

What Is Evidence-Based Behavior Analysis?

Tristram Smith

University of Rochester Medical Center

Although applied behavior analysts often say they engage in evidence-based practice, they express differing views on what constitutes “evidence” and “practice.” This article describes a practice as a service offered by a provider to help solve a problem presented by a consumer. Solving most problems (e.g., increasing or decreasing a behavior and maintaining this change) requires multiple intervention procedures (i.e., a package). Single-subject studies are invaluable in investigating individual procedures, but researchers still need to integrate the procedures into a package. The package must be standardized enough for independent providers to replicate yet flexible enough to allow individualization; intervention manuals are the primary technology for achieving this balance. To test whether the package is effective in solving consumers’ problems, researchers must evaluate outcomes of the package as a whole, usually in group studies such as randomized controlled trials. From this perspective, establishing an evidence-based practice involves more than analyzing the effects of discrete intervention procedures on behavior; it requires synthesizing information so as to offer thorough solutions to problems. Recognizing the need for synthesis offers behavior analysts many promising opportunities to build on their existing research to increase the quality and quantity of evidence-based practices.

Key words: evidence-based practice, clinical trials, treatment effectiveness evaluation, behavior analysis, behavior modification

A Google search for *evidence-based behavior analysis* yields 16,500,000 hits. Most of the top hits contain affirmations that behavior analysis is indeed evidence based. Similar statements appear regularly in traditional media (e.g., Smith, 2012). Despite this refrain, many behavior analysts express misgivings about published operational definitions of the term *evidence based* (Green, 2008; O’Donohue & Ferguson, 2006). Such definitions usually include guidelines for rating the quality of individual studies, aggregating findings across studies, and classifying the overall level of empirical support for an intervention. For example, guidelines in education and psychology list specific indicators of high-quality studies, set a threshold for the number of such studies needed to establish an intervention as evidence

based, and delineate categories such as “probably” or “possibly” efficacious (Chambless & Hollon, 1998; Kratochwill & Stoiber, 2002; Odom et al., 2005; West et al., 2002). Many guidelines in medicine specify that the gold standard for identifying evidence-based interventions is a systematic, statistical analysis of data from multiple randomized controlled trials (RCTs), in which investigators assign large numbers of participants by chance to treatment or control groups (Guyatt, Oxman, et al., 2008). By examining how closely the available research approaches this standard and how favorable the results are, reviewers can rank or grade the intervention (United States Preventive Services Task Force, 2007).

Behavior analysts have exposed many limitations of RCTs (Johnston, 1988; Johnston & Pennypacker, 1993) and statistical methods for combining data from different studies (Baron & Derenne, 2000). One practical problem is that individuals who consent to enroll in RCTs and risk being assigned to the control group may differ from individuals who decline to enroll (Christenson, Carlson, &

Correspondence concerning this article should be addressed to Tristram Smith, Division of Neurodevelopmental and Behavioral Pediatrics, Department of Pediatrics, University of Rochester Medical Center, 601 Elmwood Ave., Box 671, Rochester, New York 14642 (e-mail: Tristram_Smith@URMC.Rochester.edu).

Valdez, 2002). Quasiexperimental studies, in which treatment and control groups are matched on participant characteristics but are assigned nonrandomly (e.g., according to the availability of providers to deliver the experimental treatment), may minimize this problem. Nevertheless, a more fundamental difficulty remains: Both RCTs and quasiexperimental studies test the average change within an intervention group instead of the more clinically relevant analysis of change made by each individual participant as a function of a specific intervention (Johnston, 1988).

Statistical metrics that merge data across studies compound the problems in RCTs (Salzberg, Strain, & Baer, 1987; Strain, Kohler, & Gresham, 1998). Basically, these metrics represent an average of the average change reported in different studies. The studies almost always vary from one another in terms of their intervention protocols, outcome measures, and criteria for enrolling participants. Thus, the metrics are derived from an amalgamation of data obtained by an assortment of methods. False precision is apt to be the result.

Because of these limitations of group studies (RCTs and quasiexperimental research), behavior analysts favor studies with single-subject experimental designs (SSEDs). SSED studies involve conducting repeated observations to compare an individual's behavior during a baseline period when the individual receives no intervention to the behavior in one or more intervention phases. They are intended to detect changes that occur as soon as an intervention begins. Thus, SSED studies are much better suited than group studies for establishing a functional relation between a particular intervention and change of a specific behavior (Johnston, 1988; Johnston & Pennypacker, 1993). If the functional relation is robust, the change should be evident from visual inspection of a graphical

display of the data, obviating the need for statistical analyses. Replications of intervention procedures within and across participants will elucidate how and when the relation occurs (Johnston, 1996).

In relation to evidence-based practice, behavior analysts advocate for the recognition of SSED studies as important sources of data (Detrich, 2008; Green, 2008) and have proposed criteria for classifying an intervention as evidence based on the strength of findings from such studies (Horner et al., 2005; Kratochwill et al., 2010; National Autism Center [NAC], 2009). These guidelines include indicators of high-quality reports, notably clearly described intervention procedures and multiple replications of intervention effects, assessed by direct, reliable observations of the target behavior (Horner et al., 2005; NAC, 2009). Reviews of SSED studies find numerous applied behavior-analytic (ABA) procedures that meet criteria for classification as evidence based (e.g., Kurtz, Boelter, Jarmolowicz, Chin, & Hagopian, 2011; Lee, 2005; NAC, 2009; Sulzer-Azaroff & Austin, 2000; Vegas, Jenson, & Kircher, 2007).

From such work, some behavior analysts conclude that we have persuasively rebutted standard definitions of evidence-based practice (Keenan & Dillenburger, 2011), replaced them with definitions of our own (e.g., Horner et al., 2005), and documented that many of our interventions merit this designation (e.g., NAC, 2009). However, such conclusions are too hasty, in this writer's judgment. Inspection of the ABA interventions that are described as evidence based and the nature of empirical support adduced for them reveal that some basic questions remain unanswered.

First of all, what is a practice? Lists of evidence-based ABA practices are an incongruous lot. For example, after a meticulous literature search and coding process, the National

Standards Project (NAC, 2009) on interventions for individuals with autism spectrum disorders (ASD) classified 11 practices as evidence based:

1. Antecedent package
2. Behavioral package
3. Comprehensive behavioral treatment for young children
4. Joint attention intervention
5. Modeling
6. Naturalistic teaching strategies
7. Peer training package
8. Pivotal response treatment
9. Schedules
10. Self-management
11. Story-based intervention package (NAC, p. 43)

Two items in this list, modeling and schedules, are specific cuing procedures that have a wide range of applications for individuals with and without ASD. Most other items are combinations of procedures that vary from study to study. However, three items refer to more standardized packages. Comprehensive behavioral treatment originated in the work of Lovaas (1987) on early intensive behavioral intervention (EIBI) for toddlers and preschoolers with ASD. It has evolved into many different models (Handleman & Harris, 2001), all of which combine a multiplicity of intervention procedures that are delivered 20 to 40 hr per week for 2 to 3 years (Smith, 2011). Pivotal response treatment (Koegel & Koegel, 2006) is one approach to using naturalistic teaching strategies, with priority given to behaviors identified by the developers as especially important for promoting children's development. Story-based intervention is derived from a trademarked program, Carol Gray's Social Stories (Gray, 2004). As acknowledged in an incisive commentary, the only common element of all 11 of these items is that each is comprised of intervention procedures or combinations that seem to share core characteristics (NAC, 2009).

What is evidence based? Going beyond our longstanding contention that SSED studies have advantages over between-groups studies (Sidman, 1960), behavior analysts assert that such studies can stand alone as the sole source of empirical support for a practice (Horner et al., 2005). Is this more expansive assertion tenable? Going still further, some writers argue that many ABA intervention approaches (e.g., differential reinforcement and shaping) are so well established and are monitored so closely for each individual who receives them that they can be considered evidence based even without reference to a particular set of SSED studies (Keenan & Dillenburger, 2011). Does this line of reasoning withstand scrutiny? Also, many systems have been proposed for classifying practices as evidence based on account of findings from either SSED or between-groups studies (Chambless & Hollon, 1998; NAC, 2009; National Secondary Transition Technical Assistance Center, 2010; Odom et al., 2005; Reichow, Volkmar, & Cicchetti, 2008). How can two such disparate methodologies, often characterized as separate research traditions by behavior analysts (e.g., Johnston, 1988), both be used to identify evidence-based practices?

These questions are not merely about semantics. The purpose of identifying evidence-based practices is to help consumers and providers choose among intervention approaches (Detrich, 2008). Essentially, a practice is a service offered by providers to consumers, and representing a service as evidence based is a form of endorsement or recommendation, backed by an appeal to science. Insurers and other third-party payers increasingly use information about evidence-based practices to determine which services to cover (Green, 2008).

The pliable definitions and criteria for evidence-based ABA practices now in use are troublesome because they put us at risk for overestimating our accomplishments and helpfulness

to consumers and providers while we overlook topics in need of more study. The goal of the present article is to help behavior analysts provide useful guidance to consumers, providers, and third-party payers by clarifying what is or is not a practice and what is or is not suitable evidence. Accordingly, the article examines what constitutes a practice (a procedure or package of procedures, standardized or not) and persuasive evidence for the practice (findings from SSED or group studies). In so doing, prior discussions of these issues by behavior analysts are reconsidered in light of the much larger literature currently available, and ABA interventions that behavior analysts have identified as evidence-based practices are critiqued. These considerations are relevant for ABA providers, who need to recognize which services they offer have adequate scientific support and which do not. Moreover, they raise critical questions to be addressed by ABA researchers, who must determine what is known and unknown about interventions and set priorities for future programs of inquiry.

PROCEDURES OR PACKAGES?

In their formative description of ABA as a discipline, Baer, Wolf, and Risley (1968) emphasized that ABA is technological, meaning that its intervention techniques or procedures are precise, detailed, and replicable. A key aspect of ABA research ever since has been to discover such procedures and understand how and why they work. However, Baer (1975, 2004) pointed out that, as research progresses, it may become appropriate to combine procedures into a package or program and test the combination. A single intervention procedure (i.e., a format for teaching a new skill or tactic to decrease a behavior) is seldom enough to fix a problem by itself; as Baer (2004) remarked, "People's problems seem

to require a program" (p. 310). Azrin (1977) added that a package or program should contain strategies for deploying techniques effectively, efficiently, and credibly, even when confronted with real-world constraints such as resource limitations, variations in skill level and enthusiasm of interventionists, and competition from alternative programs. In other words, a package should help providers make decisions on how to use procedures in their practice.

The need for intervention packages led Azrin (1977) to recommend moving quickly to conducting research on packages. However, Birnbrauer (1979) countered that it is premature to evaluate packages unless their components have been thoroughly investigated. Johnston (1996) added that evaluations of components should go through a sequence beginning with investigations in controlled settings that allow rigorous experimental analyses and proceeding to field settings that might require compromises (e.g., reliance on inexperienced interventionists or introduction of an intervention after a baseline that was too short to establish a stable rate of behavior).

Although they warned against forming packages comprised of incompletely understood procedures, both Birnbrauer (1979) and Johnston (1996) agreed that movement from procedures to packages is essential. Birnbrauer averred that "no one can argue with" (p. 19) the value of packages that improve outcome, minimize costs, and are acceptable to consumers. He added that SSED studies on an individual procedure are "at best, slower" and "more difficult to sell to consumers" (Birnbrauer, p. 19) than research on packages. Likewise, Johnston (1993, 1996) pointed out that the goal of ABA research is "fully identifying all of the individuals and behaviors of interest" and "understanding the web of social contingencies" that influence these individuals and behaviors

(Johnston, 1993, p. 329). He further observed,

The challenge is to design the bits and pieces of a procedure into an integrated and often complex whole. ... At some point, however, everything that has been learned should suggest a reasonable combination of procedural elements as a focus for further analytic attention. (p. 330)

In short, analysis of procedures yields the information needed to form a coherent package that can itself be studied.

Nevertheless, controversies over efforts to create intervention packages show that behavior analysts continue to struggle with how and when to do so. For example, Carr et al. (2002) reproached behavior analysts for overrating studies of isolated procedures:

For decades, applied behavior analysts have prided themselves on the publication of many successful research demonstrations that involve the application of single interventions. These demonstrations have made for great science but ineffective practice. A comprehensive approach involving multicomponent intervention is necessary to change the many facets of an individual's living context that are problematic. (p. 9)

In a rejoinder, Johnston, Foxx, Jacobson, Green, and Mulick (2006) disputed the charge that ABA procedures look impressive in a research context yet fail in practice. They also criticized many aspects of the framework advocated by Carr et al. for creating packages to address problem behavior (positive behavior intervention and support; PBIS). It is important to note, however, that they too described effective intervention for problem behavior as involving a combination of procedures and indeed worried that advocates of PBIS have downplayed some key procedures, especially consequence-based strategies such as differential reinforcement. Moreover, they admitted that behavior analysts have much to learn about how to package interventions that nonspecialists would be

willing and able to adopt. Even more strikingly, they acknowledged a disconnect between what appears in an outlet such as the *Journal of Applied Behavior Analysis (JABA)* and what happens in practice: "Although some of the features of *JABA* research may certainly be found in routine practice, the pages of that journal do not represent the full range of practices in ABA" (Johnston et al., 2006, p. 61).

The exchanges between Azrin (1977) and Birnbrauer (1979) and between Carr et al. (2002) and Johnston et al. (2006), along with the comments by Baer (1975, 2004), suggest that a broad consensus has emerged on what constitutes a practice. Formulating and evaluating practices, however, remain contentious. Thus, it is worth exploring the consensus on what constitutes a practice and then to consider research approaches that would increase success in developing practices.

There is general accord among behavior analysts that a practice is a service aimed at solving people's problems. Further, although no one has ruled out the possibility that the solution might consist of a single procedure in some cases, behavior analysts concur that the solution is more likely to involve a combination of procedures, along with strategies for deploying the practice in typical service settings (i.e., a package). For example, a practice aimed at reducing a problem behavior is likely to include a functional analysis or assessment, followed by the use of function-based interventions that involve altering both antecedents and consequences of the behavior while strengthening replacement behaviors (Johnston et al., 2006). A practice aimed at overcoming skill deficits is likely to include a careful evaluation of the individual's current skill level and the skills that would be most helpful for the individual to learn, as well as a task analysis of those skills, followed by the implementation of a

range of teaching strategies to help the individual acquire, generalize, and maintain the skills (Greer, 2002). Either kind of practice probably also involves considering contextual factors such as the individual's family environment, peer networks, and expectations in school or at the workplace. In addition, the practice requires determining what is feasible given constraints such as resource limitations and preferences of the people involved.

These views are consistent with standard definitions of a practice across professions. Particularly when used in the phrase *evidence-based practice*, a practice refers to a fusion of research findings, consumer preferences, and clinical assessment (American Psychological Association [APA], 2005; Sackett, Rosenberg, Muir Gray, Haynes, & Richardson, 1996). This meaning is encapsulated in one helpful synopsis:

There is an abundance of definitions of evidence-based practice (EBP). Fortunately, most of them say essentially the same thing. ... The goal of EBP is the integration of (a) clinical expertise/expert opinion, (b) external scientific evidence, and (c) client/patient/caregiver values to provide high-quality services. (American Speech-Language Hearing Association, n.d.)

Thus, evidence-based practice in psychology is defined as “the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences” (APA, 2005, p. 5). Similarly, evidence-based medicine is defined as an integration of clinical expertise and research evidence for “the conscientious, explicit and judicious use of current best evidence” (Sackett et al., 1996, p. 71) to guide decisions about service delivery. In these definitions a practice is not an isolated procedure. Rather, it is a procedure or (more commonly) a set of procedures that a skilled provider can adapt to meet the needs of an individual case in the context of a

service setting and the individual's life circumstances (i.e., a package).

Definitions of evidence-based practice by behavior analysts use different words to make virtually the same point. Detrich (2008) described the function of evidence-based practice as “consumer protection.” Detrich, Keyworth, and States (2007) stressed that the identification of evidence-based practices involves the review of research findings to guide providers and consumers in their choice of interventions and the development of strategies to help them implement the interventions even in the presence of “social, political, fiscal, and human resource” (p. 4) barriers that might exist. Similarly, Horner and Kratochwill (2011) referred to a practice as “any operationally defined set of procedures that are used by a specified target audience, under defined conditions/contexts, to achieve valued outcomes for one or more defined populations” (p. 2). In keeping with definitions in other professions, these definitions characterize a practice as a package for implementing a procedure or procedures. Horner and Kratochwill's definition is especially clear that an operationally defined procedure is only one part of a practice.

Practices can vary in the range of behaviors they target. For example, as emphasized by Carr et al. (2002), practices developed within PBIS often aim to reduce a problem behavior by improving the individual's overall quality of life. In contrast, functional communication training is a practice that focuses more specifically on replacing a problem behavior with a communication skill that serves the same function as the problem behavior (Carr & Durand, 1985). As another example, EIBI is intended to accelerate all aspects of development in young children with ASD (Lovaas, 1987), whereas the picture exchange communication system is directed mainly toward helping children with ASD begin to communicate (Bondy & Frost, 2002).

However, it is a misnomer to describe an individual procedure (e.g., modeling) as a practice, as has been done in some ABA reports (e.g., NAC, 2009), unless the procedure by itself offers a solution to a problem and providers have guidance on how to use it in practice. If research shows that a procedure reliably changes behavior, it may become a component of a practice. It may also be worthy of dissemination in its own right. Embry and Biglan (2008) described some procedures as “kernels”—fundamental units that can be incorporated into a wide range of intervention or prevention packages. Examples include “time-out, written praise notes, self-monitoring, framing relations among stimuli to affect the value of a given stimulus, and physiological strategies such as nasal breathing when upset” (p. 77). They noted that kernels tend to be simpler than complete packages and thus may be more readily disseminated in some circumstances. However, they acknowledged that users will need to combine kernels with other approaches and that kernels are not a substitute or replacement for packages.

Even under the best of circumstances, community providers and consumers would have great difficulty determining how to use procedures in practice. For example, expanding on the NAC (2009) report, Odom, Hume, Boyd, and Stabel (2012) identified 24 evidence-based teaching procedures for children with ASD such as discrete-trial training, prompt delay, and visual schedules. They also laid out step-by-step instructions for setting up and implementing each procedure, along with written tutorials and instructional videos to show how to use the procedures correctly. To supplement this information, Odom et al. offered general guidelines on the kinds of skills that can be taught with each procedure (e.g., noting that visual schedules have been found to be especially useful for promoting independent comple-

tion of self-help activities). However, they did not identify a way to select procedures that match the needs of an individual child with ASD, the scope and sequence of skills to teach, the frequency and length of time to devote to each procedure, or methods to deploy the procedures expediently in community settings such as schools. Therefore, although Odom et al.’s set of procedures is a valuable resource, it mainly consists of pieces (i.e., kernels) that need to be assembled by providers and consumers. Thus, it constitutes an important step toward formulating a practice but remains a work in progress.

Going beyond procedures, packages (e.g., intervention programs developed within PBIS) do include specific manuals for assessment of the needs of individual consumers (Sugai, Lewis-Palmer, Horner, & Todd, 2005), development of an intervention plan for an individual (Dunlap et al., 2010) or a group (Sugai et al., 2010), and evaluation of the fit between the intervention and the context in which it takes place (Horner, Salentine, & Albin, 2003). As discussed above, this kind of integration of procedures into packages is crucial for creating practices. Moreover, given the voluminous literature of SSED studies on the domain targeted in PBIS (problem behavior), one might hope that investigators have met Birnbrauer’s (1979) and Johnston’s (1996) criteria for studying procedures thoroughly before packaging them. However, it is possible that the literature may still contain gaps. As Johnston (1996) commented, many SSED studies are aimed at responding to an immediate need that arises in a service setting, rather than advancing a systematic program of research. Consequently, “the considerable size of the applied literature may suggest more than it can actually deliver” (p. 39). Accordingly, a careful review of the literature may be warranted to determine whether the research is sufficiently far along to guide the creation of a package. Another issue to consider is

that, because packages encompass a multiplicity of procedures and strategies for their deployment, they can take an almost infinite variety of forms (Baer, 2004). Thus, if existing packages (e.g., PBIS approaches) raise concerns (Johnston et al., 2006), investigators can propose alternatives.

Investigators may debate whether to conduct SSED studies to fill in gaps in the literature on problem behavior, focus on PBIS, or develop new intervention packages. Such debates also may arise in connection with interventions for other problems. What is clear, however, is that the key to resolving these debates is to identify ways to promote progress from isolated procedures to integrated packages. The establishment of evidence-based practices in ABA depends on achieving this integration.

To summarize, a practice is a solution to a problem presented by a consumer. Most problems require a combination of procedures, along with strategies for implementing them in a service setting (i.e., a package). The process of implementation necessitates the integration of evidence about the practice with the preferences of the individual who is being served and the skills of the service provider. Although SSED studies are invaluable in the identification of individual procedures to include in a practice, researchers still must find ways to put procedures together into a package for use in a practice setting. The next section describes methodologies for examining whether a procedure is likely to be useful in a practice setting and for packaging procedures.

Research Implications

The potential utility of procedures in practice. Behavior analysts often emphasize the need to study the effects of ABA procedures in the context of typical practice settings (e.g., Johnston, 1996). However, reviews indicate that the large majority

of our research focuses on interventions delivered by study personnel, usually in tightly controlled environments such as laboratories, specialized ABA classrooms, or distraction-free areas set up to provide one-to-one instruction (e.g., Kasari & Smith, in press; Rehfeldt, 2011; Trahan, Kahng, Fisher, & Hausman, 2011). This discrepancy may reflect a dilemma that behavior analysts have had trouble resolving: We recognize that conducting studies in practice settings may require sacrificing some scientific rigor because the primary mission of such settings is to deliver services rather than conduct research (Johnston, 1996), yet we regard the quality of many studies in these settings as unacceptable (Johnston et al., 2006).

The root of this dilemma is that findings from poorly designed studies are difficult or impossible to interpret, no matter where they take place. Thus, the best way to move forward is to seek opportunities to carry out rigorous studies in practice settings. Of course, this is easier said than done, but several areas of research offer guidance on how to proceed. One such area is implementation science (Fixsen, Naoom, Blase, Friedman, & Wallace, 2005), which grew out of studies on ABA programs such as Achievement Place. Related areas of investigation, notably diffusion of innovations (Rogers, 2003), socio-technical technology transfer (Backer, David, & Soucy, 1995), and improvement models (Langley et al., 2009), arose from efforts to deploy other technologies in community settings. Research in these areas has highlighted many factors that contribute to success in working with community agencies. Particularly important elements include the establishment of an ongoing partnership with the agency, identification of a champion within the agency for introducing a new technology, assembly of a team to work together to implement the technology, securing support from administrators to ensure that adequate

resources are available, obtaining approbation from other influential individuals in the agency, and starting with small pilot projects (Damschroder et al., 2009). A commendable example in ABA research is the Behavior-Analytic Consultation to Schools program, which has completed a large number of high-quality functional analyses in public schools (Mueller, Nkosi, & Hine, 2011). Other examples are the studies of EIBI by Howard, Sparkman, Cohen, Green, and Stanislaw (2005) and Cohen, Amerine-Dickens, and Smith (2006), which were made possible when one of the authors (Cohen) negotiated a four-way agreement among state agencies, local school districts, private service providers, and parents on a system for publicly funding EIBI.

Along with the need to test the generality of intervention effects across service settings and providers, it is also important to test the generality of effects across individuals who receive interventions and to identify factors that predict success or failure of an intervention for a given individual. Despite the caveats noted in the introduction, meta-analyses can help fulfill this objective by systematically organizing data obtained across many SSED studies. (Because meta-analyses are also used in group studies, they may have the added advantage of presenting results from SSED studies in a way that is understandable and convincing to researchers outside ABA.) However, reliable information on generality of effects is difficult to obtain from the ABA literature because failures are hardly ever reported. For example, data from all treatment articles in Volumes 40 through 44 of *JABA* indicate that every study participant responded favorably to intervention, ordinarily within a few sessions. To be sure, the results were tempered in some cases by individual variations in response (e.g., Dunn, Sigmon, Thomas, Heil, & Higgins, 2008), incomplete generalization (e.g., Shillings-

burg, Kelley, Roane, Kisamore, & Brown, 2009), or a need to augment the original intervention (e.g., Moore & Fisher, 2007). Still, a 100% rate of improvement that usually occurs right away belies the practical reality that many individuals make slow or minimal progress, even in exemplary, intensive ABA intervention programs (cf. Hagopian & Wachtel, n.d.; Leaf, Taubman, McEachin, Leaf, & Tsuji, 2011).

The tendency to publish only positive results may be due to the priority given in ABA research to the establishment of experimental control that shows a functional relation between an intervention and a behavior. Beginning with Skinner (1956), behavior analysts have held that it is more informative to present data that epitomize experimental control than to present data on a large number of participants or on an average response to the intervention. As a corollary, null findings, which reflect a lack of experimental control, may be uninformative and hence unworthy of publication.

However, behavior analysts are interested not only in establishing experimental control but also in performing replication studies to determine the generality of findings (Johnston & Pennypacker (2009). Given that ABA interventions are not universally effective, this process must include documentation of both successes and failures. Although, to this writer's knowledge, behavior analysts have not described ways to provide such documentation, we can consider approaches developed in other disciplines as a starting point. In group studies on intervention, the standard approach is to use an "intention-to-treat" analysis (Schultz, Altman, Moher, & CONSORT Group, 2010). In this kind of analysis, researchers report outcome data from all participants who entered the study, not just those who completed it. To put these data in context, they also describe the flow of participants

through the study. This description includes the recruitment procedures (e.g., in person or by telephone or mail, from one agency or more), the inclusion and exclusion criteria for entry into the study, the number of potential participants who were screened for eligibility but did not enroll, the reasons why they did not enroll (e.g., declining to give consent), and the number of dropouts, along with the timing and reasons for dropping out (Schultz et al., 2010). The goal is to disclose the full breadth of outcomes of intervention, including the possibility of not receiving the intervention at all, rather than presenting only successes.

In other social sciences, researchers who conduct studies with small samples incorporate “purposeful sampling strategies” (Creswell, 2006; Miles & Huberman, 1994). For example, researchers sometimes seek maximum variation in the target problem (e.g., problem behavior that ranges from mild to severe) or actively look for disconfirming cases. Alternatively, they may seek participants who are considered typical or representative in some way of a larger population. Such approaches could be valuable in SSED studies. As an illustration, although experimental functional analysis has been highly productive in ABA research, it is seldom used in community practice (Dixon, Vogel, & Tarbox, 2012). Researchers have attempted to bridge the gap between research and practice by streamlining functional analysis procedures (Iwata & Dozier, 2008), showing that they can be completed in real-world environments (Mueller et al., 2011), and testing their social validity (Langthorne & McGill, 2011). However, another barrier may be that, because functional analysis emerged from work in inpatient or day-treatment programs with individuals who exhibit severe problem behavior (e.g., aggression or self-injury), it is not clear whether such analysis is necessary or helpful with individuals who

display routine problem behavior such as talking out of turn, refusing to complete tasks, or breaking rules. Deliberately sampling such individuals may be an important step toward ascertaining the utility of functional analysis across a wide spectrum of problem behavior.

Overall, although many ABA intervention procedures show considerable promise, the tendency of SSED studies on these procedures to take place in controlled settings, rely on study personnel to deliver interventions, and focus on successes may exaggerate the potential benefits of the procedures in practice. As a corrective, researchers would do well to take advantage of findings from implementation science and related areas of research on how to carry out high-quality research in practice settings. In addition, it may be advisable to use intention-to-treat analyses or purposeful sampling strategies (or both).

Packaging intervention procedures. Baer, Wolf, and Risley (1987) drew attention to a formidable technological challenge in developing intervention packages: the need to determine which procedures in the package should be “followed faithfully, no matter where or when the program is used,” and which procedures “users should be allowed, and even encouraged, to modify ... to fit their local situations and contingencies” (p. 321). In the years since Baer et al. highlighted this challenge, investigators have made little progress in elucidating principles that govern how to strike an appropriate balance between standardization and individualization. The differing levels of standardization in ABA intervention packages for individuals with ASD (NAC, 2009) or other problems may reflect this uncertainty.

Nevertheless, investigators have honed a technology for proposing a balance that can be tested: an intervention manual. The advent of manuals has transformed research on

behavioral, psychological, and educational treatments (McHugh & Barlow, 2012). Manuals standardize intervention packages by giving step-by-step instructions for each stage of implementation, along with problems that the user is likely to encounter and possible solutions to these problems (Hibbs et al., 1997). They also define the problem or population for whom the intervention package is intended and the qualifications that providers who implement the manual should possess (Hibbs et al., 1997). At the same time, they allow “constrained flexibility” (MacMahon, 2004) by delineating a limited set of acceptable variations. For example, the package may be divided into modules with decision rules or assessment procedures for selecting which modules to implement and under what circumstances (Weisz et al., 2012). The manual may describe different ways for providers to deliver the package (e.g., procedures for implementing the intervention one-to-one or in groups), incorporate the consumer’s own interests into activities, and collaborate with the consumer to set goals (cf. Kendall & Chu, 2000; Kendall, Chu, Gifford, Hayes, & Nauta, 1998; Tee & Kazantzis, 2011).

Manuals may take a variety of forms. For example, a manual for outpatient therapy might lay out a sequence of 1-hr sessions, whereas a manual for an educational program might consist of a curriculum in which the learner has to master each step before going to the next, along with techniques for the teacher to use in delivering instruction. For an ABA intervention package, it might be organized around functional relations (e.g., a set of procedures for problem behavior that serves to escape or avoid situations, another set of procedures for behavior that serves to gain access to tangible items). The manual might present a complete script for a session or merely outline key points to cover

(Frankel, 2004). It may include one or more types of supporting materials such as handouts, worksheets, videos for training interventionists or for illustrating concepts for consumers, data sheets for monitoring progress, and so on. A manual may undergo several revisions to take into account findings from initial testing of its acceptability (social validity) and usability (extent to which providers implement procedures with fidelity and recipients of the intervention adhere to the protocol). As described previously, the manual may encompass one intervention procedure or many, provided that it offers a solution to a problem presented in practice and strategies that community providers can use to implement the procedures effectively and efficiently. If the manual centers on only a few procedures, it may be possible to embed it within a research article; otherwise, it is likely to require a separate publication (Chambless & Hollon, 1998).

The availability of a manual is widely viewed as a requirement for classifying an intervention as evidence based (Chambless & Hollon, 1998; Society for Prevention Research, 2004). As stated by Chambless and Hollon, “research projects for which a treatment manual was not written and followed are of limited utility in terms of assessment of treatment efficacy” (p. 11). This view is well founded. It would be virtually impossible to generate replicable findings about unstandardized packages such as an “antecedent package” or a “behavioral package” (NAC, 2009) to reduce problem behavior. There are no criteria for determining when to use one or the other package, how many elements to include, which elements should always be included in the package and which can vary across individuals, in what sequence to introduce them, what recourse is available if the initial package is insufficient, or how to monitor fidelity of implementation.

Without such criteria, researchers and providers cannot determine whether or not they have accurately replicated an unstructured package. In contrast, as previously noted, interventions with manuals, such as those developed within PBIS (e.g., Dunlap et al., 2010), do include specific methods to address these issues. Thus, manuals create the opportunity to test a well-defined intervention package across consumers and providers, as must occur to amass a body of research that would justify classifying an intervention as evidence based. Although future researchers may develop technologies other than manuals for devising replicable intervention packages, manuals are currently the only such technology (McHugh & Barlow, 2012).

In this writer's judgment, the role of manuals in establishing evidence-based practices may explain why the efficacy of schoolwide PBIS interventions has been evaluated much more broadly (i.e., in much larger studies, with independent researchers) than unstructured ABA packages for problem behavior (see Horner, Sugai, & Anderson, 2010) or even ABA packages that have been described in more depth but not presented in manuals (e.g., functional communication training; Tiger, Hanley, & Bruzek, 2008). The same pattern holds for other ABA interventions with manuals compared to unstructured packages for the same problems. The efficacy of the UCLA–Lovaas model of EIBI has been replicated much more broadly than any other comprehensive ABA intervention for children with ASD (Smith, 2011), pivotal response treatment more broadly than other approaches to incidental teaching (Koegel, Koegel, Vernon, & Brookman-Frazer, 2010), the picture exchange communication system more broadly than other forms of augmentative or alternative communication (Flippin, Reszka, & Watson, 2010), integrative behavioral marital therapy more broadly than unstandardized forms

of acceptance and commitment therapy (Christensen et al., 2004), the community reinforcement approach plus vouchers more often than other contingency management systems for substance abuse (Higgins et al., 2003), and the comprehensive behavioral intervention for tics more broadly than other packages that involve habit reversal (Piacentini et al., 2011).

The requirement for replicability by independent providers (often necessitating the availability of a manual) sharply reduces the number of ABA approaches that can be regarded as evidence based. Of the 11 interventions for ASD identified in the NAC (2009) report, only the three that have been standardized in manuals might be considered to be practices, and even these may be incomplete. For example, a manual is available for writing stories to include in a story-based intervention package (Gray, 2004), but not for how to use the stories to change the behavior of an individual with ASD. Thus, there are many ABA procedures and unstructured packages, but relatively few practices.

Nevertheless, some possible drawbacks of manuals deserve mention. A potential weakness of practices that involve intervention packages with manuals is that the step-by-step instructions are liable to give the impression that unskilled providers can implement the package (Johnston et al., 2006). In so doing, they may tempt agencies to employ such providers, thus diminishing the effectiveness of the package. It is certainly true that, when demand has grown for ABA interventions such as EIBI, unskilled providers have proliferated (Green, 1999). Thus, it may be necessary to anticipate this unintended consequence by specifying minimum qualifications for providers and disseminating information to consumers about these qualifications (e.g., Autism Special Interest Group of the Association for Behavior Analysis, 2007; Shook, Johnston, &

Mellichamp, 2004). A manual is an aid to clinical decision making, not a substitute. With its balance of standardization and flexibility, it structures the package yet gives providers latitude to tailor it to individual cases.

Another concern is that manuals may stifle expert providers (Johnson et al., 2006). Analogous concerns have been voiced by professionals in other disciplines (see Addis & Krasnow, 2000). Research in these disciplines provides scant evidence that expert providers achieve better outcomes if left to their own devices than if given a manual (e.g., Durlak & DuPre, 2008; Woolf & Johnson, 2005). However, the large and complex literature of SSED studies on many ABA procedures raises the possibility that there really could be an advantage to letting expert providers rely on the literature and their experience rather than a manual. Indeed some writers have portrayed “behavioral artists” (Foxy, 1996) who have an uncanny knack for creating and administering ABA interventions. It may be worth evaluating whether these providers are more effective without a manual. If so, consistent with the technological focus of ABA, the next step would be to conduct research that seeks to encapsulate what these providers do so that others can replicate it. Thus, this line of inquiry would lead toward improving manuals rather than discarding them.

An additional concern is that, because standardized packages are more detailed than unstructured packages, they may be more difficult to disseminate in community settings. Unstructured packages can be (and frequently are) promulgated widely in the form of practice guidelines or recommendations. For example, relying largely on ABA research, the American Academy of Pediatrics (Committee on Psychosocial Aspects of Child and Family Health, 1998) developed guidelines for pediatricians

to counsel parents on how to discipline their typically developing children. Parents are encouraged to use positive reinforcement frequently for appropriate behavior and to discourage acting-out behavior with procedures such as time-out rather than verbal reprimands or corporal punishment. However, in keeping with the previous discussion about the difficulty of replicating unstandardized packages, little is known about how best to promote accurate implementation of such guidelines (Grimshaw et al., 2004).

In sum, interventions that are bundled into a package and detailed in a manual are not the be-all-and-end-all of treatment research. Individual procedures in the form of behavioral kernels and unstructured packages in the form of guidelines also have utility. Nevertheless, interventions with manuals are central to evidence-based practice because they provide the integration necessary to address a problem thoroughly and the standardization needed to replicate findings across studies.

SSEDS OR GROUP DESIGNS

Behavior analysts concede that group designs have some valid uses, particularly for addressing actuarial questions such as whether one intervention tends to be more effective than another (Johnston, 1988; Skinner, 1968). Nevertheless, we have repeatedly asserted that SSED studies are sufficient to establish an intervention as evidence based. To highlight the idiographic tradition in ABA (i.e., the focus on evaluation of interventions for each individual rather than for a group), Keenan and Dillenburger (2011) described ABA as evidence based because interventions are evaluated carefully and continuously for each individual who receives them, and the standard intervention procedures are part of a general-purpose technology distilled from more than a century of research

in laboratory and field settings with a variety of human and nonhuman populations. Examples of such procedures include differential reinforcement, prompting and prompt fading, shaping, and chaining.

To support their position, Keenan and Dillenburger (2011) cited Reese, Howard, and Reese's (1978) general procedure for applied behavior analysis, which includes strategies for identifying target behaviors, conducting a task analysis, selecting among general-purpose ABA interventions, and testing outcomes. Comparable systems have been proposed by others. For example, Kameenui and Simmons (1990, p. 87) proposed "five generic design-of-instruction features" for academic tasks. These features involve the use of ABA principles to assess what children need to learn, to select curriculum materials, to plan lessons, to generalize skills to new contexts, and to pinpoint successful and unsuccessful aspects of the lessons. Reese et al.'s and Kameenui and Simmons's systems present their frameworks for implementation of ABA technology without reference to any specific studies to support their deployment for a particular problem or with a particular population of individuals.

Keenan and Dillenburger (2011) used the example of shaping to illustrate why a procedure can be considered evidence based without the need for prior RCTs or other studies to document its effectiveness. They note that, when implemented for an individual learner, shaping is not based on "fixed treatment recipes" (p. 8) copied from previous research. Rather, it is a dynamic process that is individualized for the learner and that evolves continuously based on data obtained on the individual's progress. These data constitute the evidence on the efficacy of shaping for that individual.

Certainly, it is necessary to develop an individualized intervention plan for each recipient of services, to

monitor progress regularly, and to adjust the plan as new information becomes available. In the process, it makes sense to start with tried-and-true procedures such as shaping. "But the dangers of drifting too far toward the idiographic are that we end up reinventing the wheel every time a new client comes into the office or clinic" (Barlow, 1997, p. 448). As shown in the preceding section, reinventing the wheel is a formidable undertaking. Thus, investigators have recommended ways to standardize even dynamic processes such as shaping (Galbicka, 1994). Moreover, documenting that the wheel is working requires more than just showing an uptick on a graph of an individual's behavior. A behavior may change but remain a problem for the individual, or may be only a small component of a much larger cluster of problems such as addiction or delinquency (Baer et al., 1987). Also, a procedure may be appropriate for an individual yet can be administered in the context of a program that may not be inappropriate.

Such complications abound in ABA intervention, as illustrated by the following examples:

1. Many ABA interventions have been directed toward increasing the rate at which preschoolers answer questions appropriately (Ingvarsson, Tiger, Hanley, & Stephenson, 2007) or the rate at which children with disabilities initiate or respond to overtures from peers (Odom et al., 1999). However, such gains may not be enough to overcome the children's initial problems in these areas. Furthermore, to be liked and make friends, children are likely to need many other skills (e.g., skills for sustaining interactions around shared interests and for resolving conflicts).

2. Most ABA interventions for teaching children with disabilities to read focus on sight reading, but literacy also entails other critical reading skills such as decoding, phonemic

awareness, and comprehension (Browder, Wakeman, Spooner, Ahlgrim-Dezell, & Algozzine, 2006).

3. Some ABA studies have shown success in teaching children to handle guns safely, but, as investigators have emphasized, it is at least as important to help families store firearms securely (Miltenberger et al., 2005).

4. Most studies of ABA interventions for feeding difficulties document an increase in acceptance of novel foods. However, some populations that are prone to feeding difficulties, such as children with ASD, are seldom malnourished or underweight (Hyman et al., 2012). Further, investigators seldom show that a particular child with ASD in a study on ABA feeding interventions has health problems (but see McCartney, Anderson, & English, 2005, for a notable exception). Thus, for many of these children, outcomes other than acceptance of novel foods are probably more important. Such outcomes might include reduced stress at mealtimes for children and caregivers or a more varied overall diet that promotes long-term health.

5. The goal of EIBI is to normalize the functioning of children with ASD by helping them catch up to peers (Lovaas, 1987). However, despite the acquisition of many new skills, some children do not reach this goal (Smith, 2011). For these children, providers and families need to go beyond the data on acquisition of individual skills to determine whether adjusting the intensity, content or method of instruction, or goals would make the program a better fit (Delmolino & Harris, 2011).

In principle, providers could attempt to address these issues by conducting exceptionally rigorous, thorough evaluations of an individual's progress, with SSED studies of all behaviors of concern and data on social validity (the extent to which others see the intervention as acceptable and useful). In reality, however,

the exigencies of field settings seldom allow providers the luxury of spending so much time on data collection and analysis (Johnston, 1993). Thus, some behavior analysts emphasize the need for a sequence of studies on an intervention, in addition to case-by-case evaluation. Johnston (1993, 1996) recommended beginning the sequence with studies in relatively controlled settings and then moving to more typical field settings. The sequence consists entirely of SSED studies instead of actuarial group studies.

To illustrate this sequence, Johnston (1996) cited a series of studies that led to the creation of an intervention for teaching women to perform breast self-examinations. The first studies elucidated the characteristics of lumps in breast tissue. Later studies refined procedures for teaching women to identify these lumps in laboratory settings. Finally, studies tested the procedures under conditions that increasingly resembled real-world situations (Pennypacker & Iwata, 1990). Johnston described the teaching procedure as "the state of the art in breast self-examination" (p. 41). Pennypacker (2010) lamented that the Cochrane Collaboration, which conducts influential reviews of the medical literature to identify evidence-based practices, concluded that breast self-examination failed to qualify as evidence based. Indeed, SSED studies on ABA procedures for breast self-examination provide impressive analyses of what to teach and how to teach it (see Pennypacker & Iwata, 1990).

However, the status of breast self-examination as an evidence-based intervention hinges on a quintessential actuarial question: Does it prevent deaths from breast cancer? The Cochrane review examined RCTs that addressed this question and found that the answer was no, perhaps because learning a discrete procedure for self-examination fails to address barriers to performing the

examinations consistently or accurately (Köster & Göttsche, 2008).

It is possible that outcomes would improve with greater reliance on the ABA teaching approach cited by Pennypacker and Iwata (1990). In any case, further SSED studies could help find ways to increase adherence to or accuracy of self-examination procedures. However, the only way to determine whether doing so prevents deaths from breast cancer is to study large groups of women. Correlational studies can examine whether, all other things being equal, women who perform self-examinations have lower mortality rates than other women. RCTs can compare groups that are taught to conduct self-examinations against groups that do not receive this instruction. In contrast, single-case studies, which involve a small number of participants, all of whom receive the intervention (instruction on how to conduct self-examinations), cannot show whether or not the intervention reduces mortality.

As another example that may be more pertinent to the customary scope of practice for behavior analysts, Baer (1975) drew on the history of a program for juvenile delinquents (Achievement Place). He suggested that research might begin with SSED studies that test the effects of a single procedure. For instance, the studies might use a multiple baseline design across participants. In this approach, all participants start in a baseline (no intervention) condition, and then the intervention procedure is introduced for one participant at a time. The next stage of research might involve testing a package of procedures (as opposed to an individual procedure) in a multiple baseline design across participants, and a subsequent stage might be to evaluate the package in a multiple baseline design across groups of participants, with the package being introduced to one group at a time. Although actual research programs do not follow such a neat progression, SSED studies that

involve large groups of participants do exist (Biglan, Ary, & Wagenaar, 2000). However, such studies are equivalent to commonly used group designs. As recognized by Bailey and Burch (2002), Baer's example corresponds to a group design, referred to as a wait-list control, wherein one group starts intervention at the outset of the study and is compared to another group that is placed on a waiting list to start intervention when the first group finishes. Biglan et al. (2000) demonstrated that their example could be described as a group design called an interrupted time series study (Shadish, Cook, & Campbell, 2002).

In these examples, SSED studies are pivotal in the development and refinement of interventions, but group studies are required to establish the intervention package as evidence based (i.e., to show that the intervention helps to solve the target problem). Thus, SSED studies may serve as "demonstration projects" (Kazdin, 2005) or "proof of concept" (Smith et al., 2007), providing a foundation for larger studies but not standing alone as the source for the establishment of an intervention as evidence based. Stated differently, the process of development and validation of an intervention may require multiple stages, as in Baer's (1975) example, with SSED studies having a central role in earlier stages and group studies becoming increasingly important in later stages (cf. Forness, 2005; Robey, 2004; Schlosser, 2009).

The reason that this progression is likely to be necessary is that most problems pose important actuarial questions (e.g., determining whether an intervention package is more effective than community treatment as usual; deciding whether to invest in one intervention package or another, both, or neither; and determining whether the long-term benefits justify the resources devoted to the intervention). Hence, group

studies, which are designed to address such questions (Johnston, 1988), are needed. A particularly important actuarial issue centers on the identification of the conditions under which the intervention is most likely to be effective. Although SSED studies can explore the effects of systematically altering experimental conditions (Sidman, 1960), group studies can better test the generality of these effects through formal analyses of moderators and mediators (variables that influence the effects of the intervention; Kasari & Smith, in press).

This perspective may help to resolve an apparent contradiction that has emerged from efforts to identify evidence-based practices in education and psychotherapy. Even though many published criteria in psychology (Chambless & Hollon, 1998), education (Odom et al., 2005), and other professions (e.g., Romeiser Logan, Hickman, Harris, & Heriza, 2008) count both single-case and group studies as acceptable sources of evidence, group studies are much more commonly cited in support of classifying an intervention as evidence based (e.g., Nathan & Gorman, 2002). For example, the What Works Clearinghouse, which evaluates research on educational practices, has not identified any practices as evidence based from SSED studies as of this writing, although it may do so in the future, in that it has recently broadened its review criteria to incorporate such studies (Kratochwill et al., 2010). Reviews of rehabilitation interventions for neurological disorders mention SSED studies as a useful research design but do not specify evidence-based practices that have emerged from such studies (e.g., Beeson & Robey, 2006; Slifer & Amari, 2009). The initial report on evidence-based practices by the Clinical Psychology Division of the American Psychological Association (Chambless et al., 1996) listed three evidence-based interventions with

support from SSED studies: behavior modification for encopresis, token economies for individuals with schizophrenia in inpatient settings, and behavior modification for developmental disabilities. However, the last two items were deleted from the next report because they were judged to be insufficiently specific (Chambless & Hollon, 1998), leaving only behavior modification for encopresis. Although encopresis can be a source of substantial distress and functional impairment, it is a circumscribed difficulty that can usually be overcome in a few weeks (O'Brien, Ross, & Christopherse, 1986), and no known alternatives to ABA interventions have been tested other than medical treatments. As this example suggests, SSED studies can be the sole source of evidence for a brief intervention to address a narrow problem, but such interventions are the exception, not the rule.

In contrast to criteria in psychology and education, criteria for evidence-based medicine usually give greater weight to RCTs than to any other design (e.g., Guyatt, Oxman, et al., 2008). Interestingly, however, medical researchers regard some SSED approaches as types of RCTs, which they call "*N*-of-1 randomized trials" (Guyatt et al., 1986). For example, a reversal design involves observation of a participant's behavior in baseline, then the introduction of an intervention, then withdrawal of the intervention to see whether the behavior reverts to its preintervention level, and then reintroduction of the intervention. In this design, participants' behaviors, rather than the participants themselves, are assigned to the baseline and intervention conditions (Guyatt et al., 1986; see also Kratochwill & Levin, 2010). Medical researchers explicitly borrowed the reversal design from ABA research (Guyatt et al., 1986) and describe it as being at least as methodologically sound as a group RCT (Guyatt, Rennie, Meade, & Cook, 2008). Nevertheless, important

limitations have also been recognized (Evans & Ildstad, 2001). For example, it may be possible to obtain a clear baseline and reversal for chronic conditions such as diabetes or hypertension, but not for acute or rapidly deteriorating conditions such as injuries or infections. Moreover, it may be possible to observe clear changes with fast-acting treatments, but not with treatments that are slower to take effect (e.g., drugs that take several weeks to build up in the body to a level needed to ameliorate the condition being treated) or that are intended to improve long-term outcomes (e.g., reduction of an individual's risk of developing coronary heart disease later in life). Thus, in principle, SSED studies (*N*-of-1 randomized trials) could be the primary approach used to establish a medical intervention as evidence based, but they seldom are.

Consistent with the perspective presented here, medical investigators conduct a variety of types of studies before RCTs are undertaken. For example, new drugs are initially developed in the laboratory and then tested in humans in three or four phases. Phase 1 studies evaluate safety and appropriate dosing in small groups of participants. Phase 2 studies provide initial tests of safety and efficacy. Phase 3 studies are large RCTs, and Phase 4 studies are RCTs that test different possible uses of the drug. Investigators may go on to conduct additional RCTs to determine how best to use the drug in community settings (Szilagyi, 2009) and how to help providers prescribe it appropriately (Thornicroft, Lempp, & Tansella, 2011). Thus, although different designs are used in earlier stages of developing interventions, RCTs become indispensable in later phases.

Medical researchers also concur with behavior analysts that group RCTs often include heterogeneous groups of participants and regularly show wide variability in response to a medical treatment. They further

agree that this variability may limit the generalizability of study findings to a given individual (Larson, 1990). Therefore, following group RCTs, SSED studies may be recommended for testing the efficacy of the intervention in different individuals (Larson, 1990). Barlow and Nock (2009) likewise advocated such studies after group RCTs of behavioral interventions. Thus, SSED studies may be important not only for innovation (as noted above) but also for later stages of research on individualizing interventions. Still, group RCTs usually need to occur in between.

RCTs are not the only design needed to establish an intervention as evidence based, but neither are SSEDs. The process of development and validation involves multiple stages and is likely to require multiple kinds of studies. RCTs tend to be better suited to the later stages of this process than are SSEDs.

DISCUSSION

Identification of evidence-based practices entails more than the analysis of behavior. It requires synthesis of findings into a package that independent providers can adopt and that offers a thorough solution to problems presented by consumers. Although SSED studies are vital for analysis, other approaches are necessary for synthesis, especially the creation of a manual that combines procedures into a standardized yet flexible format and carrying out group studies such as RCTs to evaluate effects that are broader than a change in a single target behavior. Behavior analysts have achieved such a synthesis for diverse problems that range from ASD (Mayville & Mulick, 2011), substance abuse (Higgins et al., 2003), Tourette syndrome (Piacentini et al., 2011), and marital discord (Christensen et al., 2004). However, at times, we have become so mired in analysis that we endorse interventions that are too fragmen-

tary, protean, or incompletely tested to be of much use to providers or consumers.

Fortunately, recognition of the need for synthesis immediately opens up a plethora of exciting opportunities for behavior analysts (Lutzker & Whitaker, 2005). Many problems of interest to us have attracted a large number of SSED studies on procedures that appear to be ready to compile into packages for larger scale testing. The following are some examples, and readers are likely to identify others:

Intellectual and developmental disabilities (IDD). ABA intervention for IDD is a prominent area of SSED research. However, parent training for the promotion of adaptive behavior and reduction of problem behavior (Baker, Ambrose, & Anderson, 1989) is the only ABA intervention package that has been developed specifically for such individuals, standardized in a manual, and tested in group studies (cf. Handen & Gilchrist, 2006). Thus, it is unknown whether ABA interventions might help individuals with IDD catch up to peers or improve their quality of life.

Parenting. Because of the obvious importance of parenting, Skinner (1945) and subsequent behavior analysts have eagerly sought to use ABA principles to help parents raise their children in ways that are mutually fulfilling and stress free. Group studies show that parent training derived from ABA principles can help parents to reduce their children's severely oppositional, aggressive, or delinquent behaviors (Dretzke et al., 2009) and can lower the risk that some parents will abuse or neglect their children (Edwards & Lutzker, 2009). However, ABA interventions with manuals have not been developed and tested for counseling parents in primary care pediatric settings about milder problem behaviors displayed by typically developing children. This is a significant gap, because such behaviors are among the most com-

mon concerns that parents express (Friman, 2010). Also, although many SSED studies have centered on ABA interventions for other common concerns such as difficulties with feeding (Sharp, Jacquess, Morton, & Herzinger, 2010), this research has not progressed to group studies to evaluate the impact of intervention packages on parents' stress or on children's success in learning to feed independently and appropriately.

Health-related behaviors. Obesity is among the most urgent public health concerns in developed countries in the 21st century, and ABA procedures to promote physical exercise and healthy eating are widely viewed as especially promising intervention strategies (Freedman, 2011). SSED studies are available on such procedures (e.g., De Luca & Holborn, 1992), and more such studies are underway (Freedman, 2011). There have also been at least 40 SSED studies on ABA procedures to improve performance in various sports (Martin, Thompson, & Regehr, 2004). However, group studies on the extent to which such procedures actually reduce obesity, promote fitness, and enhance health are only just beginning (e.g., Stark et al., 2011).

Education. Low academic achievement has been a perennial concern among researchers, including behavior analysts beginning with Skinner (1954). Many SSED studies have been conducted on ABA interventions for education, and some interventions have been packaged into curricula such as Direct Instruction (Engelmann, 1980), Headsprout (Twyman, Layng, Stikeleather, & Hobbins, 2004), and EdMark (EdMark Reading Program, 2004). However, Direct Instruction is the only one to have been evaluated in controlled group studies. In addition, a federally sponsored review indicated that the quality of all but one of these studies was low and that the overall level of evidence was small (What Works Clearinghouse, 2007). Thus, the impact that

ABA curricula might have on academic achievement remains uncertain.

Besides the myriad possibilities for scaling up ABA interventions, behavior analysts are well positioned to address some critical overarching issues. As previously noted, one such issue is the relative merit of disseminating intervention packages with manuals versus unstructured packages (guidelines) or individual procedures (kernels). Another is whether expert providers are helped or hindered by manuals. A third is whether the limitations of SSED studies in medicine (i.e., difficulty using such studies in research on unstable conditions or slow-acting treatments) are also relevant to research on behavioral or educational problems.

Other issues pertain to the manuals themselves. Most manuals that have been used in research are comprised of approximately 10 sessions in which a provider works one-to-one with an individual or his or her parent (Kazdin, 1997). Despite the successes of these manuals, they have important limitations. Interventions that adhere to the manuals often reduce consumers' target problems but are seldom a cure (Kendall, 2012). Thus, the search for more effective manuals must continue. Also, the manuals have greatly facilitated dissemination of evidence-based practices to providers, but even now only a small percentage of individuals who might benefit from such practices have access to them (Kazdin 2011). For this reason, ways to disseminate evidence-based practices more widely are still needed (Embry & Biglan, 2008; Kazdin, 2011). As experts in operationalizing procedures, behavior analysts may be uniquely qualified to develop innovative ways to package intervention procedures. Indeed, although the format of some manuals for ABA interventions follows the customary 10-session model (e.g., Piacentini et al., 2011), others are strikingly different. For example, behavior analysts have developed manu-

als for intervention programs such as EIBI that involve many more than 10 sessions (Lovaas, 1987) and for parent training programs that adjust the amount of service depending on need (Edwards & Lutzker, 2009).

An additional topic of interest is whether the progression from SSED studies to manuals to group studies (as outlined in this article) offers true advantages. Although medical researchers have outlined sequences of steps to develop and validate treatments such as new medications, nonmedical researchers have had little to say about this process (Mace, 1991). For example, some writers describe how to create a manual but do not discuss what kind of evidence, if any, is needed as a prerequisite to this activity (Hibbs et al., 1997; Kendall et al., 1998). Some indirect evidence suggests that starting with SSED studies can improve outcomes. For example, Seligman (1993) observed that psychotherapies for anxiety disorders appear to be more successful than psychotherapies for depression, and he hypothesized that anxiety may be intrinsically more malleable than depression. An alternative possibility, however, is that interventions for anxiety are more effective because they are grounded in SSED studies (see Barlow, 2002), whereas interventions for depression are not (Beck, 1987). Similarly, Guralnick (1998) reviewed studies indicating that early intervention appears to produce larger gains in intellectual functioning for children with ASD than for children with other disabilities. Notably, early intervention programs for ASD emerged from SSED studies (Mayville & Mulick, 2011), whereas programs for children with other disabilities did not (Guralnick, 1997). More direct comparisons of interventions derived from SSED studies and those derived from other sources would be informative.

Despite the potential benefits of moving from SSED studies to manuals and group studies, some cau-

tions must be acknowledged. First, there is no guarantee that this process will come to fruition. For instance, investigators reported substantial success in SSED studies on token economies to improve the behavior of individuals with schizophrenia (Paul & Lentz, 1977) and created a manual for this approach (Paul & Menditto, 1992). However, the intervention was seldom tested in larger scale studies (McMonagle & Sultana, 2009) or adopted in practice (Paul & Menditto, 1992). Commentators debate whether the intervention languished because investigators and providers hold distorted views about it (Wong, 2006) or because the data are less favorable than proponents claim (Wakefield, 2006). Regardless, the implication is that research and dissemination may stall even for interventions that appear to be promising.

An additional caution is that group studies demand far more resources than most SSED studies. Because SSED studies involve few participants, they can be performed by independent providers and small teams (Hayes, Barlow, & Nelson-Gray, 1999). In contrast, group studies may be feasible only in large research centers. As an illustration, an RCT of the breast self-examination procedure discussed earlier would probably need to involve experts in the procedure, oncologists, statisticians, and data managers across multiple sites with access to a large recruitment base. Collaboration, then, will be essential (Johnston, 2000).

In short, an evidence-based practice is a service that helps to solve a consumer's problem. Thus, it is likely to be an integrated package of procedures, operationalized in a manual, and validated in studies of socially meaningful outcomes, usually with group designs. This definition is more restrictive than what many behavior analysts have proposed, but it is in keeping with behavior analysts' view that ABA is a technology. Although Baer et al. (1968) and subsequent

behavior analysts have rightly emphasized that, as a technology, ABA focuses on procedures that are well defined and thoroughly analyzed, they have also pointed out that the goal of technology is to make useful products (Baer, 1991; Johnston, 2000). Achieving this goal requires a complex, often challenging process of breaking down the product into its component parts and then assembling the parts into a consumer-friendly package—analysis and synthesis. Behavior analysts can increase our success in making one type of product, evidence-based practices, by attending to the second stage of this process as we have to the first.

REFERENCES

- Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes towards psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology, 68*, 331–339.
- American Psychological Association. (2005). *Policy statement on evidence-based practice in psychology*. Retrieved from <http://www.apa.org/practice/resources/evidence/evidence-based-statement.pdf>
- American Speech-Language Hearing Association. (n.d.). *Introduction to evidence-based practice: What it is (and what it isn't)*. Retrieved from <http://www.asha.org/Members/ebp/introl>
- Autism Special Interest Group of the Association for Behavior Analysis. (2007). *Consumer guidelines for identifying, selecting, and evaluating behavior analysts working with individuals with autism spectrum disorders* (2nd rev.). Retrieved from http://www.abainternational.org/Special_Interests/Aut_Guidelines.pdf
- Azrin, N. H. (1977). A strategy for applied research: Learning based but outcome oriented. *American Psychologist, 32*, 140–149.
- Backer, T. E., David, S. L., & Soucy, G. (Eds.). (1995). *Reviewing the behavioral science knowledge base on technology transfer* (NIDA Research Monograph 155, NIH Publication No. 95-4035). Rockville, MD: National Institute on Drug Abuse.
- Baer, D. M. (1975). In the beginning, there was the response. In E. Ramp & G. Semb (Eds.), *Behavior analysis: Areas of research application* (pp. 16–30). Englewood Cliffs, NJ: Prentice Hall.
- Baer, D. M. (1991). Tacting “to a fault.” *Journal of Applied Behavior Analysis, 24*, 429–431.

- Baer, D. M. (2004). Program evaluation: Arduous, impossible, and political. In H. E. Briggs & T. L. Rzepnicki (Eds.), *Using evidence in social work practice: Behavioral perspectives* (pp. 310–322). Chicago, IL: Lyceum Books.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1*, 91–97.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 20*, 313–327.
- Bailey, J. S., & Burch, M. R. (2002). *Research methods in applied behavior analysis*. Thousand Oaks, CA: Sage.
- Baker, B. L., Ambrose, S. A., & Anderson, S. R. (1989). Parent training and developmental disabilities. *Monographs of the American Association on Mental Retardation, 13*.
- Barlow, D. H. (1997). It's yet another empirical question: Commentary on "Can Contextualism Help?" *Behavior Therapy, 28*, 445–448.
- Barlow, D. H. (2002). True alarms, false alarms, and learned (conditioned) anxiety: The origins of panic and phobia. In D. H. Barlow (Ed.), *Anxiety and its disorders: The nature and treatment of anxiety and panic* (2nd ed, pp. 219–251). New York, NY: Guilford.
- Barlow, D. H., & Nock, M. K. (2009). Why can't we be more idiographic in our research? *Perspectives on Psychological Science, 4*, 19–21.
- Baron, A., & Derenne, A. (2000). Quantitative summaries of single-subject studies: What do group comparisons tell us about individual performances? *The Behavior Analyst, 23*, 101–106.
- Beck, A. T. (1987). Cognitive therapy. In J. K. Zeig (Ed.), *The evolution of psychotherapy* (pp. 149–178). New York, NY: Brunner/Mazel.
- Beeson, P. M., & Robey, R. R. (2006). Evaluating single-subject research: Lessons learned from the aphasia literature. *Neuropsychological Review, 16*, 161–169.
- Biglan, A., Ary, D. V., & Wagenaar, A. C. (2000). The value of interrupted time-series experiments for community intervention research. *Prevention Science, 1*, 31–49.
- Birnbrauer, J. S. (1979). Applied behavior analysis, service, and the acquisition of knowledge. *The Behavior Analyst, 2*, 15–21.
- Bondy, A., & Frost, L. (2002). *A picture's worth: PECS and other visual communication strategies in autism*. Bethesda, MD: Woodbine House.
- Browder, D. M., Wakeman, S. Y., Spooner, F., Ahlgrim-Dezell, L., & Algozzine, B. (2006). Research on reading instruction for individuals with significant cognitive disabilities. *Exceptional Children, 72*, 392–408.
- Carr, E. G., Dunlap, G., Horner, R. H., Koegel, R. L., Turnbull, A. P., Sailor, W., ... Fox, L. (2002). Positive behavior support: Evolution of an applied science. *Journal of Positive Behavior Interventions, 4*, 4–16.
- Carr, E. G., & Durand, V. M. (1985). Reducing behavior problems through functional communication training. *Journal of Applied Behavior Analysis, 18*, 111–126.
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*, 7–18.
- Chambless, D. L., Sanderson, W. C., Shoham, V., Bennett Johnson, S., Pope, K. S., Crits-Christoph, P., ... Woody, S. R. (1996). An update on empirically validated therapies, I. *The Clinical Psychologist, 49*(2), 5–18.
- Christensen, A., Atkins, D. C., Berns, S., Wheeler, J., Baucom, D. H., & Simpson, L. E. (2004). Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples. *Journal of Consulting and Clinical Psychology, 72*, 176–191.
- Christenson, S. L., Carlson, C., & Valdez, C. R. (2002). Evidence-based interventions in school psychology: Opportunities, challenges, and cautions. *School Psychology Quarterly, 17*, 466–474.
- Cohen, H., Amerine-Dickens, M. T., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Journal of Developmental Behavioral Pediatrics, 27*, 143–155.
- Committee on Psychosocial Aspects of Child and Family Health. (1998). American Academy of Pediatrics guidance for effective discipline. *Pediatrics, 101*, 723–728.
- Creswell, J. W. (2006). *Qualitative inquiry and research design: Choosing among five approaches* (2nd ed.). Thousand Oaks, CA: Sage.
- Damschroder, L., Aron, D., Keith, R., Kirsh, S., Alexander, J., & Lowery, J. (2009). Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. *Implementation Science, 4*, 50.
- De Luca, R. V., & Holborn, S. W. (1992). Effects of a variable-ratio reinforcement schedule with changing criteria on exercise in obese and nonobese boys. *Journal of Applied Behavior Analysis, 25*, 671–679.
- Delmolino, L., & Harris, S. L. (2011). Matching children on the autism spectrum to classrooms: A guide for parents and professionals. *Journal of Autism and Developmental Disorders, 42*, 1197–1204.
- Detrich, R. (2008). Evidence-based education: Can we get there from here? *ABAI Newsletter, 31*(3). Retrieved from <http://www.abainternational.org/ABA/newsletter/vol313/Detrich.asp>
- Detrich, R., Keyworth, R., & States, J. (2007). A roadmap to evidence-based education:

- Building an evidence-based culture. *Savage Controversies*, 1(1), 3–5.
- Dixon, D. R., Vogel, T., & Tarbox, T. (2012). A brief history of functional analysis and applied behavior analysis. In J. L. Matson (Ed.), *Functional assessment of challenging behavior* (pp. 3–23). New York, NY: Springer.
- Dretzke, J., Davenport, C., Frew, E., Barlow, J., Stewart-Brown, S., Bayliss, S., ... Hyde, C. (2009). The clinical effectiveness of different parenting programmes for children with conduct problems: A systematic review. *Child and Adolescent Psychiatry and Mental Health*. Retrieved from <http://www.caphm.com/content/3/1/7>
- Dunlap, G., Iovannone, R., English, C., Kincaid, D., Wilson, K., Christiansen, K., ... English, C. (2010). *Prevent-teach-reinforce: A school-based model of positive behavior support*. Baltimore, MD: Brookes.
- Dunn, K. E., Sigmon, S. C., Thomas, C. S., Heil, S. H., & Higgins, S. T. (2008). Voucher-based contingent reinforcement of smoking abstinence among methadone-maintained patients: A pilot study. *Journal of Applied Behavior Analysis*, 41, 527–538.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Psychology*, 41, 327–350.
- EdMark Reading Program Level 1 and 2 Combo Kit [Computer software]. (2004). Austin, TX: Pro-Ed.
- Edwards, A., & Lutzker, J. R. (2009). Iterations of the SafeCare model. An evidence-based child maltreatment prevention program. *Behavior Modification*, 32, 736–756.
- Embry, D. D., & Biglan, A. (2008). Evidence-based kernels: Fundamental units of behavioral influence. *Clinical Child and Family Psychology*, 11, 75–113.
- Engelmann, S. (1980). *Direct instruction*. Englewood Cliffs, NJ: Educational Technology.
- Evans, C. H., & Ildstad, S. T. (Eds.). (2001). *Small clinical trials: Issues and challenges*. Washington, DC: National Academy of Sciences.
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication 231).
- Flippin, M., Reszka, S., & Watson, L. R. (2010). Effectiveness of the picture exchange communication system (PECS) on communication and speech for children with autism spectrum disorders: A meta-analysis. *American Journal of Speech-Language Pathology*, 19, 179–185.
- Forness, S. R. (2005). The pursuit of evidence-based practices in special education for children with emotional or behavioral disorders. *Behavioral Disorders*, 30, 311–330.
- Fox, R. M. (1996). Translating the covenant: The behavior analyst as ambassador and translator. *The Behavior Analyst*, 19, 147–161.
- Frankel, F. (2004, May). *Children's friendship training*. Invited address given at the meeting of the NIH Working Group on Methodological Challenges in Autism Treatment Research, Sacramento, CA.
- Freedman, D. H. (2011, February). How to fix the obesity crisis. *Scientific American*, 40–47.
- Friman, P. C. (2010). Come on in, the water is fine: Achieving mainstream relevance through integration with primary medical care. *The Behavior Analyst*, 33, 19–36.
- Galbicka, G. (1994). Shaping in the 21st century: Moving percentile schedules into applied settings. *Journal of Applied Behavior Analysis*, 27, 739–760.
- Gray, C. (2004). Social Stories 10.0: The new defining criteria. *Jenison Autism Journal*, 15, 1–21.
- Green, G. (1999). Science and ethics in early intervention for autism. In P. M. Ghezzi, W. L. Williams, & J. E. Carr (Eds.), *Autism: Behavior analytic perspectives* (pp. 11–28). Reno, NV: Context Press.
- Green, G. (2008). Single-case research methods for evaluating treatments for ASD. In S. C. Luce, D. S. Mandell, C. Mazefsky, & W. Seibert (Eds.), *Autism in Pennsylvania: A symposium issue of the Speaker's Journal of Pennsylvania Policy* (pp. 119–132). Harrisburg, PA: Legislative Office for Research Liaison, Pennsylvania House of Representatives.
- Greer, R. D. (2002). *Designing teaching strategies: An applied behavior analysis systems approach*. San Diego, CA: Academic Press.
- Grimshaw, J. M., Thomas, R. E., MacLellan, G., Fraser, C., Ramsey, C. R., Vale, L., ... Donaldson, C. (2004). Effectiveness and efficiency of guideline dissemination and implementation strategies. *Health Technology Assessment*, 8(6).
- Guralnick, M. J. (Ed.). (1997). *The effectiveness of early intervention*. Baltimore, MD: Brookes.
- Guralnick, M. J. (1998). Effectiveness of early intervention for vulnerable children: A developmental perspective. *American Journal on Mental Retardation*, 102, 319–345.
- Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., ... Schunemann, H. J. (2008). GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *British Medical Journal*, 336, 924–926.
- Guyatt, G., Rennie, D., Meade, M. O., & Cook, D. J. (2008). *Users' guides to the medical literature: A manual for evidence-based clinical practice* (2nd ed.). Chicago, IL: American Medical Association.

- Guyatt, G., Sackett, D., Taylor, D. W., Chong, J., Roberts, R., & Pugsley, S. (1986). Determining optimal therapy—Randomized trials in individual patients. *New England Journal of Medicine*, *314*, 889–892.
- Hagopian, L., & Wachtel, L. (n.d.). *Neurobehavioral unit*. Retrieved from <http://www.kennedykrieger.org/patient-care/patient-care-programs/inpatient-programs/neurobehavioral-unit-nbu>
- Handen, B. L., & Gilchrist, R. H. (2006). Mental retardation. In E. J. Mash & R. A. Barkley (Eds.), *Treatment of childhood disorders* (3rd ed., pp. 411–454). New York, NY: Guilford.
- Handleman, J. S., & Harris, S. L. (2001). *Preschool education progress for children with autism* (2nd ed.). Austin, TX: Pro-Ed.
- Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The scientist-practitioner: Research and accountability in the age of managed care* (2nd ed.). Boston, MA: Allyn & Bacon.
- Hibbs, E. D., Clarke, G., Hechtman, L., Abikoff, H. B., Greenhill, L. L., & Jensen, P. S. (1997). Manual development for the treatment of child and adolescent disorders. *Psychopharmacology Bulletin*, *33*, 619–629.
- Higgins, S. T., Sigmon, S. C., Wong, C. J., Heil, S. H., Badger, G. J., Donham, B. D., ... Anthony, S. (2003). Community reinforcement therapy for cocaine-dependent outpatients. *Archives of General Psychiatry*, *60*, 1043–1052.
- Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Council for Exceptional Children*, *71*, 165–179.
- Horner, R. H., & Kratochwill, T. R. (2011). Using single-case research to identify evidence-based practices. *Savage Controversies*, *5*(1), 2–5.
- Horner, R. H., Salentine, S., & Albin, R. W. (2003). *Self-assessment of contextual fit in schools*. Eugene, OR: University of Oregon.
- Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptional Children*, *42*(8), 1–15.
- Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities*, *26*, 359–383.
- Hyman, S., Stewart, P., Schmidt, B., Cain, U., Lemcke, L., Foley, J., ... Ng, P. (2012). Nutrient intake from food in children with autism. *Pediatrics*, *130*, S143–S153.
- Ingvarsson, E. T., Tiger, J. H., Hanley, G. P., & Stephenson, K. M. (2007). An evaluation of intraverbal training to generate socially appropriate responses to novel questions. *Journal of Applied Behavior Analysis*, *40*, 411–429.
- Iwata, B. A., & Dozier, C. L. (2008). Clinical application of functional analysis methodology. *Behavior Analysis in Practice*, *1*, 3–9.
- Johnston, J. M. (1988). Strategic and tactical limits of comparison studies. *The Behavior Analyst*, *11*, 1–9.
- Johnston, J. M. (1993). A model for developing and evaluating behavioral technology. In R. Van Houten & S. Axelrod (Eds.), *Effective behavioral treatment: Issues and implementation* (pp. 323–343). New York, NY: Plenum.
- Johnston, J. M. (1996). Distinguishing between applied research and practice. *The Behavior Analyst*, *19*, 35–47.
- Johnston, J. M. (2000). Behavior analysis and the R & D paradigm. *The Behavior Analyst*, *23*, 141–148.
- Johnston, J. M., Foxx, R. M., Jacobson, J. W., Green, G., & Mulick, J. A. (2006). Positive behavior support and applied behavior analysis. *The Behavior Analyst*, *29*, 51–74.
- Johnston, J. M., & Pennypacker, H. S. (1993). *Readings for strategies and tactics of behavioral research* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Johnston, J. M., & Pennypacker, H. S. (2009). *Strategies and tactics of behavioral research* (3rd ed.). New York, NY: Routledge.
- Kameenui, E. J., & Simmons, D. C. (1990). *Designing instructional strategies: The prevention of academic learning problems*. Columbus, OH: Merrill.
- Kasari, C., & Smith, T. (in press). Interventions in schools for children with ASD: Methods and recommendations. *Autism*.
- Kazdin, A. E. (1997). A model for developing effective treatments. Progression and interplay of theory, research, and practice. *Journal of Child Clinical Psychology*, *26*, 114–129.
- Kazdin, A. E. (2005). *Parent management training: Treatment for oppositional, aggressive, and antisocial behavior in children and adolescents*. Cary, NC: Oxford University Press.
- Kazdin, A. E. (2011). Evidence-based treatment research: Advances, limitations, and next steps. *American Psychologist*, *66*, 685–698.
- Keenan, M., & Dillenburger, K. (2011). When all you have is a hammer ...: RCTs and hegemony in science. *Research in Autism Spectrum Disorders*, *5*, 1–13.
- Kendall, P. C. (2012). Guiding theory for therapy with children and adolescents. In P. C. Kendall (Ed.), *Child and adolescent therapy: Cognitive-behavioral procedures* (4th ed., pp. 3–24). New York, NY: Guilford.
- Kendall, P. C., & Chu, B. C. (2000). Retrospective self-reports of therapist flexibility in a manual-based treatment for youths with anxiety disorders. *Journal of Clinical Child Psychology*, *29*, 209–220.

- Kendall, P. C., Chu, B., Gifford, A., Hayes, C., & Nauta, M. (1998). Breathing life into a manual: Flexibility and creativity with manual-based treatments. *Cognitive and Behavioral Practice, 5*, 177–198.
- Koegel, R. L., & Koegel, L. K. (2006). *Pivotal response treatments for autism: Communication, social, & academic development*. Baltimore, MD: Brookes.
- Koegel, R. L., Koegel, L. K., Vernon, T. W., & Brookman-Frazee, L. I. (2010). Empirically supported pivotal response treatment for autism spectrum disorders. In J. R. Weisz & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (2nd ed., pp. 327–344). New York, NY: Guilford.
- Köster, J. P., & Göttsche, P. C. (2008). Regular self-examination or clinical examination for early detection of breast cancer. *Cochrane Database of Systematic Reviews*. Issue 4. 2. Art. No.: CD003373. doi: 10.1002/14651858.CD003373
- Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case design technical documentation version 1.0 (Pilot). Retrieved from http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf
- Kratochwill, T. R., & Levin, J. R. (2010). Enhancing the scientific credibility of single-case intervention research: Randomization to the rescue. *Psychological Methods, 2*, 124–144.
- Kratochwill, T. R., & Stoiber, K. C. (2002). Evidence-based interventions in school psychology: Conceptual foundations of the *Procedural and Coding Manual* of Division 16 and the Society for the Study of School Psychology Task Force. *School Psychology Quarterly, 17*, 341–389.
- Kurtz, P. F., Boelter, E. W., Jarmolowicz, D. P., Chin, M. P., & Hagopian, L. P. (2011). An analysis of functional communication training as an empirically supported treatment for problem behavior displayed by individuals with intellectual disabilities. *Research in Developmental Disabilities, 32*, 2935–2942.
- Langley, G. J., Moen, R. D., Nolan, K. M., Nolan, T. W., Norman, C. L., & Provost, L. P. (2009). *The improvement guide: A practical approach to enhancing organizational performance*. San Francisco, CA: Jossey-Bass.
- Langthorne, P., & McGill, P. (2011). Assessing the social acceptability of the functional analysis of problem behavior. *Journal of Applied Behavior Analysis, 44*, 403–407.
- Larson, E. B. (1990). N-of-1 clinical trials: A technique for improving medical therapeutics. *The Western Journal of Medicine, 152*, 52–56.
- Leaf, R. B., Taubman, M. T., McEachin, J. J., Leaf, J. B., & Tsuji, K. H. (2011). A program description of a community-based intensive behavioral intervention program for individuals with autism spectrum disorders. *Education and Treatment of Children, 34*, 259–285.
- Lee, D. L. (2005). Increasing compliance: A quantitative synthesis of applied research on high-probability sequences. *Exceptionality, 3*, 141–154.
- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*, 3–9.
- Lutzker, J. R., & Whitaker, D. J. (2005). The expanding role of behavior analysis and support: Current status and future directions. *Behavior Modification, 29*, 575–594.
- Mace, F. C. (1991). Technological to a fault or faulty approach to technology development. *Journal of Applied Behavior Analysis, 24*, 433–435.
- MacMahon, R. J. (2004, May). *The Fast Track Project*. Invited address delivered at the meeting of the NIH Working Group on Methodological Challenges in Autism Treatment Research, Sacramento, CA.
- Martin, G. L., Thompson, K., & Regehr, K. (2004). Studies using single-subject designs in sport psychology: 30 years of research. *The Behavior Analyst, 27*, 263–280.
- Mayville, E. A., & Mulick, J. A. (Eds.). (2011). *Behavioral foundations of effective autism treatment*. Cornwall-on-Hudson, NY: Sloan.
- McCartney, E. J., Anderson, C. M., & English, C. L. (2005). Effect of brief clinic-based training on the ability of caregivers to implement escape extinction. *Journal of Positive Behavior Interventions, 7*, 18–32.
- McHugh, R. K., & Barlow, D. H. (Eds.). (2012). *Dissemination and implementation of evidence-based interventions*. Cary, NC: Oxford University Press.
- McMonagle, T., & Sultana, A. (2009). Token economy for schizophrenia (review). *Cochrane Database of Systematic Reviews*. Issue 1. Art. No. CD001473. doi: 10.1002/14651858.CD001473
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: A sourcebook of new methods* (2nd ed.). Thousand Oaks, CA: Sage.
- Miltenberger, R. G., Gatheridge, B. J., Satterlund, M., Egemo-Helm, K. R., Johnson, B. M., Jostad, C., ... Flessner, C. A. (2005). Teaching safety skills to children to prevent gun play: An evaluation of in situ training. *Journal of Applied Behavior Analysis, 38*, 395–398.
- Moore, J. W., & Fisher, W. W. (2007). The effects of videotape modeling on staff acquisition of functional analysis methodology. *Journal of Applied Behavior Analysis, 40*, 197–202.
- Mueller, M. M., Nkosi, A., & Hine, J. F. (2011). Functional analysis in public schools: A summary of 90 functional analyses. *Journal of Applied Behavior Analysis, 44*, 807–818.

- Nathan, P. C., & Gorman, J. M. (Eds.). (2002). *A guide to treatments that work* (2nd ed.). Cary, NC: Oxford University Press.
- National Autism Center. (2009). *National Standards Project: Findings and conclusions*. Randolph, MA: Author.
- National Secondary Transition Technical Assistance Center. (2010). *Levels of evidence*. Retrieved from <http://www.nsttac.org/sites/default/files/assets/pdf/pdf/ebps/DecisionRules.2pdf.pdf>
- O'Brien, S., Ross, L. V., & Christophersen, E. R. (1986). Primary encopresis: Evaluation and treatment. *Journal of Applied Behavior Analysis, 19*, 137–145.
- Odom, S. L., Brantlinger, E., Gersten, R., Horner, R., Thompson, B., & Harris, K. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 71*, 137–148.
- Odom, S. L., Hume, K., Boyd, B., & Stabel, A. (2012). Moving beyond the intensive behavior therapy versus eclectic dichotomy: Evidence-based and individualized learning programs for learners with ASD. *Behavior Modification, 36*, 270–297.
- Odom, S. L., McConnell, S. R., McEvoy, M. A., Peterson, C., Ostrosky, M., Chandler, L. K., ... Favazza, P. C. (1999). Relative effects of interventions supporting the social competence of young children with disabilities. *Topics in Early Childhood Special Education, 19*, 75–91.
- O'Donohue, W., & Ferguson, K. E. (2006). Evidence-based practice in psychology and behavior analysis. *The Behavior Analyst Today, 7*, 335–350.
- Paul, G. L., & Lentz, R. J. (1977). *Psychosocial treatment of chronic mental patients: Milieu versus social-learning programs*. Cambridge, MA: Harvard University Press.
- Paul, G. L., & Menditto, A. A. (1992). Effectiveness of inpatient treatment programs for mentally ill adults in public psychiatric facilities. *Applied and Preventive Psychology: Current Scientific Perspectives, 1*, 41–63.
- Pennypacker, H. S. (2010). *What is evidence?* Retrieved from <http://www.behavior.org/resource.php?id=351>
- Pennypacker, H. S., & Iwata, M. (1990). Mamma Care: A case history in behavioral medicine. In D. E. Blackman & H. Lejeune (Eds.), *Behavior analysis in theory and practice: Contributions and controversies* (pp. 259–288). Hillsdale, NJ: Erlbaum.
- Piacentini, J., Woods, D. W., Scchill, L., Wilhelm, S., Peterson, A. L., Chang, S., ... Walkup, J. T. (2011). Behavior therapy for children with Tourette disorder: A randomized controlled trial. *Journal of the American Medical Association, 303*, 1929–1937.
- Reese, E., Howard, J., & Reese, T. W. (1978). *Human operant behavior: Analysis and application*. Madison, WI: Brown & Benchmark.
- Rehfeldt, R. A. (2011). Toward a technology of derived stimulus relations: An analysis of articles published in *JABA*, 1992–2009. *Journal of Applied Behavior Analysis, 44*, 109–119.
- Reichow, B., Volkmar, F. R., & Cicchetti, D. V. (2008). Development of the evaluative method for evaluating and determining evidence-based practices in autism. *Journal of Autism and Developmental Disorders, 38*, 1311–1319.
- Robey, R. R. (2004). A five-phase model for clinical-outcome research. *Journal of Communication Disorders, 37*, 401–411.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York, NY: Free Press.
- Romeiser Logan, L., Hickman, R. R., Harris, S. R., & Heriza, C. B. (2008). Single-subject research design: Recommendations for levels of evidence and quality rating. *Developmental Medicine and Child Neurology, 50*, 99–103.
- Sackett, D. L., Rosenberg, W. M. C., Muir Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *BMJ, 312*, 71–72.
- Salzberg, C. L., Strain, P. S., & Baer, D. M. (1987). Meta-analysis of single-subject research: When does it clarify, when does it obscure? *Remedial and Special Education, 8*, 43–48.
- Schlosser, R. W. (2009). The role of single-subject research designs in evidence-based practice times. *FOCUS, 22*, 1–8.
- Schultz, K. F., Altman, D. G., Moher, D., & CONSORT Group. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *PLoS Medicine, 7*(3), e1000251.
- Seligman, M. E. P. (1993). *What you can change and what you can't: The complete guide to successful self-improvement*. New York, NY: Knopf.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin.
- Sharp, W. G., Jacquess, D. L., Morton, J. F., & Herzinger, C. V. (2010). Pediatric feeding disorders: A quantitative synthesis of treatment outcomes. *Clinical Child and Family Psychology Review, 13*, 348–365.
- Shillingsburg, M. A., Kelley, M. E., Roane, H. S., Kisamore, A., & Brown, M. R. (2009). Evaluation and training of yes-no responding across verbal operants. *Journal of Applied Behavior Analysis, 42*, 209–223.
- Shook, G. L., Johnston, J. M., & Mellichamp, F. (2004). Determining essential content for applied behavior analyst practitioners. *The Behavior Analyst, 27*, 67–94.
- Sidman, M. (1960). *Tactics of scientific research: Evaluating experimental data in psychology*. New York, NY: Basic Books.

- Skinner, B. F. (1945). Baby in a box: Introducing the mechanical baby tender. *Ladies' Home Journal*, 62(10), 30–31, 135–36, 138.
- Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review*, 24, 86–97.
- Skinner, B. F. (1956). A case history in the scientific method. *American Psychologist*, 11, 221–233.
- Skinner, B. F. (1968). *The technology of teaching*. New York, NY: Appleton-Century-Crofts.
- Slifer, K. J., & Amari, A. (2009). Behavior management for children and adolescents with brain injury. *Developmental Disabilities Research Reviews*, 15, 144–151.
- Smith, T. (2011). Applied behavior analysis and early intensive intervention. In D. G. Amaral, G. Dawson, & D. H. Geschwind (Eds.), *Autism spectrum disorders* (pp. 1037–1055). New York, NY: Oxford University Press.
- Smith, T. (2012). *Making inclusion work for students with autism spectrum disorders: An evidence-based guide*. New York, NY: Guilford.
- Smith, T., Scabill, L., Dawson, G., Guthrie, D., Lord, C., Odom, S., & Wagner, A. (2007). Designing research studies on psychosocial interventions in autism. *Journal of Autism and Developmental Disorders*, 37, 354–366.
- Society for Prevention Research. (2004). *Standards of evidence: Criteria for efficacy, effectiveness, and dissemination*. Retrieved from <http://www.preventionresearch.org/StandardsOfEvidencebook.pdf>
- Stark, L. J., Spear, S., Boles, R., Kuhl, E., Ratcliff, M., Scharf, C., ... Rauch, J. (2011). A pilot randomized controlled trial of a clinic and home-based behavioral intervention to decrease obesity in preschoolers. *Obesity*, 19, 134–141.
- Strain, P. S., Kohler, F. W., & Gresham, F. (1998). Problems in logic and interpretation with quantitative syntheses of single-case research: Mathur and colleagues (1998) as a case in point. *Behavioral Disorders*, 24, 74–85.
- Sugai, G., Horner, R. H., Algozzine, R., Barrett, S., Lewis, T., Anderson, C., ... Simonsen, B. (2010). *School-wide positive behavior support: Implementers' blueprint and self-assessment*. Eugene, OR: University of Oregon.
- Sugai, G., Lewis-Palmer, T., Horner, R. H., & Todd, A. W. (2005). *School-wide evaluation tool version 2.1*. Eugene, OR: University of Oregon.
- Sulzer-Azaroff, B., & Austin, J. (2000). Does BBS work? Behavior-based safety and injury reduction: A survey of the evidence. *Professional Safety*, 45(7), 19–24.
- Szilagyi, P. G. (2009). Translational research and pediatrics. *Academic Pediatrics*, 9, 71–80.
- Tee, J., & Kazantzis, N. (2011). Collaborative empiricism in cognitive therapy: A definition and theory for the relationship construct. *Clinical Psychology: Science and Practice*, 18, 47–61.
- Thornicroft, G., Lempp, H., & Tansella, M. (2011). The place of implementation science in the translational medicine continuum. *Psychological Medicine*, 41, 2015–2021.
- Tiger, J. H., Hanley, G. P., & Bruzek, J. (2008). Functional communication training: A review and practical guide. *Behavior Analysis in Practice*, 1, 16–23.
- Trahan, M. A., Kahng, S. W., Fisher, A. B., & Hausman, N. L. (2011). Behavior-analytic research on dementia in older adults. *Journal of Applied Behavior Analysis*, 44, 687–691.
- Twyman, J. S., Layng, T. V. J., Stikeleather, G., & Hobbins, K. A. (2004). A non-linear approach to curriculum design: The role of behavior analysis in building an effective reading program. In W. L. Heward, T. E. Heron, N. A. Neef, S. M. Peterson, D. M. Sainato, G. Cartledge, R. Gardner III, et al. *Focus on behavior analysis in education: Achievements, challenges, and opportunities* (pp. 55–68). Upper Saddle River, NJ: Prentice Hall.
- United States Preventive Services Task Force. (2007). *Grade definitions*. Retrieved from <http://www.uspreventiveservicestaskforce.org/uspstf/grades.htm>
- Vegas, K. C., Jenson, W. R., & Kircher, J. C. (2007). A single-subject meta-analysis of the effectiveness of time-out in reducing disruptive classroom behavior. *Behavioral Disorders*, 32, 109–121.
- Wakefield, J. C. (2006). Is behaviorism becoming a pseudo-science? Power versus scientific rationality in the eclipse of token economies by biological psychiatry in the treatment of schizophrenia. *Behavior and Social Issues*, 15, 202–221.
- Weisz, J., Chorpita, B., Palinkas, L., Schoenwald, S. K., Miranda, J., Bearman, J., ... Givvons, R. D. (2012). Testing standard and modular designs for psychotherapy treating depression, anxiety, and conduct problems in youth. *Archives of General Psychiatry*, 69, 274–282.
- West, S., King, V., Carey, T. S., Lohr, K. N., McKoy, N., Sutton, S. F., & Lux, L. (2002). *Systems to rate the strength of scientific evidence* (AHQR Publication No. 02-E016). Rockville, MD: Agency for Healthcare Research and Quality. Retrieved from <http://www.ahrq.gov/clinic/evrptfiles.htm#strength>
- What Works Clearinghouse. (2007). *Direct Instruction, DISTAR, and Language for Learning*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/intervention_reports/WWC_Direct_Instruction_052107.pdf
- Wong, S. E. (2006). Behavior analysis of psychotic disorders: Scientific dead end or casualty of the mental health political economy? *Behavior and Social Issues*, 15, 152–177.
- Woolf, S. H., & Johnson, R. E. (2005). The break-even point: When medical advances are less important than improving the fidelity with which they are delivered. *Annals of Family Medicine*, 3, 545–552.