trivette, and P. A. Cutspec. 2002. "Toward an Operational Definition of Evidence-Based Practices." *Centerscope* 1 (1): 1–10.

Edgington, E. S. 1992. "Nonparametric Tests for Single-Case Experiments." In *Single-Case Research Design and Analysis: New Direction for Psychology and Education*, edited by T. R. Kratochwill and J. R. Levin, 133–158. Hillsdale, NJ: Lawrence Erlbaum Associates.

European Commission. 2013. *Commission Recommendation of 20.2.2013. Investing in Children: Breaking the Cycle of Disadvantage*. Brussels: European Commission.

European Platform for Investing in Children. 2013. http://europa.eu/epic/practices-that-work/index_en.htm.

Fallon, L. M., M. A. Collier-Meek, D. M. Maggin, L. M. H. Sanetti, and A. H. Johnson. 2015. "Is Performance Feedback for Educators an Evidence-Based Practice? A Systematic Review and Evaluation Based on Single-Case Research." *Exceptional Children* 81: 227–246. doi:10.1177/0014402914551738.

Fletcher, J. M., and D. J. Francis. 2004. "Scientifically Based Educational Research: Questions, Designs, and Methods." In *The Voice of Evidence in Reading Research*, edited by P. McCardle and V. Chhabra, 59–80. Baltimore, MD: Paul H. Brookes.

Gast, D. L. 2009. *Single Subject Research Methodology in Behavioral Sciences*. New York: Routledge Taylor & Francis Group.

Gersten, R., L. S. Fuchs, D. Compton, M. Coyne, C. Greenwood, and M. S. Innocenti. 2005. "Quality Indicators for Group Experimental and Quasi-Experimental Research in Special Education." *Exceptional Children* 71: 149–164.

Gliner, J. A., G. A. Morgan, and R. J. Harmon. 2000. "Single-Subject Designs." *Journal of the American Academy of Child and Adolescent Psychiatry* 39 (10): 1327–1329.

Gorman, B. S., and D. A. Allison. 1997. "Statistical Alternatives for Single-Case Designs." In *Design and Analysis of Single-Case Research*, edited by R. D. Franklin, D. B. Allison, and B. S. Gorman, 159–214. Mahwah, NJ: Lawrence Erlbaum Associates.

Horner, R. H. 2006. "Single-Case Research: Documenting Evidence-Based Practice." Paper presented at the 32nd annual convention of the Association for Behavior Analysis, Atlanta, GA.

Horner, R. H., E. G. Carr, J. Halle, G. McGee, S. Odom, and M. Wolery. 2005. "The use of Single Subject Research to Identify Evidence-Based Practice in Special Education." *Exceptional Children* 71: 165–179.

Horner, R. H., and T. R. Kratochwill. 2012. "Synthesizing Single-Case Research to Identify Evidence-Based Practices: Some Brief Reflections." *Journal of Behavioral Education* 21: 266–272. doi:10.1007/s10864-012-9152-2.

Individuals with Disabilities Education Improvement Act. 2004. PL 108-466, 118 Stat. 2647.

IES (Institute of Education Sciences). 2015. *About IES: Connecting Research, Policy and Practice*. Washington, DC: Institute of Education Sciences. http://ies.ed.gov/aboutus/.

Jenson, W. R., E. Clark, J. C. Kircher, and S. D. Kristjansson. 2007. "Statistical Reform: Evidence-Based Practice, Meta-Analyses, and Single Subject Designs." *Psychology in the Schools* 44: 483–493. doi:10.1002/Pits.20240.

Johnson, J. M., and H. S. Pennypacker. 1993. *Strategies and Tactics of Behavioral Research*. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

Jones, R. R., R. S. Vaught, and M. Weinrott. 1977. "Time-Series Analysis in Operant Research." *Journal of Applied Behavior Analysis* 10: 151–166.

Kazdin, A. E. 2011. *Single-Case Research Designs: Methods for Clinical and Applied Settings*. 2nd ed. New York: Oxford University Press.

Kennedy, C. H. 2005. *Single-Case Designs for Educational Research*. Boston, MA: Allyn and Bacon.

Kilburn, R. M., and T. Mattox. 2014. *Using Evidence to Strengthen Policies for Investing in Children*. http://europa.eu/epic/practices-that-work/evidence-guide/evidence-guide_en.htm.

Kratochwill, T. R., J. Hitchcock, R. H. Horner, J. R. Levin, S. L. Odom, D. M. Rindskopf, and W. R. Shadish. 2010. *Single-Case Designs Technical Documentation*. What Works Clearinghouse. http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Kratochwill, T. R., and J. R. Levin. 2010. "Enhancing the Scientific Credibility of Single-Case Intervention Research: Randomization to the Rescue." *Psychological Methods* 15: 124–144. doi:10.1037/a0017736.

Kratochwill, T. R., and J. R. Levin. 2014. *Single-Case Intervention Research: Methodological and Statistical Advances*. Washington, DC: American Psychological Association.

Kratochwill, T. R., and K. C. Stoiber. 2002. "Evidence-Based Interventions in School Psychology: Conceptual Foundations of the Procedural and Coding Manual of Division 16 and the Society for the Study of School Psychology Task Force." *School Psychology Quarterly* 17: 341–389.

Layng, J. T. V., G. Stikeleather, and J. S. Twyman. 2004. "Scientific Formative Evaluation. The Role of Individual Learners in Generating and Predicting Succesful Educational Outcomes." *The Laboratory for Student Success Review* 3 (2): 5–6.

Matson, J. L., N. C. Turygin, J. Beighley, and M. L. Matson. 2012. "Status of Single-Case Research Designs for Evidence-Based Practice." *Research in Autism Spectrum Disorders* 6: 931–938. doi:10.1016/j.rasd.2011.12.008.

McDonnell, J., and R. O'Neill. 2003. "A Perspective on Single/Within Subject Research Methods and 'Scientifically Based Research'." *Research and Practice for Persons with Severe Disabilities* 28 (3): 138–142.

Morris, R. D. 2004. "Clinical Trials as a Model for Intervention Research Studies in Education." In *The Voice of Evidence in Reading Research*, edited by P. McCardle, and V. Chhabra, 127–152. Baltimore, MD: Paul H. Brookes.

No Child Left Behind Act. 2001. 20 U.S.C. § 6301 *et seq*. (2012).

NRC (National Research Council). 2002. "Scientific Research in Education. Committee on Scientific Principles for Education Research." In *Center for Education. Division of Behavioral and Social Sciences and Education*, edited by R. J. Shavelson and L. Towne, 50–79. Washington, DC: National Academy Press.

Odom, S. L., E. Brantlinger, R. Gersten, R. H. Horner, B. Thompson, and K. R. Harris. 2005. "Research in Special Education: Scientific Methods and Evidence-Based Practices." *Exceptional Children* 71: 137–148.

Parsonson, B. S., and D. M. Baer. 1992. "The Visual Analysis of Data, and Current Research into the Stimuli Controlling It." In *Single-case Research Design and Analysis: New Direction for Psychology and Education*, edited by T. R. Kratochwill and J. R. Levin, 15–40. Hillsdale, NJ: Lawrence Erlbaum Associates.

Plavnick, J. B., and S. F. Ferreri. 2013. "Single-Case Experimental Designs in Educational Research: A Methodology for Causal Analyses in Teaching and Learning." *Educational Psychology Review* 25: 549–569.

Reyna, V. F. 2004. "Why Scientific Research? The Importance of Evidence in Changing Educational Practice." In *The Voice of Evidence in Reading Research*, edited by P. McCardle and V. Chhabra, 47–58. Baltimore, MD: Paul H. Brookes.

Schlosser, R. W. 2009. "The Role of Single-Subject Experimental Designs in Evidence-Based Practice Times." *Focus: Technical Brief* 22: 1–8. www.ncddr.org/kt/products/focus/focus22/.

Shadish, W. R., L. V. Hedges, R. H. Horner, and S. L. Odom. 2015. *The Role of between-Case Effect Size in Conducting, Interpreting, and Summarizing Single-Case Research*. Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Shadish, W. R., and D. M. Rindskopf. 2007. “Methods for Evidence-Based Practice: Quantitative Synthesis of Single-Subject Designs.” *New Directions for Evaluation* 113: 95–109.

Shavelson, R. J., D. C. Phillips, L. Towne, and M. J. Feuer. 2003. “On the Science of Education Design Studies.” *Educational Researcher* 32 (1): 25–28.

Sheridan, S. M. 2014. “Single-Case Designs and Large-N Studies: The Best of Both Worlds.” In *Single-Case Intervention Research: Methodological and Statistical Advances*, edited by T. R. Kratochwill and J. R. Levin, 299–308. Washington, DC: American Psychological Association.

Sidman, M. 1960. *Tactics of Scientific Research: Evaluating Experimental Data in Psychology*. New York: Basic Books.

Skinner, B. F. 1938. *The Behavior of Organisms: An Experimental Analysis*. New York: Appleton-Century.

van Stolk, C., and R. M. Kilburn. n.d. *Fact Sheet: European Platform for Investing in Children*. Brussels: RAND Europe. www.randeurope.org.

Tawney, J. W., and D. L. Gast. 1984. *Single Subject Research in Special Education*. Columbus, OH: Charles E.

What Works Clearinghouse. 2014. *Procedures and Standards Handbook* (Version 3.0). Institute of Education Sciences, U.S. Department of Education. http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf.