Using Single-Subject Research to Establish the Evidence Base of Special Education

Melody Tankersley
Sanna Harjusola-Webb
Timothy J. Landrum

Research in the field of special education often incorporates single-subject designs to investigate the effectiveness of educational practices for students with disabilities. As such, it is important that educators and educational professionals understand the characteristics of single-subject research methodologies and how those characteristics allow conclusions to be drawn about effectiveness of practices. Because conclusions about whether an intervention causes changes in student outcomes can be derived from single-subject research, it has much to offer to discussion of evidence-based practice and the ultimate identification of evidence-based practices for students with disabilities.

Keywords: single-subject design research

The term evidence-based practice generally refers to educational programs or instructional procedures that have been determined to produce consistent patterns of positive results in student outcomes. Although the concept of evidence based seems straightforward, many questions arise when practitioners try to determine which practices in fact have a positive evidence base. Among the more fundamental questions to arise are those regarding what constitutes evidence and, more specifically, what types of research methods produce the type of evidence that can support the effectiveness of a given practice. Indeed, not all types of research are intended to address questions about effectiveness.

To establish that an educational practice is effective, or causes the desired or intended results, researchers seek to establish that there is a systematic, or functional, relationship between variables. Experimental research is, by and large, regarded as providing the most credible evidence of the effectiveness of a practice (B. G. Cook, Tankersley, Cook, & Landrum, 2008 [this issue]). Experimental research includes two general types of methodological approaches: group experimental designs and single-subject research designs. In group experimental studies, researchers (a) use meaningful comparison groups, such that the outcomes of those receiving a particular intervention are compared to the outcomes of those who do not receive it, and (b) actively and systematically implement the educational practice being tested (L. Cook, Cook, Landrum, & Tankersley, 2008 [this issue]). In this type of research, the effectiveness of a practice is examined by making comparisons between the outcomes of the two groups. The hallmark characteristics of group experimental research—using a comparison group and providing the intervention to only one of the groups—allows researchers to draw
conclusions about whether a practice causes the desired changes in participants’ outcomes.

Like group experimental designs, single-subject research designs also allow researchers to draw conclusions about whether an intervention causes desired changes in student outcomes. Despite this conceptual similarity, single-subject designs differ from group experimental designs in important ways. First, instead of using groups to make comparisons, in single-subject research, participants provide their own comparison. Their performance is compared across conditions in which they are and are not participating in the intervention under study. For example, a researcher might measure a student’s math problem solving when teachers give frequent corrective feedback on their performance, compared to a condition in which no corrective feedback is given.

Most typically, researchers who use single-subject research methods study individual learning or behavioral outcomes under at least two experimental conditions, a baseline condition and an intervention condition. A baseline condition generally involves the status quo, or observations of behavior or performance in a typical setting (e.g., a classroom) before the intervention in question is implemented. These observations occur repeatedly to determine a reliable current level of performance. Once baseline performance is stable and predictable, the educational practice or intervention is introduced, and performance is again measured frequently while it is in place during a period of time. Researchers measure the target behaviors repeatedly across baseline conditions and then across intervention conditions to establish student performance in the absence and in the presence of the intervention. Comparisons of data across baseline and intervention conditions provide the basis for establishing whether there is in fact a functional relationship between the independent and dependent variables. Horner et al. (2005) recommended that single-subject designs include at least three demonstrations of a functional relationship. Obviously, the more demonstrations a design includes, the more evidence researchers have that a true functional relationship exists between a given practice and student performance.

Because changes within a student’s behavior or performance are demonstrated from baseline to intervention conditions, multiple examples of single-subject research results can establish whether a particular practice has an evidence base. That is, causal inferences—that the intervention causes the change in students’ behavior or performance—can be ascertained from the results of single-subject research designs. And because causal inferences are generally recognized as necessary for establishing evidence-based practices, single-subject research can be used to determine evidence-based practices.

Essential features of single-subject research designs, those related to measuring target behaviors, introducing the intervention, and analyzing the effects, allow researchers to make determinations about whether a practice is effective. These features can be illustrated through their applications in special education literature. Single-subject research designs are also closely aligned with special education’s core principles of individualized instructional decisions and frequent monitoring of student progress. Both of these contributions—identifying the evidence base and providing meaningful information about individualized instruction—make single-subject research methods critical considerations for special educators.

**Essential Features of Single-Subject Research**

The purpose of single-subject research is to establish whether a functional relationship exists between a practice and student outcomes at the level of individual participants. Despite its name, single-subject research designs rarely involve only one participant. Even so, the data are organized and analyzed according to a within-participant orientation that focuses on the individual. That is, each participant’s behavior or performance is compared to his or her own behavior or performance across multiple conditions. Although there are a number of different single-subject research designs, three key components are common to all single-subject designs: (a) Target behaviors must be assessed repeatedly using trustworthy measures, (b) interventions must be systematically introduced and withdrawn, and (c) the effects across baseline and intervention conditions must be analyzed for each participant (Kazdin, 1982, 1992).

**Repeated, Trustworthy Measurement of the Target Behaviors**

The repeated measurement of target behaviors is critical to single-subject research. By measuring target behaviors repeatedly, researchers can be more confident that the collective measurement is a true representation of the participant’s performance and the effects of random conditions in the environment are minimized. Because the target behavior is measured repeatedly within both baseline and intervention conditions, it must be measured in a way that allows for multiple assessments, and measurement must occur the same way each time. Thus, assessments that change over time or cannot
be given in quick succession (e.g., reading probes that increase in difficulty, standardized tests that are designed to be given only once per year) are poor measurement choices for single-subject research.

Repeatedly measuring the target behavior over time is also important so that performance patterns and trends both prior to intervention and during intervention can be identified (Kazdin, 1992). For example, single-subject researchers often continue a baseline phase until the target behavior appears to be stable. A stable baseline provides confidence that the observed behavior is occurring in a typical way or at a typical rate for that participant and allows for a relatively unambiguous assessment of the impact of an intervention. For example, to examine the effect of social stories on the screaming behavior of a child with autism, Agosta, Graetz, Mastroperii, and Scruggs (2004) observed and recorded the number of times the child screamed during a total of 36 days. These researchers collected initial baseline data until a stable pattern was established in the first 8 of those days. In an academic behavior example, Boulineau, Fore, Hagan-Burke, and Burke (2004) recorded the percentage of story-grammar elements children represented correctly in story maps across 13 days, including 4 days of an initial, stable baseline phase.

In addition to measuring behavior repeatedly, single-subject researchers must ensure that their measures of behavior are trustworthy. In other words, they must ensure that the measurements used reflect what has really occurred. The most common way this is accomplished is to assess some form of interrater or interobserver agreement (sometimes referred to as interrater or interobserver reliability). To determine interobserver agreement, two independent observers record the same student behaviors (e.g., hand raising) or score the same products (e.g., students’ spelling performance). A calculation is then made as to the extent to which the two observers agreed that a behavior occurred or a student’s performance was correct or incorrect. Agreements are then typically expressed as a percentage (e.g., 90%).

Many single-subject studies use some form of direct observation of the target behaviors. Direct observation of student performance can occur in domains such as academic skill areas (e.g., reading, writing, math), peer interactions (e.g., social skills, cooperation, peer tutoring), and behavioral skills (e.g., self-regulation, on-task behavior, disruptive behavior). Researchers strive to define target behaviors in measurable terms that can be reliably observed and recorded. Clear behavioral or operational definitions of the target behavior help to promote replicability of the study, interobserver reliability, and therefore the overall trustworthiness of the research. An operational definition provides enough detail about a target behavior or response to allow observers to agree a majority of the time on whether the behavior occurred. For example, the phrase is disruptive is not well-defined, whereas makes irrelevant comments during a teacher-led lesson is more operationally clear.

In the example mentioned previously, Agosta et al. (2004) defined and measured “screaming behavior” in their examination of the effects of social stories with a young child with autism as including yelling, crying, or loud humming during group circle time. In the other example, Boulineau et al. (2004) evaluated whether story mapping would increase reading comprehension of 6 third- and fourth-grade students with learning disabilities. To measure the outcome variable, these researchers counted and then calculated a percentage of identified story-grammar elements students correctly represented on their story maps after they read passages.

Repeated and Systematic Introduction of the Intervention

In addition to assessing target behaviors repeatedly using trustworthy measurements, single-subject research designs also require that interventions are systematically introduced and withdrawn to determine whether the intervention (or something else) caused any observed change in the target behavior. By purposefully and systematically making changes in whether and how an intervention is implemented (e.g., not present, provided with more or less intensity) and then evaluating how the target behavior responds to those changes, researchers can make claims as to the extent to which the intervention causes changes in student outcomes. It is this causal relationship between the use of a practice and changes in student performance or behavior that is necessary for determining evidence-based practices.

There are several different single-subject designs that may be used to evaluate whether the intervention causes changes in the target behavior. One of the common features of most designs is that they use at least two conditions, a baseline and an intervention. Baseline data are collected in the same manner across several different sessions and describe how the target behavior occurs without the intervention in place, similar to a pretest measure. Once baseline levels of the target behavior are established, most single-subject designs then require an active intervention condition in which the educational practice is implemented over time and multiple measurements of the target behavior continue. Comparison of an individual’s performance of the target behavior during baseline to his or her performance during the intervention condition determines the effectiveness of...
the intervention. When changes in the target behavior are associated with changes in the intervention, the research demonstrates a functional relationship (Barlow & Hersen, 1984).

Single-subject research requires that at least one replication of the functional relationship (i.e., changes between baseline and intervention conditions) is included within the design (Alberto & Troutman, 2006). There are many designs used in single-subject research that show functional relationships, and each design differs in terms of the procedures for applying, withdrawing, or altering the intervention. Perhaps the two most straightforward designs used in single-subject research are the reversal and the multiple-baseline designs.

Reversal design. The reversal design (ABAB) is one of the most powerful single-subject research designs because it can clearly show the relationship between the implementation of the intervention and changes in the target behavior. This design requires that the intervention be sequentially introduced and taken away in different conditions. The design usually uses four conditions. Initial baseline performance of the target behavior is collected first (the first A phase), followed by assessment of the target behavior while the intervention is in place (the first B phase). These are followed by the withdrawal of the intervention or a “return-to-baseline” condition (the second A phase) and, finally, the reinstatement of the intervention (the second B phase). This design is illustrated in Figure 1. In this illustration, a functional relationship between the target behavior and the intervention is established by repeatedly measuring the target behavior during all baseline and intervention conditions and showing that the target behavior consistently occurred at different levels depending on whether the intervention was in place (the two B phases) or not (the two A phases). Confidence in the functional relationship increases when student behavior changes meaningfully and in the predicted direction corresponding to the introduction and withdrawal of the intervention.

Dettmer, Simpson, Myles, and Ganz (2000) used a reversal or ABAB design to test the effects of visual schedules designed to help 2 boys with autism transition faster from one classroom activity to another with fewer teacher prompts. Portable schedules with icons were used to remind the children where they had been, where they were going, and the next activity in the schedule. During baseline conditions, the time spent in transition averaged 2.5 min, but once the visual schedules were introduced, the time decreased to an average of 0.7 min. Moreover, the number of verbal prompts required from teachers decreased from an average of 17 during baseline to 6 during the first intervention condition. Once the target behaviors were stable during the intervention condition, the visual schedules were removed (the reversal or return to baseline in this second A phase), and the target behaviors increased to average levels similar to where they had been during the initial baseline conditions (2.3 min and 15 prompts). The intervention of visual schedules was then reinstated, and the target behaviors again decreased in frequency. By evaluating changes in student behavior in relation to the introduction and withdrawal of the intervention, the authors were able to conclude that visual schedules caused a reduction in the time spent between activities and the number of teacher prompts necessary for students to make transitions successfully.

Multiple-baseline designs. Although the reversal design establishes a functional relationship between the target behavior and the intervention, there are practical concerns that may arise from withdrawing interventions from participants who have learned a particular behavior (e.g., behaviors that cannot be reversed, such as reading) and ethical concerns when participants experience success with an intervention in place and the intervention should not be withdrawn (e.g., successful reduction of self-injurious behaviors) (Baer, Wolf, & Risley, 1968). The multiple-baseline design was developed in part in response to these concerns and is now one of the more commonly used designs in special education and intervention research. The multiple-baseline design incorporates a baseline and an intervention condition (AB) across participants (i.e., more than one participant), behaviors (i.e., more than one target behavior), or settings (i.e., more than one environment). Once baseline levels of the target behavior have been established for participant, behavior, or setting, the intervention is introduced in a staggered sequence (at a different time) for each next participant, behavior, or setting. This sequential introduction continues until the intervention is introduced in succession for each.
participant, behavior, or setting (Alberto & Troutman, 2006). In this fashion, a functional relationship is established if the target behavior changes when and only when the intervention is introduced to another person, targeting another behavior, or in another environment. This is illustrated in Figure 2.

Stormont, Smith, and Lewis (2007) used a multiple-baseline-across-participants design to investigate the effects of a teacher-implemented schoolwide positive behavior support (PBS) intervention on the reduction of students’ problem behavior. Three Head Start teachers were trained in PBS techniques, and their implementation of the use of precorrective and praise statements was individually staggered across baseline and intervention conditions. The teachers’ use of PBS strategies increased from an average of 30 during baseline to an average of 92 during the intervention condition, and the rate of students’ problem behavior decreased each time the teachers began using the PBS strategies in small-group settings. Because the interventions were introduced at different times and each introduction of the intervention was associated with a unique change in student outcomes, one can be confident that it was the intervention, rather than some other event (e.g., the hiring of a new principal), that caused the decrease in students’ problem behaviors.

**Analysis of Effects Within Participants**

Although single-subject research results can be examined through statistical methods, the traditional way of assessing intervention effectiveness is through visual inspection, a process with systematic rules for evaluating a graphic representation of the continuous data gathered for baseline and intervention conditions. As stated previously, these comparisons are made within participants. What is assessed is how each individual participant’s target behavior changes from baseline to intervention conditions. The within-participant changes in performance are typically evaluated according to strength or magnitude of the target behavior (mean and level) across conditions and the rate of these changes (trend and latency; Kazdin, 1992). Each of these criteria is described in the following sections.

**Mean.** The mean performance is calculated and compared for each condition in the design to determine the impact of the intervention. If mean performance during the intervention condition(s) is meaningfully better than performance during the baseline condition(s), the intervention shows evidence of effectiveness (Alberto & Troutman, 2006). For example, if a child’s mean completion of math problems per class session increases from 3.0 during a baseline condition to 13.5 when a self-monitoring intervention is implemented, we have one piece of evidence that the intervention was effective.

**Level.** The change in performance that occurs just after the intervention is implemented or withdrawn is referred to as a change in level. Target behaviors that show an immediate or abrupt change in level from the end of one condition (e.g., the last data point in baseline) to the beginning of the next condition (e.g., the first data point in intervention) demonstrate a strong reaction to the intervention. Researchers look at both the immediacy and the magnitude of change in level as indicators of the strength of the intervention’s effect on the target behavior (Kazdin, 1992).

**Trend.** Another aspect of determining the effectiveness of an intervention is the change in trend or direction of the data as the intervention is applied and withdrawn. A trend in data refers to the tendency of a series of data points to systematically increase (an ascending trend) or decrease (a descending trend) over time. Interventions that cause an increase in a positive behavior should, generally speaking, result in an ascending trend in outcomes after being implemented. Consider an example of baseline data showing a decreasing trend, followed by an intervention that results in an increasing trend in data during the intervention phase. This change in trend provides solid evidence of the intervention’s effect on target behavior.

**Latency of change.** Latency of change refers to the quickness with which the behavior changes at the termination of one condition (baseline or intervention) and
onset of another. The shorter the time frame in which change occurs, the clearer the effect of the intervention on the target behavior. If the behavior takes a long time to change after the intervention is introduced, for example, there is less confidence in determining that the effect is the result of the intervention alone (Alberto & Troutman, 2006). It is important for researchers to consider the nature of their interventions when evaluating latency, however, as in some cases, a response to a particular intervention may take several trials to become evident (e.g., when a child is learning a new problem-solving routine that takes time to master or use fluently).

**Importance of Single-Subject Research in Special Education**

Special education encompasses a range of specialized, often individualized, educational services designed to meet the unique needs of individuals with disabilities. As such, the practice of special education in schools lends itself in several ways to single-subject research designs, a hallmark of which is the assessment of individual performance and behavior change over time. Two elements of single-subject research methods, in particular, that should be part of special education teachers’ routines and repertoires are (a) frequent measurement of behavior and (b) monitoring of the effects of interventions.

It would be hard to imagine an effective special education teacher who does not collect data on students’ performance on a frequent basis. Although new emphases on standards and accountability have educators focusing on issues such as progress monitoring, special education teachers need to collect and analyze performance data on an almost daily basis for many routine academic and social behavioral tasks, particularly, those that are included on a student’s individualized education program. Accordingly, the repeated measurement of performance and behavior required by single-subject research should be a regular part of most special educators’ instructional routines. Recording and even graphing the number of words students read correctly, the number of math problems they complete, the number of times they raise their hand before being called on—any behaviors that are of particular concern for an individual student—should be the norm.

In tandem with recording data, special education teachers should also systematically evaluate the effects of their interventions. This requires at least a basic understanding of the fundamental tenet of single-subject research designs, namely, the assessment of a target behavior under both baseline and treatment conditions. Although it is probably not necessary that teachers be required to implement full-blown research studies with elaborate designs (e.g., multiple baseline) for each instructional intervention they implement, for the purpose of communicating with others (e.g., administrators, parents, general educators) as well as being confident about what to use in their own classrooms, expert special educators should document the effectiveness of interventions for individual students by comparing repeated measures of student performance prior and subsequent to the implementation of interventions. In this way, single-subject research can play an important role in fulfilling the goal of providing an individualized education for students with disabilities.

**Conclusion**

As outlined in Table 1, there are several key elements of single-subject research methods: (a) measuring target behaviors repeatedly, (b) applying interventions
TABLE 1
Considerations Regarding Single-Subject Research

- Single-subject research designs are appropriate for answering research questions that allow comparisons to be made within participants and for which interventions can be introduced and/or withdrawn systematically.
- Target behaviors must be assessed repeatedly using trustworthy measurements.
- Researchers must measure target behaviors the same way on repeated occasions.
- Measurement of the target behavior must occur multiple times during each baseline and intervention condition.
- Interventions must be systematically introduced and withdrawn repeatedly.
- The effects of the intervention are evaluated when conditions change (e.g., from baseline to intervention, from intervention to baseline).
- The effectiveness of the intervention is evaluated within participants.
- Changes between baseline and intervention conditions are assessed within individuals rather than between individuals or groups.
- Changes in a participant’s performance across conditions are evaluated according to the mean, level, trend, and latency of the changes in his or her observed behavior.
- Single-subject research designs include explicit evaluation of causality and thus can be used to determine whether a practice is evidence based.
- Given typically small sample sizes, multiple single-subject studies are needed to determine definitively whether a practice is effective.

systematically, and (c) evaluating outcome effectiveness. And these elements parallel the practice of special education. Although the rich history of applied behavior analysis and single-subject research methodology is well beyond the scope of this article (see Kazdin, 1978; Tawney & Gast, 1984), it is nevertheless important to note that as an experimental approach to research, single-subject designs offer much to the discussion and ultimate validation of practices as evidence based. Horner et al. (2005) outlined a series of indicators that researchers may use to evaluate the quality of single-subject research studies. Their thorough overview of the characteristics of single-subject research included discussion of criteria for evaluating various aspects of a single-subject research study, including the rigor of the design, methods of measurement and the trustworthiness of those measures, and description of participants and the interventions being examined.

Horner et al. (2005) further discussed the ways in which bodies of single-subject research studies of a particular intervention may be used, taken together, to establish a practice as evidence based. It is important to emphasize this point: Because individual single-subject research studies typically involve relatively few participants, results must be replicated across several studies before generalizations about the evidence of a practice can be drawn. Yet Horner et al.’s assessment that single-subject research has much to offer to the discussion of evidence-based practice because these designs allow conclusions about whether a practice is effective remains accurate. As special education grapples with and helps to shape the evidence-based practice movement, single-subject research will play a significant role. The current educational improvement agenda calling on teachers to use the best available scientific evidence as a basis for formulating educational interventions for individual children seems to point in particular to the use of single-subject research methods by both researchers and practitioners as interventions are developed, tested, and translated into practice.

About the Authors

Melody Tankersley, PhD, is a professor of special education at Kent State University. Her current scholarship interests are related to the prevention of emotional and behavioral disorders, promoting positive classroom behavior, and bringing research to bear upon practice. Sanna Harjusola-Webb, PhD, is an assistant professor of early childhood special education at Kent State University. Her research interests include prevention and early intervention for young children, teacher preparation, evidence-based practices, and issues related to intervention fidelity. Timothy J. Landrum, PhD, is an associate professor on the general faculty and senior scientist in the Curry School of Education at the University of Virginia. His research and writing focus on classroom and behavior management, the identification of evidence-based practices in special education, and the translation of research into practice. Address: Melody Tankersley, 405 White Hall, Kent State University, Kent, OH 44242; e-mail: mtankers@kent.edu.

References


