# Single-Case Experimental Design: Current Standards and Applications in Occupational Therapy

Justin D. Lane, Jennifer R. Ledford, David L. Gast

Occupational therapy is a field with a long-standing history of recommending and implementing interventions designed to improve the quality of life of clients with disabilities. Often, the interventions are individualized to meet the needs of this diverse group of clients in dynamic settings. Identifying effective and efficient interventions for such a diverse group of clients and settings requires a flexible research approach. Single-case experimental designs (SCEDs) allow practitioners and researchers to answer experimental questions within the context of rigorous research designs. The purpose of this article is to highlight the similarities between the mission of occupational therapy and SCEDs. Recommendations for designing single-case studies with the framework provided by the Single-Case Reporting Guideline in Behavioral Interventions are provided. In addition, common problems and proposed solutions, along with implications for practitioners and researchers, are provided.

Lane, J. D., Ledford, J. R., & Gast, D. L. (2017). Single-case experimental design: Current standards and applications in occupational therapy. *American Journal of Occupational Therapy, 71,* 7102300010. https://doi.org/10.5014/ajot.2017.022210

**Justin D. Lane, PhD, BCBA-D,** is Assistant Professor, Department of Early Childhood, Special Education, and Rehabilitation Counseling, University of Kentucky, Lexington; justin.lane@uky.edu

**Jennifer R. Ledford, PhD, BCBA-D,** is Assistant Professor, Department of Special Education, Vanderbilt University, Nashville, TN.

**David L. Gast, PhD,** is Professor, Department of Communication Sciences and Special Education, University of Georgia, Athens.

In occupational therapy and related fields, identification of effective and efficient interventions for people with, or at-risk for, disabilities is critical. Given that practitioners typically have a finite number of resources available, use of evidence-based practices is advantageous for all involved parties: When practitioners use high-quality interventions, it increases the likelihood of significant changes in clients' adaptive and functional skills, which, in turn, leads to an improved quality of life for clients (e.g., Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996). Moreover, the use of these interventions has the potential to result in favorable client outcomes in the short-term, and it may also increase positive perception of the field and increase client retention, allowing for continued long-term improvements. Three important procedural steps can improve the use of evidence-based practice. The first is the identification of effective practices in the context of methodologically rigorous research. The next is to disseminate findings to practitioners. Finally, training practitioners to systematically evaluate interventions using rigorous designs may promote data-based decision making in the field—a skill set that allows practitioners to identify effective and ineffective interventions for specific clients and to make modifications as needed. The purpose of this article is to provide an overview of the uses of single-case experimental designs (SCEDs), to review current standards for the use of SCEDs, and to describe their use in occupational therapy, including potential difficulties and possible solutions.

## Selecting an Experimental Design

Several research approaches are available for evaluating interventions: (1) group experimental research, (2) qualitative research approach, and (3) SCEDs (Gast, 2014; Kennedy, 2005). Group experimental research typically includes at least two equivalent groups of participants who are assigned to either treatment or control conditions or to a wait-list control condition. Possible differential effects of receiving or not receiving an intervention are compared and evaluated with appropriate statistical tests. A randomized controlled trial is typically considered the gold standard of experimental research. Given the potential generality of findings to the broader population, the impact of these designs is without question; however, they are better suited for answering population-based inquiries. Findings are based on mean responding of a group, which does not account for the individual variability common among clients typically served by occupational therapy practitioners. Another option is to use a qualitative research approach, which encompasses a collection of descriptive designs (e.g., case study, narrative) that allows researchers to evaluate common themes in nature and to develop a better, more detailed understanding of certain phenomena. Although informative, a qualitative research approach does not allow for intervention evaluation, which is typically of interest in occupational therapy and related fields.

In contrast, SCEDs are well suited for formatively evaluating the therapeutic benefits of an intervention within and across clients. SCEDs have a long-standing history in the intervention literature. SCEDs are not the same as case studies, which are part of a qualitative research approach (Kazdin, 2010; Kennedy, 2005), and they are also not pretest–posttest designs. When using SCEDs, researchers can answer questions about the effectiveness of an intervention (initial attempts to demonstrate and replicate effects across participants) as well as compare and identify superior interventions (e.g., compare an established intervention with a novel treatment approach). Each client serves as his or her own control, meaning that his or her performance under one condition is compared with performance under a different condition.

In *demonstration* designs, analyses are typically conducted for data collected during adjacent baseline and intervention conditions. In *comparison* designs, analyses are conducted between two different intervention conditions, although a baseline condition may also be included. The same dependent variable, or target behavior, is measured repeatedly in each primary comparison condition; this characteristic differentiates SCEDs from pretest–posttest designs (even those with small numbers of participants), which operate according to the nomothetic approach used in group experimental designs. Researchers attempt to replicate therapeutic effects across participants, behaviors, or conditions (e.g., arrangements, settings) to strengthen the external validity (or applicability) of findings to clients who display similar preintervention characteristics. Extending findings beyond an individual SCED is a matter of replication within a study and across studies (e.g., conducted by other researchers, with different participants, and in various settings).

The extent to which findings from SCEDs are generalizable is oftentimes criticized, but Sidman (1960) highlighted the utility of replicating an effect within a participant and across participants, stating that "replication of an experiment with two subjects establishes greater generality . . . among the individuals of a population than does replication with two groups [whose data have been combined]" (p. 75). Such criticisms of SCEDs are addressed through independent and systematic replications of studies (Gast, 2014). Studies in which SCEDs are used typically include multiple participants; the number is dependent on the research question and design type.

SCEDs are well suited for evaluating interventions conducted in applied settings where practitioners provide therapeutic services, such as community clinics, schools, and homes (Kennedy, 2005; Mechling & Ayres, 2012). Compared with highly controlled settings, such as research clinics, these dynamic environments present researchers with the unique challenge of not only determining whether an intervention is effective but also determining whether it is effective when implemented in ever-changing, real-world environments (Odom et al., 2005; Umeda & Deitz, 2011). Under these circumstances, it is recommended that applied researchers formatively evaluate change over time by closely monitoring progress and, as needed, make modifications to an intervention. In addition, researchers interested in answering questions related to the impact of an intervention might want to know under what conditions (e.g., treatment room with parent or home with parent) and for whom (e.g., adolescents with autism spectrum disorder) interventions are effective.

When answering these questions, SCEDs allow researchers to evaluate the impact of an intervention by collecting data on a day-by-day or session-by-session basis. Another benefit of an SCED is that it allows for ongoing formative evaluation, which enables the adoption of changes or modifications if an intervention is not working as planned (cf. Cox, Gast, Luscre, & Ayres, 2009). This type of decision making corresponds well with ethical clinical practice. Large-scale group studies generally involve summative evaluation and thus do not involve session-by-session data collection and decision making, although some recent work in adaptive designs allows longer term, data-based decision making (e.g., reassigning group after a set period of time with little or no growth; Nahum-Shani et al., 2012). However, even these group comparison designs do not solve associated problems with the need for large numbers of participants, and they do not allow for the type of formative day-to-day decision making that is typical of practice.

Researchers have multiple options when selecting an SCED to evaluate the effects of an intervention. Three common types of SCEDs include (1) introduction and removal of an intervention at different points in time (withdrawal design), (2) time-lagged introduction of an intervention (multiple baseline and multiple probe designs), or (3) a rapid iterative introduction and withdrawal of interventions (alternating treatments design and adapted alternating treatments design). Descriptions of the uses of these design types and their common variations are shown in Table 1, and in-depth information about design types is available elsewhere (cf. Gast & Ledford, 2014; Shadish, Hedges, Horner, & Odom, 2015; for applied examples, see, e.g., Costigan & Light, 2010; Crowe et al., 2014; Mechling & Ayres, 2012; Umeda & Deitz, 2011).

Regardless of the design, the integrity of a study can become compromised when specific sources of error are not controlled during the course of a study. Common threats to internal validity (Table 2) include unplanned events outside the study that influence findings (e.g., client enrolls in a treatment program designed to target the same target behavior), human error or bias (e.g., researcher feels strongly that a treatment package that he or she designed is influencing outcomes, which results in purposeful or inadvertent data collection inaccuracies), procedural infidelity (e.g., researcher fails to implement experimental condition procedures as delineated in the method section of the research proposal), and other issues such as attrition (participants leave a study). Careful planning requires researchers to consider all potential sources of error and to control for them to the fullest extent possible. Current standards for SCEDs include procedures for minimizing these threats.

## Current Standards for Single-Case Experimental Designs

The Single-Case Reporting Guideline in Behavioral Interventions (SCRIBE) provides information on how to guide and structure reporting SCED research through the use of a checklist for researchers (Tate & Perdices, 2015; Tate, Perdices, McDonald, Togher, & Rosenkoetter, 2014). Items include design considerations and reporting standards. Reviewing each of the items included in the checklist is beyond the scope of this article; however, a summary of key points, especially as they relate to conducting an SCED, is warranted. Moving through a linear sequence of planning, implementation, and dissemination, the first set of items stresses the importance of building a rationale for a study with previously published literature. In some instances, an intervention may be commonly recommended, but it has little to no support in the experimental literature. If so, the need for initial research to support common practice should be highlighted. In addition, researchers must provide clearly defined research questions that allow readers to understand the question or questions the study is attempting to answer within the context of the selected design. Different formats are available for writing research questions; however, at a minimum, researchers should write questions that indicate the (1) independent variable, (2) dependent variable, (3) population (including age range and diagnoses), (4) context (conditions under which the independent variable will be implemented), and (5) an indication of how data will be quantified (e.g., increase in percentage of steps completed independently).

Next, clearly defined, well-written procedures are the crux of a method section when planning for and disseminating findings. Throughout this process, researchers need to continually consider, "With a baseline level of knowledge about a topic, could another person or team of researchers replicate these procedures with the same level of precision?" Researchers should describe the inclusion criteria and prerequisite skills required for participation. In addition, researchers should provide general information that describes participants' current levels of functioning and descriptive information related to the purpose of the study (e.g., strengths and needs in the

**Table 1. Purpose of Commonly Used Single-Case Experimental Designs**

| Design Type | Use |
|---|---|
| A–B–A–B (withdrawal) | Demonstrates the effectiveness of an intervention on reversible behaviors by alternating two conditions |
| A–B–C–B–C (multitreatment) | Compares the effectiveness of two interventions on reversible behaviors by alternating two conditions |
| Multiple baseline | Demonstrates the effectiveness of an intervention on reversible behaviors by introducing the intervention in a time-lagged fashion across at least 3 participants, behaviors, or contexts |
| Multiple probe | Demonstrates the effectiveness of an intervention on nonreversible behaviors by introducing the intervention in a time-lagged fashion across at least 3 participants, behaviors, or contexts |
| Changing criterion | Demonstrates the effectiveness of an intervention on reversible behaviors by introducing stepwise intervention requirements in a time-lagged fashion; generally used for reinforcement-based interventions to increase responding for behaviors already in a learner's repertoire |
| Alternating treatments | Compares the effectiveness of two interventions on reversible behaviors by rapidly alternating sessions |
| Adapted alternating treatments | Compares the effectiveness of two interventions on nonreversible behaviors by rapidly alternating sessions |

**Table 2. Threats to the Internal Validity of a Study**

| Threat | Definition (*Example*) | Procedures for Controlling or Detecting Threat (*Example*) |
|---|---|---|
| History | Events occurring during a study that affect the outcome and that are unrelated to planned condition changes. *When participating in a therapist-designed intervention to improve on-task behavior, a classroom teacher implements a new behavior plan. The teacher says that her behavior plan resulted in the changes, and she wants to discontinue the occupational therapy program.* | Include three or more potential demonstrations of effect. When change occurs, when and only when condition changes occur, there is confidence in the intervention. *The therapist measures on-task behavior during small group sessions with and without the intervention in an A–B–A–B design. The only difference between conditions is the therapist-designed program, allowing for assessment of whether that (and only that intervention) is responsible for behavior change* (Crowe et al., 2014; Umeda & Deitz, 2011). |
| Implementer bias and drift | Apparent changes in behavior that are due to a change in measurement, including an inadvertent change. *The therapist implementing the program designed to improve on-task behavior counted staring into space as off-task behavior during baseline conditions, but then she decided this behavior was indicative of listening and, thus, began coding it as on-task.* | Include data collection by a second observer who collects dependent variable data independently but simultaneously with the primary data collector (interobserver agreement [IOA] data). *To ensure she uses the same definitions and coding rules over time and is consistent with her written coding procedures, the therapist enlists a colleague to collect data on the same behaviors once or twice per week. When disagreements occur, they have discrepancy discussions to ensure they are coding according to written procedures* (Umeda & Deitz, 2011). |
| Maturation of child | Development over time of behaviors or skills outside of planned procedures. *A therapist uses a handwriting program to improve letter-writing for a kindergarten student, and she conducts weekly assessments to determine effectiveness. A parent questions whether the student would have made equal progress during the year without the program.* | Include three or more potential demonstrations of effect, and choose designs with rapid iteration when possible. Choose dependent variables likely to change when your independent variable is implemented, including tasks analyzing complex and slow-changing behaviors. *The therapist conducts baseline assessments and targets specific behaviors in the context of a multiple probe design across behaviors. When behaviors change, when and only when those behaviors are targeted, confidence that maturation is not responsible for behavior change is increased* (Crowe et al., 2014). |
| Procedural infidelity | Inability to attribute outcomes to changes in conditions because of a failure of implementers to follow planned procedures. *When using the handwriting intervention, the therapist implements all planned procedural components but also implements supplemental components, such as response-prompting procedures.* | Include observation and data collection on the implementation of planned procedures, including baseline and intervention procedures. *During baseline and intervention sessions, the second observer who collects IOA data also collects data regarding the steps of the intervention completed in each session. After one intervention session, data suggest that the therapist added unplanned steps; after this identified error, the therapist implemented the procedures with fidelity* (Costigan & Light, 2010). |
| Multiple treatment interference | The inability to determine which treatment is responsible for changes in behavior when two or more treatments are compared. *A therapist is interested in whether using a traditional chair versus a therapy ball results in more accurate fine motor performance and better attention. The child's accuracy and attention improves over time in both conditions.* | For reversible behaviors (e.g., attention), include a best alone condition after the comparison condition to minimize alternation effects and to increase the discriminability between conditions and the amount of time between sessions to minimize carryover effects. For nonreversible behaviors (e.g., accuracy), use an adapted alternating treatments design and use different but equivalent behaviors in each condition. *The therapist teaches the child to write five specific letters in the ball condition and five similarly difficult letters in the chair condition* (Mechling & Ayres, 2012). |

*Note.* References listed are sample articles with adequate controls for the specific threat listed, but they may include inadequate experimental control or other methodological problems.

domain of interest) as well as any scores obtained from formal assessments and evaluations. More detailed descriptions of participants allow other researchers and practitioners to identify for whom interventions might or might not be effective and to extend findings beyond recruited participants.

When using a multiple probe or baseline design, researchers should consider randomizing cases (participants, behaviors, settings) to different tiers, assuming that data from all participants are stable. Researchers should also provide detailed descriptions of the setting, including people within the setting, and materials used during the study. As with participant descriptions, setting, implementer, and materials descriptions allow practitioners and researchers to determine in what contexts the intervention is likely to be successful.

A technological description of prescribed procedures for all conditions in a study (e.g., baseline, intervention, generalization) allows other research teams, and consumers of research, to better understand each component of a study, which reduces ambiguity and increases the likelihood that research teams can appropriately attempt to replicate effects. In addition, precise reporting of procedures assists in closing the gap between research and practice by allowing practitioners access to precise descriptions and effective interventions. When reporting results of a study, detailed descriptions of how data were analyzed also reduce ambiguity and allow for a clear understanding of how data were evaluated and interpreted.

### *Defining and Measuring Dependent Variables*

An SCED research approach involves repeated measurement of observable behaviors. Commonly, SCED researchers operationally define each dependent variable (describe a behavior with clearly defined and observable actions) and provide examples and nonexamples of each behavior (Wolery & Lane, 2014). Thus, when selecting a dependent variable (or variables), researchers who have never used an SCED might be hesitant or unsure of how to measure or represent broader conceptual ideas, such as motor functioning or self-efficacy. For example, *active engagement* refers to a broad construct that can be defined as a person meaningfully attending to and responding in his or her immediate environment. This construct can be anchored to or represented by observable and measurable actions, such as responding to adult directives within 5 s and orienting toward an adult when he or she is speaking. When these specific observable behaviors are measured, they can be considered a proxy for the broader concept of active engagement. By selecting specific operationally defined behaviors, researchers can ensure objective, valid, and reliable measurement of the construct of interest. Another procedure by which researchers can select dependent variables is to consider recommendations that a therapist makes to parents or teachers who work with their child or student at home or at school; what observable and measurable behavior(s) can they target in these applied settings that address or serve as a proxy for the overarching construct of interest?

When defining behaviors of interest, researchers should consider to which of two broad categories a behavior belongs: (1) reversible or (2) nonreversible. A *reversible behavior* is one that will increase or decrease on the basis of the presence or absence of an intervention, which means that the behavior is sensitive to changes in the environment (Gast, 2014). A *nonreversible behavior* is a behavior that, once learned, is expected to remain after an intervention is removed (e.g., developmental skills). Carefully considering operational definitions of behaviors and categorization as either reversible or nonreversible will assist in the selection of the most appropriate SCED. Although one dependent variable should be identified as the primary variable and the one that will drive experimental decisions, additional behaviors can also be monitored (cf. Reichow, Barton, Sewell, Good, & Wolery, 2010), such as related nontarget behaviors or generalized tendencies. In addition, construct validity is improved through the use of multiple measures; it can be further improved by conducting reviews by experts or blind observers.

After selecting a dependent variable, it is necessary to determine what aspect (or dimension) of a behavior should be measured during the study (e.g., how often or for how long a behavior occurs). Direct observation systems are commonly used to count (number, rate) or time (duration, latency) each occurrence of a behavior (Ayres & Ledford, 2014) or to estimate the duration of a behavior (i.e., momentary time sampling; Ledford, Ayres, Lane, & Lam, 2015). These direct observational recording systems allow researchers to summarize a participant's performance for a given day or session. After selecting the appropriate dimension and recording system, interobserver agreement (IOA) data should be collected for at least 25% of the sessions for each condition (Wolery, Dunlap, & Ledford, 2011) to ensure reliable measurement of the dependent variable. Ideally, independent reviewers, blind to the purpose of the study, should be selected to collect data to reduce potential biases that might arise during the course of the study (e.g., Umeda & Deitz, 2011). If this is not possible, we recommend that at least one independent reviewer collect IOA data. When developing behavior definitions, two independent observers should be able to reliably note the occurrence of a behavior (e.g., should "agree" when it occurred or did not occur by separately measuring and then comparing data). This measurement of IOA should occur regularly throughout baseline and treatment conditions, and observers should be retrained if decrements in agreement are noted. Evaluating agreement in this way ensures that observers are measuring the same behaviors over time, as intended, and that changes in behavior between conditions are due to the intervention rather than observer bias.

### Interpreting Data

The most common method for evaluating data collected within the context of an SCED is visual analysis of graphic displays of data. A line graph is the most commonly used mode for presenting change over time within and across participants, behaviors, or settings. Providing a detailed task analysis of how to visually analyze data is beyond the scope of this article, but a "how to" article is available in the literature (see Lane & Gast, 2014). Investigators should identify a priori what types of changes are expected (i.e., changes in level, trend, or variability). Most often, researchers are interested in changes in level, but they may expect an initial change in trend, which eventually leads to a higher or lower level when compared with baseline data (e.g., when teaching letter-writing, a child may perform no behaviors correctly in baseline and may gradually increase accuracy over time with intervention).

For some behaviors, an immediate or near-immediate change in level should be expected (e.g., when manipulating environmental characteristics to improve on-task behaviors). A *behavior change* occurs when data are consistently and predictably different from one condition to another adjacent condition. A *functional relation* is present when these consistent and predictable changes occur every time a condition change occurs, with at least three demonstrations of effect. These demonstrations can occur when two conditions are alternated at least twice each with multiple sessions per alternation (e.g., A–B–A–B or A–B–C–B–C but not A–B–A–C), when two or more conditions are alternated in single sessions (e.g., alternating treatments design or adapted alternating treatments design with rapid alternation between conditions and at least three data points in every condition), or when an intervention is introduced in a time-lagged fashion (e.g., multiple baseline or multiple probe designs with at least three tiers and at least three start points).

When an SCED is used, consistency and replication are essential characteristics for a functional relation determination—large differences in level are not. You may decide to quantify differences in behavior by reporting a median level and range in each condition (mean or average is only appropriate if there are no trends and limited variability with no outliers) or by reporting the extent to which data from one condition overlap with another (e.g., percentage of non-overlapping data; Scruggs, Mastropieri, & Casto, 1987). More complicated overlap-based metrics (e.g., improvement rate difference, Tau-U) also exist, but these metrics are generally not appropriate because values have problems similar to those inherent in percentage of non-overlapping data (cf. Wolery, Busick, Reichow, & Barton, 2010) and because values are sensitive to variables such as the use of interval systems and session length (Ledford et al., 2015; Pustejovsky, 2015). More complicated methods for comparing data are also available (see Shadish et al., 2015), but most of these methods include assumptions that are not likely to be true when SCEDs are used (e.g., no trend, many data points in each condition, independence of data points). When used, all calculations and conclusions drawn from mean or overlap-based procedures should be secondary to visual analysis of data. Similarly, standardized assessments, narrative records, social validity scales, and related formal and informal measures should be considered secondary to data collected within the context of an experimental design.

## Applications in Occupational Therapy

Occupational therapy practitioners address several areas of need for clients in clinics, homes, schools, and related settings. Each client presents his or her own unique set of strengths and challenges; this variation leads to the need for individualization of interventions. As the field of occupational therapy moves forward in identifying scientifically validated interventions and supports, practitioners and applied researchers need an experimental approach that is adaptable; SCEDs offer a cadre of experimental designs that are flexible, but rigorous, and can meet these needs of the field. SCEDs are a reputable set of experimental designs with decades of support, but, as with any experimental approach, questions or concerns related to how to adequately and appropriately answer research questions are likely to arise in the planning process (see Table 3). The following sections highlight optimal choices when planning studies, whether as an applied researcher in the field or as a researcher objectively evaluating a practice.

### Clients With Unrelated Goals

Oftentimes, clinicians and researchers interact and work with clients with different needs and, as such, believe that they cannot conduct an experimental study. Although a case study or A–B design could be selected (lacking adequate experimental control to answer a research question), it is possible to select an SCED that can experimentally answer a research question with only one or two clients (see Table 3). For example, a withdrawal design or multiple probe design across behaviors can be used with a single client (e.g., Costigan & Light, 2010; Umeda & Deitz, 2011). A practitioner or researcher could attempt to replicate effects with an additional client, but these designs do not require concurrent measurement across clients. One of the benefits of SCEDs is the ability to experimentally answer research questions when limited numbers of clients are available.

### Nonreversible Behaviors

In practice, consumers and clients often request interventions leading to lasting change in target behaviors (i.e., accurate performance is maintained after intervention is removed). Some skills tend to be nonreversible in nature (e.g., writing letters and using clothing fasteners). When evaluating acquisition of behaviors that are not readily reversible, it is critical to select an SCED that has adequate experimental control for evaluating those behaviors (i.e., multiple probe rather than A–B–A–B designs; see Table 3). Although it is preferable to evaluate maintenance of these behaviors, withdrawal of the intervention should not be conceptualized as the use of an A–B–A design because we expect that these behaviors will not change when intervention is withdrawn; moreover, at least one additional condition change (i.e., A–B–A–B) is needed to meet contemporary guidelines (Shadish et al., 2015). Even when reversible behaviors are measured, nonwithdrawal designs can be beneficial because they do not require

**Table 3. Common Issues in Applied Research With Possible and Superior Solutions**

| Potential Problem | Possible and Common Solution | Superior Solution | Examples | Example Solution |
|---|---|---|---|---|
| Concerned with single participants with individualized therapy procedures and goals | Do not attempt study with experimental control; use case study or A–B design | Use designs that require 1 or 2 participants | 1. Two clients need to learn specific fine motor behaviors | 1. Use a combination multiple probe across participants and behaviors design to assess the use of visual supports to teach chained fine motor tasks |
| | | | 2. One client needs assistance with staying on task during large group activities | 2. Use an A–B–A–B withdrawal design to assess the use of therapy balls on on-task behavior |
| Interested in target behaviors that are not readily reversible | Do not attempt study with experimental control; use A–B design | Use designs that do not require intervention withdrawal | 1. One client needs to learn to write her name | 1. Use a multiple probe design across behaviors to assess the use of fading to teach letter drawing (2 letters per tier; 6 total) |
| | | | 2. Three clients need to learn self-feeding behaviors | 2. Use a multiple probe design across participants to assess the use of graduated guidance to teach self-feeding |
| Interested in complex behaviors and concepts | Do not attempt to measure observational behaviors; rely on professional opinion | Operationalize important behavior making up complex chains | 1. Two clients need to learn to identify and manage their emotional responses | 1. Identify a specific context in which behaviors can be operationalized and measured—for example, when playing competitive games and the percentage of intervals during which positive, neutral, and negative vocalizations are directed to peers |
| Measure behaviors that change slowly (e.g., unlikely to change immediately when intervention begins) | Do not attempt to measure behavior directly; rely on subjective reports or rating scales | Use designs for which stepwise changes are appropriate | 1. One client needs to improve the ability to complete simple tasks | 1. Use a changing criterion design to assess reinforcement and shaping to improve task engagement for increasing durations |
| | | | 2. Several clients need to improve their acceptance of different varieties of food | 2. Use a combination multiple baseline across food acceptance behaviors (touch, smell, taste, eat) and changing criterion design for number of bites eaten |

removing an effective intervention, which might be objectionable to some consumers.

### *Complex Behaviors and Concepts*

Similar to situations in which a practitioner or researcher might avoid conducting a study because he or she has access to a limited number of clients, practitioners and researchers may also avoid experimentally evaluating complex behaviors within the context of a single-case design. For example, some occupational therapists might argue that the measurement of such generalized tendencies, like independence, motor functioning, or sensory processing, is unnecessary. However, occupational therapists can measure these conceptual categories by identifying observable behaviors associated with these outcomes (e.g., initiations, completion of tasks, challenging or adaptive behavior; e.g., Crowe et al., 2014). When interested in quantifying such behaviors, practitioners and researchers should (1) consider what dimension or dimensions (e.g., count- or time-based assessment of a behavior) of the behavior they are interested in measuring, (2) consider the situations in which the target behavior is likely to occur, and (3) identify all steps necessary to complete the complex behavior. For example, when evaluating a method for promoting the use of scissors during an art activity, the practitioner or researcher should determine all the steps necessary to engage in the target behavior, beginning with picking up the scissors or obtaining the scissors from a designated area through completion. This task analysis process allows for accurate measurement of complex behavior chains.

### *Slow or Stepwise Behavior Change*

Although an immediate and abrupt change in a target behavior is typically desired when conducting an SCED study, practitioners and researchers may be interested in behaviors for which stepwise changes are expected. If so, a changing criterion design (Ledford & Gast, 2014) can be selected, or specific steps can be identified for measurement to improve sensitivity to change (see Table 3). A changing criterion design involves selecting predetermined criteria for evaluating change over time. For example, when evaluating a new feeding intervention, the researcher's initial expectation is that a client will take 1 bite of a nonpreferred food for three consecutive sessions and then 2 bites and so on until a predetermined mastery criterion is reached (e.g., 10 bites of nonpreferred foods).

When a changing criterion design is used, the magnitude of change can vary from criterion to criterion. Similar to other SCEDs, a minimum number of data points should be collected with each stepwise change to ensure a consistent pattern of responding. When specific steps are involved (as in the cutting example previously described), researchers can measure completion of each step rather than independent completion of the entire chain—this method allows for sensitive measurement of behavior change over time. Moreover, task analysis of complex behaviors can allow for the use of a multiple probe across behaviors design (e.g., target writing of two letters in the first tier, two other letters in the second tier). Experimental control is demonstrated when a child learns skills being taught while not acquiring untaught skills; behavioral covariation is, however, a potential problem (cf. Cosby, McLaughlin, & Derby, 2009).

## Implications for Occupational Therapy Practice

The findings of this study have the following implications for occupational therapy practice:

- *Operationalizing behaviors.* Clearly written, precise descriptions of target behaviors ensure appropriate measurement and facilitate replication attempts; complex concepts should be measured through careful identification of related, observable behaviors.
- *Selecting a design.* The research question determines the experimental design. When asking questions about an intervention or a common or novel practice, practitioners and researchers first determine whether the question is attempting to demonstrate the effects of the intervention or practice or compare such independent variables. This decision narrows the scope of options to specific, appropriate SCEDs. Being familiar with SCEDs, along with behaviors of interest (e.g., reversible, nonreversible) ensures that a research question can be adequately answered.
- *Methodological rigor.* When answering research questions with SCEDs, practitioners and researchers are encouraged to consider guidelines for conducting methodologically rigorous studies. SCRIBE guidelines provide detailed recommendations for ensuring that researchers conduct methodologically rigorous studies and follow standard reporting practices.
- *Interpreting findings.* The most beneficial information for the field is generated when appropriate designs are used, contemporary guidelines for rigor are followed, positive outcomes occur, and researchers replicate findings within and across participants.

## Conclusion

The purpose of this article was to highlight the utility of SCEDs in applied settings when evaluating the effectiveness of interventions for remediating deficits and delays in clients who receive occupational therapy. SCEDs allow practitioners and researchers to formatively evaluate the effects of an intervention over time and to make modifications as needed. This ability is especially critical when identifying interventions for vulnerable populations who typically require intensive support to learn to function independently within their environments. Designing a study with single-case research methodology requires precision in design and reporting. This process ensures that findings are meaningful, regardless of whether an intervention is effective or ineffective.

Although SCEDs have a long-standing history in the literature as an experimental approach for evaluating interventions, confusion still exists regarding the utility of SCEDs in practice and research. Some confusion stems from the name itself. The phrase *single case* in *single-case experimental designs* leads some researchers to think that SCEDs are case studies, which they are not. SCEDs include several designs that allow practitioners and researchers to formatively evaluate the impact of an intervention with a small number of participants, when compared with a group design approach. When considering models for building an evidence base, SCEDs play a crucial role in the neurorehabilitation research process (Whyte, Gordon, & Rothi, 2009), and they can be crucial for identifying what works, for whom, and under what conditions. Recognizing the utility of SCEDs within and across disciplines provides practitioners and researchers with more than one single approach to answer experimental questions. In addition, given the dynamic nature of intervention research, especially in applied settings, SCEDs are a flexible, but methodologically strong, option for furthering the field and for improving clients' long-term outcomes. ▲

## References


Ayres, K., & Ledford, J. R. (2014). Dependent measures and measurement systems. In D. L. Gast & J. R. Ledford (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed., pp. 124–153). New York: Routledge.

Cosby, E., McLaughlin, T. F., & Derby, K. M. (2009). Using tracing and modeling with a Handwriting Without Tears worksheet to increase handwriting legibility for a preschool student with autism. *The Open Social Science Journal, 2,* 74–77. https://doi.org/10.2174/1874945300902010074

Costigan, F. A., & Light, J. (2010). Effect of seated position on upper-extremity access to augmentative communication for children with cerebral palsy: Preliminary investigation. *American Journal of Occupational Therapy, 64,* 596–604. https://doi.org/10.5014/ajot.2010.09013

Cox, A. L., Gast, D. L., Luscre, D., & Ayres, K. M. (2009). The effects of weighted vests on appropriate in-seat behaviors of elementary-age students with autism and severe to profound intellectual disabilities. *Focus on Autism and Other Developmental Disabilities, 24,* 17–26. https://doi.org/10.1177/1088357608330753

Crowe, T. K., Perea-Burns, S., Sedillo, J. S., Hendrix, I. C., Winkle, M., & Deitz, J. (2014). Effects of partnerships between people with mobility challenges and service dogs. *American Journal of Occupational Therapy, 68,* 194–202. https://doi.org/10.5014/ajot.2014.009324

Gast, D. L. (2014). General factors in measurement and evaluation. In D. L. Gast & J. R. Ledford (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed., pp. 85–104). New York: Routledge.

Gast, D. L., & Ledford, J. R. (Eds.). (2014). *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed.). New York: Routledge.

Kazdin, A. (2010). *Single-case research designs: Methods for clinical and applied settings.* New York: Oxford University Press.

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

Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation, 24,* 445–463.

Ledford, J. R., Ayres, K. M., Lane, J. D., & Lam, M. F. (2015). Identifying issues and concerns with the use of interval-based systems in single case research using a pilot simulation study. *Journal of Special Education, 49,* 104–117. https://doi.org/10.1177/0022466915568975

Ledford, J. R., & Gast, D. L. (2014). Combination and other designs. In D. L. Gast & J. R. Ledford (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed., pp. 346–376). New York: Routledge.

Mechling, L. C., & Ayres, K. M. (2012). A comparative study: Completion of fine motor office related tasks by high school students with autism using video models on large and small screen sizes. *Journal of Autism and Developmental Disorders, 42,* 2364–2373. https://doi.org/10.1007/s10803-012-1484-1

Nahum-Shani, I., Qian, M., Almirall, D., Pelham, W. E., Gnagy, B., Fabiano, G. A., . . . Murphy, S. A. (2012). Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychological Methods, 17,* 457–477. https://doi.org/10.1037/a0029372

Odom, S. L., Brantlinger, E., Gersten, R., Horner, R. H., Thompson, B., & Harris, K. R. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 71,* 137–148. https://doi.org/10.1177/001440290507100201

Pustejovsky, J. E. (2015, April). *Operational sensitivities of non-overlap effect sizes for single case designs.* Paper presented at the annual meeting of the American Educational Research Association, Chicago. Retrieved from https://osf.io/pxn24/

Reichow, B., Barton, E. E., Sewell, J. N., Good, L., & Wolery, M. (2010). Effects of weighted vests on the engagement of children with developmental delays and autism. *Focus on Autism and Other Developmental Disabilities, 25,* 3–11. https://doi.org/10.1177/1088357609353751

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *BMJ, 312,* 71–72. https://doi.org/10.1136/bmj.312.7023.71

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education, 8,* 24–33.

Shadish, W. R., Hedges, L. V., Horner, R. H., & Odom, S. L. (2015). *The role of between-case effect sizes in conducting, interpreting, and summarizing single-case research* (NCER 2015-002). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education.

Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology.* New York: Basic Books.

Tate, R. L., & Perdices, M. (2015). N-of-1 trials in the behavioral sciences. In J. Nikles & G. Mitchell (Eds.), *The essential guide to N-of-1 trials in health* (pp. 19–41). Dordrecht, the Netherlands: Springer. https://doi.org/10.1007/978-94-017-7200-6_3

Tate, R. L., Perdices, M., McDonald, S., Togher, L., & Rosenkoetter, U. (2014). The design, conduct and report of single-case research: Resources to improve the quality of the neurorehabilitation literature. *Neuropsychological Rehabilitation, 24,* 315–331.

Umeda, C., & Deitz, J. (2011). Effects of therapy cushions on classroom behaviors of children with autism spectrum disorder. *American Journal of Occupational Therapy, 65,* 152–159. https://doi.org/10.5014/ajot.2011.000760

Whyte, J., Gordon, W., & Rothi, L. J. G. (2009). A phased developmental approach to neurorehabilitation research: The science of knowledge building. *Archives of Physical Medicine and Rehabilitation, 90*(Suppl.), S3–S10. https://doi.org/10.1016/j.apmr.2009.07.008

Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *Journal of Special Education, 44,* 18–28. https://doi.org/10.1177/0022466908328009

Wolery, M., Dunlap, G., & Ledford, J. R. (2011). Single-case experimental methods: Suggestions for reporting. *Journal of Early Intervention, 33,* 103–109. https://doi.org/10.1177/1053815111418235

Wolery, M., & Lane, K. L. (2014). Literature reviews, research proposals, and final reports. In D. L. Gast & J. R. Ledford (Eds.), *Single case research methodology: Applications in special education and behavioral sciences* (2nd ed., pp. 50–84). New York: Routledge.