The Role of Single-Subject Experimental Designs in Evidence-Based Practice Times

Ralf W. Schlosser, PhD, Department of Speech-Language, Pathology and Audiology, Northeastern University

The concept of evidence-based practice (EBP) is omnipresent in medicine and allied health care and is gradually gaining a foothold in rehabilitation and disability as well (NCDDR, 2006). EBP is defined as the integration of best and current research evidence and clinical/educational expertise with relevant stakeholder perspectives to inform decisions relative to an individual client (Schlosser & Raghavendra, 2004). The emergence of EBP has brought about increased scrutiny of the importance of evidence and what constitutes high-quality research evidence. One outgrowth of this emphasis has been the declaration of randomized controlled trials (RCTs), where a sample is drawn from the population and participants are randomly allocated to the treatment group or the control group (Hahs-Vaughn & Nye, 2008), as the gold standard of treatment research (only to be superceded by meta-analyses of more than one RCT) (Sackett et al., 1997).

This superior status is reflected in the prominent place that RCTs occupy on most hierarchies of evidence (e.g., Lloyd-Smith, 1997) as well as in its attributed importance within the progression of a phase model of treatment research (Robey, 2004). It has led to pressure in rehabilitation and disability as well as in related fields such as education to produce more RCTs. Moreover, this attributed prominence of RCTs directly or indirectly calls into question any kind of designs that are non-RCTs for demonstrating whether or not a treatment works. One such group of non-RCT designs is single-subject experimental designs (SSEDs). SSEDs make up a considerable percentage of treatment studies across the fields of education and rehabilitation and disability (e.g., Schlosser & Sigafouos, 2006; Wendt, 2006). Hence, there is a need to discuss the role of SSEDs during these times when EBP seems to dominate the discourse.

The purpose of this brief is to discuss the role of SSEDs in terms of establishing empirically supported treatments and implementing EBP. First, however, SSEDs will be defined and distinguished from other designs.

What Are SSEDs?

SSEDs are often described as \( n = 1 \) designs. This is because experimental control is established within one unit rather than across units (Kennedy, 2005). Most often human beings are the unit. However, the sample size could also refer to a classroom, a school, a system, a community, or even an animal. Frequently there is the misperception that SSEDs involve only one unit such as a single human. While only one participant is needed to implement an SSED, studies with an \( n = 1 \) design are the exception rather than the rule; a larger number of participants has the potential to enhance the generality of the findings. SSEDs utilize repeated observations and measurements prior to and during/after an intervention that are consistent with a time-series design. These repeated observations are typically presented in graphic format, and the data analyses are visual in nature.
Distinguishing SSEDs From Other n = 1 Trials

SSEDs are not the only n = 1 designs. In medicine, researchers have proposed and implemented the use of n = 1 RCTs, known as n-of-1 RCTs. With this design, subjects are assigned to an active treatment or placebo and then crossed over at random during a series of treatment intervals. The subject and clinician remain blinded to treatment assignment during these intervals (Guyatt, Keller, Jaeschke, Rosenbloom, Adachi, & Newhouse, 1990). An n-of-1 RCT is similar to an SSED in that it also relies on repeated measures. At the same time, there are also important differences. For instance, n-of-1 RCTs rely on procedures used for implementing group RCTs such as the assignment of subjects to treatment conditions (Backman & Harris, 1999). Considering the questions and treatments studied using SSEDs, random allocation to a placebo condition, without the subject realizing this, seems next to impossible. Although placebo conditions are easy to camouflage in medicine (e.g., sugar pills instead of actual medicine), this rarely works with behavioral treatments. Similarly, when multiple treatments are being investigated using SSEDs, these are usually studied within the same individual. The crossing over from treatment to placebo or another treatment, as done with n-of-1 RCTs, is typically not implemented with SSEDs. Rather, once a treatment is assigned to a participant, the participant usually remains with this treatment; if a participant is assigned multiple treatments, they are applied concurrently rather than intermittently. This is the case when comparing two or more treatment strategies aimed at acquisition through an adapted alternating treatments design (Schlosser, 1999a). With this design, researchers develop as many equivalent instructional sets (e.g., manual signs) as there are treatments. The sets are assigned randomly to the treatment and then implemented concurrently with the order counterbalanced or randomized (see Schlosser & Blischak, 2004, for an example). In sum, it is clear that SSEDs share some features with n-of-1 RCTs, but there are also crucial differences that set them apart. This brief focuses on the role of SSEDs only; they tend to have much greater applicability in disability and rehabilitation research.

Role of SSEDs in Treatment Research

SSEDs and the Progression Model of Treatment Research

Robey (2004) discussed a five-phase progression model for conducting treatment research in which he commented on the utility of various design strategies, including SSEDs. In Phase I, the primary purpose is to select a therapeutic effect (if one is present) and to estimate its magnitude. Besides case studies, small-group pre/post studies, and retrospective studies, Robey lists discovery-oriented SSEDs as potential research designs in this phase. SSEDs seem well-suited for accomplishing many of the proposed tasks of Phase I, including developing a first approximation of the treatment protocol and the population definition; estimating appropriate dosage; detecting the therapeutic effect; and generating or refining hypotheses. Due to the liberal control of Type I errors in this phase along with low n requirements for implementing SSEDs and their response-guided nature, SSEDs seem highly conducive as a design for this stage.

In Phase II, the aim is to explore the dimensions of the therapeutic effect and to make the necessary preparations for conducting a clinical trial. This includes tasks such as refining the definition of the target population; assessing the therapeutic effect in terms of the range of utility (i.e., whether it might apply to other subjects); refining the treatment protocol; refining the outcome construct; finalizing operational definitions; and making point and interval estimates of effect sizes. Besides considering several other designs (i.e., case studies, small-within-group designs, case-control studies, and small-group cohort-control designs), Robey (2004) sees a role for discovery-oriented SSEDs in Phase II as structured and formal experimental designs for testing specific formulations of the treatment.

The purpose of Phase III is to test the efficacy of a treatment—that is, whether or not the treatment works under ideal conditions. Here, Robey (2004)
calls for the use of between-group designs that include one experimental group and one control group. Although not specifically named or described by Robey in terms of its characteristics, the RCT, which is a specific type of between-group design, is considered the gold standard (Lloyd-Smith, 1997; Sackett, Rosenberg, Gray, Haynes, & Richardson, 1997). As explained earlier, an RCT involves the drawing of a sample from the population and the random allocation of subjects to the experimental group or the control group (or second treatment condition). Citing Chambless and Hollon (1998), Robey (2004) argues that rigorous protocols of SSEs may be used as well to evaluate efficacy.

In Phase IV, researchers assess whether the therapeutic effect demonstrated earlier in efficacy research can be realized in day-to-day clinical practice in order to establish its effectiveness. According to Robey (2004), the aim is to expand the applicability of the treatment protocol beyond the original form in terms of population, service-delivery model, and treatment delivery method. Research designs suitable for this phase include pre/post group designs, between-group designs, and hypothesis-driven SSEs.

The purpose of Phase V, the last phase, is to determine the cost-effectiveness of a treatment. Here, SSEs play no role. SSEs, therefore, can play a role in four out of the five phases of Robey’s five-phase progression model for conducting treatment research. In sum, SSEs play a critical role in bringing a treatment from initial conceptualization to implementation in daily practice.

**SSEs and Empirically Supported Treatments**

Due to the recent advent of EBP and the promotion of RCTs as the gold standard of research in medical fields, many fields with a long-standing history of SSEs have been forced to take another look at SSEs and discuss their merit. Horner, Carr, Halle, McGee, Odom, and Wolery (2005) presented several criteria that would indicate the appropriate use of an SSE for the identification of empirically supported treatments. The criteria included the following:

- Participants and the process of their selection are described with sufficient detail to allow other researchers to select similar participants.
- Critical features of the physical setting are described with sufficient precision to allow for replication.
- The dependent variable is sufficiently operationalized and measured repeatedly using sufficient assessment occasions to allow for identification of performance patterns prior to intervention and comparison of performance patterns across conditions/phases (level, trend, variability).
- The dependent variable is assessed for consistency through interobserver agreement.
- The dependent variable is selected for its social significance.
- The independent variable is defined with replicable precision.
- The independent variable is actively manipulated.
- The fidelity of independent variable implementation is documented.
- The performance during baseline condition is compared with performance during intervention.
- The emphasis on comparison across conditions requires measurement and description of the baseline (comparison condition).
- The description of the baseline condition should be sufficiently precise to permit replication.
- The measurement of baseline data should continue until performance is sufficiently consistent before intervention is introduced to allow prediction of future performance.

---

1 Horner et al. (2005) refer to “identifying evidence-based practices,” but the argument has been made that the phrase “identifying empirically supported treatments” might be more appropriate for their self-stated purpose (see Schlosser & Sigafoos, 2008).
Experimental control is demonstrated via three demonstrations of the experimental effect (predicted change in the dependent variable varies with the manipulation of the independent variable at different points in time within a single participant or across different participants).

An experimental effect is demonstrated when predicted performance of the dependent variable co-varies with the manipulation of the independent variable.

Based on these criteria, Horner et al. (2005) proposed that a treatment meet the following standards in order to be considered empirically supported: (1) a minimum of five SSED studies on the treatment have been published in peer-reviewed journals that meet minimally acceptable methodological criteria and document experimental control; (2) the studies are conducted by at least three different investigators across three different locations; and (3) the studies include a total of at least 20 participants. Consistent with these high standards, several task forces have acknowledged that SSEDs may go a long way toward demonstrating that a treatment is considered empirically supported (e.g., Gambrill, 1999; Lonigan, Elber, & Johnson, 1998).

**SSEDs and Evidence-Based Practice**

In order to discuss the role of SSEDs in EBP, we must begin with a definition of EBP as a construct. Sackett and his colleagues (1996) defined evidence-based medicine (that is, EBP specific to the field of medicine) as “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients [emphasis added] . . . [by] integrating individual clinical expertise with the best available external clinical evidence from systematic research” (Sackett et al., 1996, p. 71). Subsequently, Schlosser and Raghavendra (2004) proposed an EBP definition for the field of augmentative and alternative communication as “the integration of best and current research evidence with clinical/educational expertise and relevant stakeholder perspectives to facilitate decisions for assessment and intervention that are deemed effective and efficient for a given direct stakeholder [emphasis added]” (p. 3). Although there are some differences between these definitions, what is important for our purpose is that both definitions emphasize that EBP should assist with decisions relative to individual clients. Therefore, the first step of the EBP process—the asking of well-built questions—should also focus on an individual rather than on a group of individuals or population (Schlosser, Koul, & Costello, 2007). If we accept that research evidence should help inform decisions relative to an individual2 it becomes prudent to discuss what kind of evidence is more conducive for doing so. We will engage in this discussion by contrasting SSEDs with RCTs.

**Extrapolation as a Shared Starting Point**

In all likelihood, the individual specified in a well-built question is not the same as the subject or subjects in a research study (or studies synthesized in a systematic review). Hence, whether the research evidence before the practitioner involves an SSED or an RCT, the practitioner will have to engage in what might be called “extrapolation.” Extrapolation may be defined as the act of inferring or estimating by extending or projecting known information. Steiner (1999) suggested that one way to fully individualize a treatment effect or make it relevant to the client is to conduct an n-of-1 RCT with each client for whom a certain treatment is being implemented. Similarly, a practitioner could implement an SSED with each client for whom a well-built treatment question has been formulated. This approach would permit the practitioner to circumvent extrapolation altogether. While the benefits of such an approach for both is obvious—avoiding extrapolation and not only talking the talk but walking the walk (regarding data collection)—there are also numerous barriers to its implementation, including resources and skill burden. Thus, in most cases we still are faced with

---

2 It is acknowledged that research evidence should never directly inform decision making; rather, research evidence should be modified by clinical or educational expertise and relevant stakeholder perspectives.
extrapolation as a common starting point. First, we will begin with the underlying facts of having to extrapolate from an RCT and discuss its ramification for the practitioner.

**Extrapolating From RCTs**

In an RCT subjects are drawn from the population, preferably at random, with the goal of selecting a sample that is representative of the population at large. For the practitioner, this has the benefit of inferential generality whereby the results can be generalized from the sample to the larger population using statistical methods. Hegde (2007) pointed out, however, that this requirement of RCTs is often not met; samples tend to be drawn from subpopulations rather than populations at large. Therefore, it is important for the practitioner to be cognizant that the notion of inferential generality cannot be evoked for each and every RCT.

In terms of enrolled subjects, RCTs tend to describe the inclusion criteria and the group means and ranges for applicable characteristics. Typically, individuals are not described. The group mean is a somewhat artificial denotation that may not be fully representative of any one particular subject enrolled in the experiment. For practitioners, this may pose a challenge in comparing the characteristics of the client to those of the subjects enrolled in the experiment—comparisons can be made relative only to the mean or the extreme scores on either end. Because only the extreme scores (ranges) and the means are provided, it is difficult to assess whether the mean applies to the individual.

Analyses tend to be conducted at the group level to assess, for example, the difference between an experimental group and a control group in terms of effect size. Steiner (1999) noted that tensions arise in the process of extrapolation (he calls it “translation”) because RCTs do not take into consideration how individual characteristics account for outcomes. This poses a difficulty for practitioners because they cannot determine whether the group-level results generalize to their clients. Subgroup analyses may alleviate this concern somewhat as they provide data on how efficacious the treatment is based on the characteristics of subgroups enrolled in the RCT. For instance, Yoder and Stone (2006) conducted subgroup analyses in an RCT comparing Responsive Education and Prelinguistic Milieu Teaching (REPMT) with the Picture Exchange Communication System (PECS) in young children on the autism spectrum. This subgroup analysis revealed that children with initially high object-exploration skills yield better outcomes with REPMT whereas children with low object-exploration skills fared better with PECS. This finding is informative to the practitioner who can make a more nuanced decision about selecting either treatment for his or her clients. Although subgroup analyses have the potential to be somewhat helpful, Steiner (1999) pointed out that they tend to be rare and, if available, are often statistically underpowered, which limits their applicability to individual decision making.

Besides issues related to extrapolating from the population to the individual, the practitioner also has to assess whether the treatment itself, the level of treatment integrity or fidelity, the skills of the treatment agent, the setting, and the measurement of outcomes as implemented in the RCT are generalizable to the individual client (for variables to consider in assessing the transportability of evidence, please consult Schlosser, 2003). RCTs are primarily used in Phase III, as described earlier, to establish whether a treatment works under optimal conditions. Optimal conditions often are generated using highly trained treatment agents, excellent treatment fidelity, settings that are very conducive to the experiment, and the like. Hence, the question becomes whether these conditions really could be replicated with the client to achieve the same end. In sum, extrapolating from RCTs to well-built questions involving individual clients poses considerable challenges for practitioners. Much of what has been said can be summarized by the following quote:

> We can rarely translate with certainty the average benefit reported in randomized clinical trials to a precise assessment of treatment benefit for an individual patient. Subgroup analyses from randomized clinical trials can refine
average effects of treatment into subgroup-specific effects, but analysis of a subgroup small enough to include all of the relevant risk factors of an individual patient may lack the precision necessary to be clinically useful. ... The language of populations can bring us closer to informing our patient about the consequences of treatment but cannot convey all that must be said (Steiner, 1999, p. 620).

**Extrapolating From SSEDs**

As mentioned earlier, SSEDs tend to be used more frequently than RCTs in disability and rehabilitation and related fields. With SSEDs, researchers tend to rely on a convenience sample. Therefore, no inferential generality is generated and practitioners cannot easily assume that the results obtained with the convenience sample would generalize to the population on statistical grounds. On the other hand, SSEDs have the potential to produce what some have called logical generality. Logical generality is established when the results of a treatment have been replicated a sufficient number of times, which would lead one to conclude that a participant with characteristics similar to those of the participants who completed the experiments, if enrolled in the treatment, would yield similar results. The practitioner could make this same extrapolation if the client was similar to the participants who completed the experiment. How many replications are sufficient to establish logical generality? Perhaps the standard set by Horner et al. (2005) for calling a treatment empirically supported might be used as well in this case. To what extent the treatment literature lives up to this standard has not yet been empirically evaluated—systematic reviews to that end would be desirable.

With SSEDs, participants tend to be described in terms of both inclusion criteria and individual characteristics. In published SSED studies, it is not uncommon to find separate sections in a manuscript devoted to each participant labeled by a pseudonym in which characteristics pertinent to that particular treatment study are described. This allows practitioners to compare the participants enrolled in the study with their client more precisely. That being said, it cannot be assumed that all SSEDs report all necessary subject selection procedures and criteria. As documented by Bedrosian (2003), many SSEDs fall short on reporting crucial language and sensory variables. It is for this reason that one of the quality criteria for SSEDs specified earlier rightfully indicates the importance of subject selection being defined with replicable precision (Horner et al., 2005). Thus, although SSEDs tend to do better than RCTs in that respect, the discerning practitioner should ascertain this on a case-by-case basis.

Data analyses in SSEDs are typically implemented at the individual subject level. Generalized conclusions across subjects are drawn as applicable and appropriate. This, along with the description of individual subject characteristics, allows for a relatively easy determination by practitioners regarding the applicability of the results to their clients. In addition, individual variations in the outcomes are discussed in terms of individual participant characteristics. Again, this has the potential to be very informative for assessing the goodness-of-fit between the client and successful study participants.

As mentioned earlier, high-quality SSEDs provide a detailed description of the physical setting and an operational definition of the independent variable to facilitate replication, as well as data to support the fidelity of treatment implementation. These methodological considerations permit the practitioner to assess whether the physical setting in which the client finds himself or herself is comparable to that of the study participants, whether the treatment is feasible in a practical setting, and whether the treatment can be implemented with a similar degree of fidelity—factors that all contribute to the transportability of the findings to the client. Although not inherent in all SSEDs, high-quality SSEDs select treatment goals that are of social significance (Horner et al., 2005) and/or evaluate treatments and outcomes that are socially valid (by relevant stakeholders) (Schlosser, 1999b). If the relevant stakeholders of the client bear similarity
to those in the experiment, this may enhance the transportability of the findings (Schlosser, 2003). Social validation assessments are possible with RCTs as well; however, they tend to be not as common with that research tradition.

**Summary and Conclusions**

It has been shown in this brief that SSEDs are deliberate, systematic, and a priori research designs that have the potential to minimize threats to internal validity and contribute to external validity through the process of replication.\(^3\) Quality standards have been proposed that help practitioners distinguish sound from poor SSEDs and delineate clearly what it takes for a treatment to be considered empirically supported (Horner et al., 2005). SSEDs also have been shown to play a critical role in a phase model of treatment research with its active contributions to four out of five phases. In Phase III, RCTs have the distinct advantage of producing population-based evidence in support of a treatment working under ideal conditions. Even though SSEDs have been likened to RCTs in terms of the overall level of evidence they can achieve (under certain conditions), it should be kept in mind that SSEDs cannot produce population-based evidence (at the most, SSEDs can produce logical generality through sufficient replications). RCTs, as deliberated, pose significant challenges to the practitioner in EBP-implementation efforts. For numerous reasons discussed in this brief, it is simply very difficult and problematic to extrapolate population-based evidence to individual clients. SSEDs, on the other hand, lend themselves to easier extrapolation to clients in practice—logical generality may facilitate this process—but they cannot offer the benefit of inferential generality like RCTs. Based on the deliberations in this brief, the following course of action is proposed. Properly conducted RCTs with samples drawn randomly from the larger population continue to be needed to establish the efficacy of treatments and produce population-based evidence. Hypotheses-driven subgroup analyses with sufficient \(n\) are desirable to yield subgroup-specific clinical implications. RCTs may be preceded by rigorous SSEDs in order to create the impetus and prepare for the implementation of an RCT. Alternatively, RCTs may be followed up with high-quality and rigorous SSEDs under ideal conditions in order to individualize population-based evidence of efficacy. Further, in Phase IV, rigorous SSEDs should be conducted under everyday conditions to help establish the effectiveness of treatments. Practitioners will find the results from these designs more amenable for extrapolation to their clients. Finally, resources permitting, it would be desirable to further individualize promising treatment effects by applying treatments to actual participants in practice-based SSEDs.

---

\(^3\)** Although SSEDs share some features with n-of-1 RCTs, it has been shown in this brief that they also bear several critical differences. Therefore, this author argues that these designs are sufficiently dissimilar to warrant separate discussions concerning their contributions to empirically supported treatments and EBP.
References


NCDDR (2006). The role of systematic reviews in evidence-based practice, research, and development. FOCUS, 15, 1–4. Austin, TX: SEDL.


Copyright © 2009 by SEDL

FOCUS: A Technical Brief From the National Center for the Dissemination of Disability Research was produced by the National Center for the Dissemination of Disability Research (NCDDR) under grant H133A060028 from the National Institute on Disability and Rehabilitation Research (NIDRR) in the U.S. Department of Education's Office of Special Education and Rehabilitative Services (OSERS). NCDDR is funded 100% by NIDRR at $750,000 per project year.

FOCUS: A Technical Brief From the National Center for the Dissemination of Disability Research was produced by the National Center for the Dissemination of Disability Research (NCDDR) under grant H133A060028 from the National Institute on Disability and Rehabilitation Research (NIDRR) in the U.S. Department of Education’s Office of Special Education and Rehabilitation Services (OSERS). NCDDR is funded 100% by NIDRR at $750,000 per project year.

SEDL operates the NCDDR. SEDL is an equal employment opportunity/affirmative action employer and is committed to affording equal employment opportunities for all individuals in all employment matters. Neither SEDL nor the NCDDR discriminates on the basis of age, sex, race, color, creed, religion, national origin, sexual orientation, marital or veteran status, or the presence of a disability. The contents of this document do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the federal government.

Available in alternate formats upon request. Available online: www.ncddr.org/kt/products/focus/focus22/