



Replication Research, Publication Bias, and Applied Behavior Analysis

Matt Tincani¹  · Jason Travers²

Published online: 18 March 2019

© Association for Behavior Analysis International 2019

Abstract

The “replication crisis” describes recent difficulties in replicating studies in various scientific fields, most notably psychology. The available evidence primarily documents replication failures for group research designs. However, we argue that contingencies of publication bias that led to the “replication crisis” also operate on applied behavior analysis (ABA) researchers who use single-case research designs (SCRD). This bias strongly favors publication of SCR D studies that show strong experimental effect, and disfavors publication of studies that show less robust effect. The resulting research literature may unjustifiably inflate confidence about intervention effects, limit researchers’ ability to delineate intervention boundary conditions, and diminish the credibility of our science. To counter problems of publication bias in ABA, we recommend that journals that publish SCR D research establish journal standards for publication of noneffect studies; that our research community adopt open sharing of SCR D protocols and data; and that members of our community routinely publish systematic literature reviews that include gray (i.e., unpublished) research.

Keywords Applied behavior analysis · Replication · Publication bias · File drawer effect · Single-case design · Behavior science

Scientific experimentation is the best, albeit imperfect, approach to understanding our surroundings, and has proven useful for clarifying functional relationships between the environment and the behavior of organisms (Skinner, 1953). The imperfections of a

✉ Matt Tincani
tincani@temple.edu

Jason Travers
jason.travers@ku.edu

¹ Department of Teaching and Learning, Temple University, 1301 Cecil B. Moore Ave., Ritter Hall 351, Philadelphia, PA 19122, USA

² Department of Special Education, University of Kansas, Joseph R. Pearson Hall, Rm. 547, 1122 W. Campus Rd, Lawrence, KS 66045-3101, USA

science are a manifestation of ignorance about how to perfect the craft, but also of human fallibility and proclivity to error. Scientists are trained to avoid or reduce the probability of error; however, they may mistakenly draw false positive and false negative conclusions from their experiments. In addition to errors of interpretation, scientists may draw inaccurate conclusions from experimental data in a way that reflects bias (Mahoney, 1977). Such contexts make the process of scientific knowledge accumulation difficult, and may lead to inflated confidence about phenomena of interest (e.g., Doyen, Klein, Pichon, & Cleeremans, 2012). Scientific claims to knowledge are, therefore, considered tentative, and confidence in claims should correspond with the amount and quality of evidence. More evidence is therefore generally preferred over less evidence, and high-quality (i.e., experimental) evidence is preferred over low-quality (e.g., anecdotal) evidence. The imperfect craft of science carried out by imperfect scientists makes replication of experimental studies a primary tenet of scientific advancement and knowledge generation. However, replication research appears to be relatively rare in some scientific fields (e.g., education; Cook, Collins, & Cook, 2016; Makel & Plucker, 2014).

Researchers in a number of disciplines over the past several years have, therefore, sought to generate studies in which the methods of previously published studies are replicated to determine if the findings are valid (e.g., Ioannidis, 2012; Makel & Plucker, 2014; Nosek, Spies, & Motyl, 2012; Pashler & Harris, 2012). Systematic replication projects have been reported in psychology (Open Science Collaboration, 2015), economics (Evanschitzky & Armstrong, 2010), medicine (Ioannidis, 2005), and business (Hubbard & Armstrong, 1994). A common outcome of replication studies is that original findings are absent or, if detected, less robust than in the original report. The so-called “replication crisis” has affected a number of scientific fields; however, psychology seems to be at the center of concern (Earp & Trafimow, 2015; Pashler & Wagenmakers, 2012; Hales, Wesselmann, & Hilgard, 2018). Although the replication crisis has been blamed, in part, on questionable research practices involving manipulation and selective reporting of data collected during group experimental research (e.g., Kerr, 1998; Simmons, Nelson, & Simonsohn, 2011), such practices likely are motivated and reinforced by publication bias, or the tendency of journals to exclusively publish studies that find statistically significant effects (Franco, Malhotra, & Simonovits, 2014; Lilienfeld, 2017; Rosenthal, 1979). One implication of the crisis is that psychological phenomena purportedly based on evidence does not hold up when subjected to further scientific scrutiny (i.e., replication studies). More broadly, a lack of sound replication studies foments public distrust in science, scientists, and scientific evidence.

In this article, we contend that applied behavior analysis (ABA) researchers are not immune to the contingencies of publication bias affecting the group design research communities. In fact, we suspect that ABA researchers who use single-case research designs (SCRD) may be exposed to the same problems associated with publication bias, namely, inflated confidence about intervention effects and the obscuring of intervention boundary conditions. We discuss contingencies in the academic environment that select for behavior that may mislead researchers and a corpus of knowledge. Finally, we explore the purposes for conducting replication research in ABA, and how applied behavior analysts, as a scientific research community, can enhance contingencies for researchers to conduct and publish replication studies, including those that demonstrate modest or no experimental control.

Replication in Applied Behavior Analysis

Following Skinner's (1938) earliest attempts to develop a science of behavior, behavior scientists and analysts have sought ways to confirm, clarify, and extend what appeared to be known about behavior (e.g., Ferster & Skinner, 1957). Initial results were subjected to scrutiny, and experimental methods, instruments, and findings were refined in ways that led to a robust scientific theory of selective pressure on behavior (Pierce & Cheney, 2017). Other sciences have proceeded in a similar way by emphasizing the verification of claims made by peer scientists. Independent replication of experimental studies is a primary tenet of scientific advancement and knowledge generation.

Replication played an important role in the discovery of operant principles of behavior and was evident in the progression of Skinner's early experiments and experiments of those who followed. Sidman (1960) offered a succinct and comprehensive definition of replication, which heralded the development of contemporary ABA research methods in which replication was an essential feature (Baer, Wolf, & Risley, 1968). Sidman carefully delineated the importance of direct and systematic replication in behavior analysis research. Direct replication is used to establish the reliability of a phenomenon of interest through repeated demonstration of experimental effect. For example, the reversal design, which entails application and removal of an intervention following an initial baseline phase (Baer et al., 1968), allows researchers to demonstrate direct (in this case, intrasubject) replication if intervention delivery reliably coincides with behavior change, and intervention removal reliably coincides with behavior returning to baseline levels. Systematic replication establishes the generality of a phenomenon "over a wide range of situations" (Sidman, 1960, p. 110). These situations may include but are not limited to varied intervention procedures, participant characteristics, implementers, and settings.

High-quality research SCRD exhibits the qualities of both direct and systematic replication (Horner et al., 2005; Kratochwill et al., 2013). Horner et al. (2005) stipulated that SCRD studies should provide, at minimum, three demonstrations of experimental effect (direct replication) across multiple participants, settings, materials and/or behaviors (systematic replication). Systematic replication is especially critical to contemporary ABA research given the variety of individual characteristics and circumstances of applied settings. For instance, a researcher who develops an intervention to improve the communicative skills of children with autism spectrum disorder (ASD) may wish to evaluate whether it is effective for children of different ages (e.g., preschool versus school age) and abilities (e.g., speaking versus nonspeaking), or with different implementers (e.g., expert clinicians versus classroom teachers) in different settings (e.g., school versus community). Greater generality of effect indicates greater treatment utility across a population.

There is little doubt that behavior scientists and behavior analysts as a research community recognize the intrinsic importance of replication in our science. A pillar of evidence-based practice in ABA is the concept that the best available research evidence should always inform clinical decisions (Slocum et al., 2014). However, there is little consensus about how much replication is necessary to establish an intervention as having sufficient empirical support for broad application. Likewise, there is little consensus about the type of replication activities needed to demonstrate satisfactory or compelling empirical support.

Publication Bias in ABA Research

The limited use of group comparison research designs and heavy dependence on SCRD in behavior analysis may convey among behavior analysts a sense of immunity from the sorts of problems that underlie the replication crisis in other fields. However, behavior analysts are human organisms and are thus subject to the same contingencies that affect scientists who primarily rely on group experimental designs (Mahoney, 1994). In recent years, the academic environment has increasingly reinforced the quantity of research studies published by academics, creating a competitive environment during an era when resources were less available due to increasing costs of coupled with dwindling public funding of higher education (Lilienfeld, 2017; Mitchell, Leachman, and Masterson, 2016). The increasing emphasis on publication metrics, at both the journal and researcher level (e.g., journal impact factor, researcher h-index, citation count) likely functions as reinforcement for publishing in particular journals, publishing at a high rate, and for attaining a large number of citations for one's publications (Lane, 2010). Under circumstances in which these contingencies are less influential, publishing in certain niche outlets—high-status journals that are affiliated with an organization attached to one's discipline—may nonetheless confer particular social prestige and recognition (e.g., Dixon, Reed, Smith, Belisle, & Jackson, 2015), attract desirable attention to one's organization, or result in more tangible reinforcers (e.g., increased invitations for speaking engagements, awards and honors, paid consulting opportunities, promotion, tenure, and enhanced prospects for future employment). Such contingencies motivate and reinforce biases in ways that can contribute to erroneous findings.

We are not suggesting that selective data reporting, manipulation, or fabrication in group research that have been documented in highly publicized cases (e.g., Gelman & Fung, 2016) are commonplace in ABA research, though some ABA researchers may engage in questionable research practices, such as selective data reporting (Shadish, Zelinsky, Vevea, & Kratochwill, 2018). However, we assert the kinds of contingencies that lead to such behavior in the broader scientific community also operate on ABA researchers and may contribute to publication bias. If such behavior occurs with prevalence sufficient to mislead the ABA community, then the integrity of our research is undermined and consumers of our research may not benefit from ABA services.

Consider a hypothetical clinical ABA researcher who developed an intervention to reduce challenging behavior of individuals with disabilities served by their agency. Development of this intervention required substantial commitments of time and resources by the researcher, their clinical team, and the agency. Thus, it is reasonable to conclude the researcher is invested in the intervention. The researcher's investment is amplified by the clinical team providing conference presentations and workshops touting the intervention's effects for reducing challenging behavior. Attention to the intervention among other researchers and clinicians accrues, and clinicians from other agencies and schools express strong interest in adopting it for their consumers. Given these reinforcing consequences, the researcher then seeks to evaluate the intervention with three new participants, each of whom will receive the intervention within three simultaneous ABAB reversal designs (Horner et al., 2005), in hopes of publishing the findings. Baseline responding (i.e., challenging behavior) of Participants 1 and 2 yields high levels and stable responding. However, the baseline of Participant 3 yields

inconsistent, variable, and unstable responding. For Participants 1 and 2, the first intervention condition reduces their challenging behavior to low and stable levels. Following a second baseline, this effect is replicated in the final intervention condition with strong evidence of experimental control for these two participants. However, the intervention does not produce a discernable effect for Participant 3, whose initial baseline was unstable, and responding across each condition of the ABAB reversal design reveals little evidence of experimental control for this participant.

As a result, the researcher who is invested in the intervention elects to exclude the dataset for Participant 3 when submitting the study to a peer-reviewed journal for publication. The researcher may justify this decision in three ways: (a) uncontrolled variables resulted in varied responding in the baseline and intervention conditions; (b) this equivocal dataset yielded no experimental control and will distract readers from the strong experimental effects for the other two participants; and (c) journal editors will likely reject the manuscript for publication if it includes a dataset that reflects poor experimental control.

We do not believe these three reasons, individually or collectively, justify dropping a dataset from an experiment or failing to submit a study for publication in which one or more datasets shows less-than-optimal experimental control. First, although Participant 3's baseline data were initially unstable, operant responding in applied conditions is often likewise unstable and, therefore, the dataset may reasonably approximate authentic conditions under which this intervention would be applied. Furthermore, Participant 3's equivocal responding may highlight important clinical boundaries of this particular intervention for similar participants. That is, data for Participant 3 may indicate the intervention is less effective for individuals who display highly variable levels of challenging behavior, or whose organismic, situational, or other circumstances mitigate intervention effects. If the researcher collected procedural fidelity data in both baseline and intervention conditions, which showed that extraneous variables did not differentially affect baseline and intervention responding, this would further support these data as reflective of intervention boundaries (Tincani & Travers, 2018). It is important to note that this information need not distract readers from the positive effects of the intervention for the other two participants, but rather it could illustrate potential differences in responsiveness to the intervention based on unique characteristics of Participant 3. This information is indispensable for secondary consumers—practitioners and parents—who may seek to adopt the intervention, but also to discover the conditions under which the intervention is likely to be effective, ineffective, or countertherapeutic. Furthermore, this information may occasion future experimental investigations (cf. Perone, 2018), which seek to isolate the variables that facilitate or reduce intervention effectiveness. As a result, clinically useful variations in the intervention may be discovered. In a broader sense, submitting all data collected within and across SCRD experiments represents an intellectually honest approach, researcher investment in the intervention notwithstanding.

The third concern, that a journal will reject the study if the data show poor experimental control, is perhaps the most legitimate given that experimental control has been widely viewed as a necessary feature of rigorous SCRD (Baer et al., 1968; Cooper, Heron, & Heward, 2007; Horner et al., 2005). Cooper et al. (2007) suggested, “An experiment is interesting and convincing, and yields the most useful information for application, when it provides an unambiguous demonstration that the independent

variable was solely responsible for the observed behavior change” (p. 230). ABA researchers, journal reviewers, and editors have likely interpreted this to mean that SCRD studies that fail to demonstrate “unambiguous” experimental control are simply flawed experiments. Thus, journal submission, reviewing, and editorial practices may be biased against publishing such studies.

One way to detect publication bias is to examine differences in effect sizes between published and unpublished (i.e., gray) research studies, including unpublished doctoral dissertations (Gage, Cook, & Reichow, 2017). For example, Sham and Smith (2014) compared effect sizes for published and unpublished SCRD studies on a behaviorally based intervention called pivotal response treatment (PRT; Koegel & Koegel, 2006), which has been characterized as an evidence-based practice based entirely on research published in referred journal articles (Wong et al., 2015). Sham and Smith reported that published studies had larger treatment effects than unpublished studies, suggesting publication bias for this intervention. In particular, their findings suggested that researchers were less likely to submit PRT studies for publication when effects were not robust, that reviewers and journal editors were less likely to accept studies for publication given lack of robust effects, or of both.

A different method for detecting publication bias involves appraising how researchers and reviewers evaluate datasets that demonstrate varying degrees of experimental control. Mahoney (1977) presented 75 reviewers for the *Journal of Applied Behavior Analysis* with hypothetical datasets demonstrating varying degrees of experimental control. He found reviewers tended to rate manuscripts demonstrating strong experimental control more positively, even though the procedures in all studies were identical. More recently, Shadish et al. (2016) presented SCRD research experts with hypothetical SCRD datasets that demonstrated varying degrees of experimental effect. A majority of researchers reported that they were more likely to recommend publication of a submitted manuscript when a dataset showed positive effects, and a minority of researchers indicated a willingness to drop datasets showing weaker experimental effect when submitting a manuscript for publication. These results collectively highlight how contingencies operate on ABA researchers in ways that produce publication bias in the manuscript submission and review process.

Replication in Contemporary Applied Behavior Analysis Research

Scientists in other fields have acknowledged widespread problems of publication bias and lack of published replication and nonreplication studies (Franco et al., 2014; Makel & Plucker, 2014; Simmons et al., 2011). In response, systematic efforts have been undertaken to evaluate the extent to which findings reported in the literature are actually reproducible (e.g., Open Science Collaboration, 2012), and to evaluate the extent to which publication bias is reflected within published reviews of research (e.g., Driessen, Hollon, Bockting, Cuijpers, & Turner, 2015). No similar efforts have been mounted in ABA, leaving the credibility of published scientific findings open to question. Thus, we believe concerted efforts to conduct and publish replication and nonreplication studies in the field are justified by three reasons. These are to (a) reveal truths about behavior, (b) establish generality of applied behavior analytic interventions, and (c) demarcate boundary conditions of applied behavior analytic interventions.

Reveal Truths about Behavior

Replication studies are necessary for revealing truths about relationships between the environment and behavior. Claims supported by evidence are provisionally accepted as true, but confidence in such claims is proportionate to the amount and quality of available evidence. That is, confidence placed in truth claims increases with more systematic replications of effect. On the other hand, failed replications may conflict with previous claims, decrease confidence about the particular behavioral phenomenon of interest, or clarify delimiting factors that impinge on the phenomenon (i.e., intervention boundary conditions).

A commitment to a scientific worldview requires researchers to conduct experiments and gather evidence from a sufficient number of studies prior to claiming an intervention is effective. Despite recent attempts to quantify the number and kind of successful replication and nonreplication studies required to reach a conclusion about a particular intervention, there is currently little consensus on the minimum threshold of credible evidence beyond the notion that more evidence is better (e.g., Council for Exceptional Children, 2014; Chambless & Hollon, 1998; Kratochwill et al., 2013; Lonigan, Elbert, & Johnson, 1998). Thus, a researcher who touts an intervention on the basis of a single study or few studies is only justified in making tentative claims while acknowledging the need for more research, regardless of how strongly the intervention confirms the researcher's theoretical and conceptual perspectives. On the other hand, a researcher who presents a large body of studies conducted by several different research teams with apparently convergent findings can make claims about an intervention with little potential damage to reputation or status. However, reporting an intervention is effective while ignorant of documented failures to replicate has undesirable implications for the researcher making the claim, the professional who applies the intervention in practice, and the recipient who is presumed to benefit from the intervention. In such cases, individuals are harmed by a failure to seek and acknowledge intervention boundaries, and exclusive attention to studies that report only positive effects likely contributes to increased risk of harm. Addressing this potential problem, which appears entwined with publication bias, requires a value for and commitment to discovering what is true.

A commitment to truth is important not only for the integrity of the researcher or others who advocate for an intervention, but also for the researcher's field or discipline in a broader sense. A field whose researchers have collectively published a large number of rigorous, well-conducted replication experiments (along with rigorous, well-conducted nonreplication studies) are permitted greater confidence about claims to scientific truth than one that does not. On the other hand, a field whose researchers or practitioners make bold claims about the efficacy of interventions in the absence of a sound body of replication and nonreplication studies at best has a tenuous understanding of the truth. Likewise, a field whose researchers ignore the possibility of and evidence for the boundaries of a particular intervention or practice prioritizes dogma over truth. The availability, evaluation, and consideration of the entire body of evidence, including rigorous studies that detected reduced or no effects, is critical for the scientific advancement of knowledge in the respective field.

Establish Generality of ABA Interventions

The importance of systematic replication in ABA research reflects a need to establish the generality of ABA interventions across a wide array of human problems. ABA researchers and practitioners often advocate for interventions that solve particular social problems, and they do so based on a sound body of experimental evidence indicating ABA-based strategies are likely to produce better consumer outcomes than other approaches. For example, ABA researchers have advocated for behaviorally based treatments for ASD on the basis of supportive research while simultaneously warning consumers to avoid other approaches that lack similar empirical support (Zane, Davis, & Rosswurm, 2014) or could potentially produce harmful outcomes (Foxy & Mulick, 2016; Travers, Tincani, & Lang, 2014).

An implicit assumption of an evidence-based approach is that therapeutic effects of ABA interventions are generalizable across a wide variety of organismic and situational variables. For example, the recommendation that young children diagnosed with ASD should receive early and intensive behavioral interventions accompanies the assumption that, although ASD is a heterogeneous neurological condition encompassing a broad spectrum of behavioral manifestations, the positive effects of ABA are sufficiently generalizable that many, if not most young children who receive this treatment will benefit. This claim requires and is supported by large and robust body of research that shows ABA interventions are effective across a wide variety of individual characteristics (Eldevik et al., 2009; Reichow, 2012).

In the earliest stages of research and development, the positive effects of an intervention are likely to be observed with only a small number of similar individuals in similar circumstances. However, to demonstrate the generality of the interventions, researchers usually seek to vary, and thereby expand, the relevant organismic and situational variables under which the intervention is evaluated. Thus, as a line of research progresses, we would expect to see an increasing number of different characteristics represented in the literature, which translates into broader generality of the intervention, as long as positive effects continue to be observed. The progression of research also might lead to clarifications under which the intervention produces diminished or no effects (i.e., the boundaries of an otherwise effective intervention).

Demarcate Boundary Conditions

A novel intervention entails a protocol that, based on theory, concepts, principles, and relevant facts, is reasonably suspected to confer benefit to the recipient. Novel intervention protocols may be composed of various procedures that are believed to be important for generating the expected result. With experience (i.e., experimental application; replication), the protocol may evolve, with dispensable procedures eliminated, essential procedures retained, and new procedures developed for experimental analysis. That is, selective pressure acts on the researcher and the research community to produce effective and efficient intervention protocols. For example, recognizing practical limitations of standard functional analysis (FA) techniques (Carr & McDowell, 1980; Iwata, Dorsey, Slifer, Bauman, & Richman, 1982, 1994), researchers sought to empirically evaluate alternative techniques that were less resource intensive and more practically feasible for clinical and naturalistic settings (Bloom, Iwata, Fritz, Roscoe,

& Carreau, 2011; Cooper, Wacker, Sasso, Reimers, & Donn, 1990; Jessel, Hanley, & Ghaemmaghami, 2016). The resulting procedures have varied from the original protocols in ways intended to enhance the efficiency and usability of the approach while retaining certain essential elements of the original protocols (e.g., experimental manipulation of contingencies). Replications can also clarify for whom a particular intervention is more or less effective, along with critical features of the setting, such as implementer skill and training, consumer values and acceptance, and available resources to support high fidelity implementation, all of which mediate the intervention's effectiveness under applied circumstances.

What Kinds of Nonreplication Studies Should Be Published?

Interventions that are supported by a large body of SCRD evidence do not ensure that all recipients will experience the same or any benefit. There are no guarantees an intervention will be similarly effective for achieving a desired outcome, regardless of whether a large number of studies exist that suggest it ought to. Well-designed studies that fail to replicate a previous and well-documented effect therefore constitute a valuable discovery and provide a potential stimulus for scientific innovation. Researchers can and should examine how varied intervention procedures, dosages, contexts, and participant responses modulate intervention effects and therefore inform practical decisions of social significance. This is achieved by considering not only those studies with positive effects, but also well-designed studies without positive effects (c.f. Perone, 2018).

ABA researchers have published studies that do not report positive effects, but the investigations have usually centered on questionable practices that have little or no credible empirical evidence to begin with, such as sensory-based interventions (Barton, Reichow, Schnitz, Smith, & Sherlock, 2015; Cox, Gast, Luscre, & Ayres, 2009; Losinski, Cook, Hirsch, & Sanders, 2017). Typical contingencies of publication bias probably do not operate here because the interventions are poorly regarded by ABA researchers, and therefore negative findings are likely to be embraced by the ABA community (e.g., Foxx & Mulick, 2016). Although necessary and important, this research does little to demarcate the boundaries of behaviorally based interventions. Therefore, we propose a different course of action that focuses on examination of behavior interventions that are established and generally accepted. For example, preliminary studies on a behaviorally based intervention, stimulus–stimulus pairing, showed that it was effective in increasing vocalizations of young children with ASD (e.g., Yoon & Bennett, 2000). In a later study, Normand and Knoll (2006) showed that stimulus–stimulus pairing did not increase vocalizations for a young child with ASD in their study. Because their findings were discrepant with previous research, they suggested differences could be attributed to a number of specific boundary conditions, including variations in participant characteristics, reinforcer potency, and number of pairing trials, along with the presence of a control condition in their study. It is important to note that identifying potential boundaries of an intervention in this manner conveys critical information to practitioners about the conditions under which an intervention may be more or less effective. Detection of boundary conditions also occasions future research aimed at isolating variables that facilitate or hinder the

effectiveness of a particular therapeutic practice, such as the stimulus–stimulus pairing procedure (Shillingsburg, Hollander, Yosick, Bowen, & Muskat, 2015).

Given the emphasis of SCRD on describing functional relations at the individual level, behavior analytic researchers are uniquely positioned to conduct this type of research. Although experimenters, particularly novice researchers, may be inclined to view replication failures with disappointment due to reduced value associated with them, such findings may clarify opportunities for new experiments aimed at isolating effects of previously unknown variables and delineating boundary conditions. This trial and error approach is, in part, what makes science exciting (see Perone, 2018). For example, Ledford et al. (2016) summarized findings from three different master's theses with varied results. One thesis involved planned errors in a simultaneous prompting procedure to evaluate whether high-fidelity implementation produced a more therapeutic response than low-fidelity implementation. Results indicated one of the two study participants responded well during both low- and high-fidelity implementation, whereas the other only responded to high-fidelity implementation. Ledford et al. suggested that high-frequency delivery of simultaneous prompting may be necessary for some learners to achieve therapeutic benefit, and that less frequent delivery of simultaneous prompting, as might occur in naturalistic settings, may be sufficiently therapeutic for other learners. This difference in responsiveness highlights important variance in the therapeutic boundaries for this particular intervention. However, Ledford et al. also explained the results may appear unpublishable due to discrepant results that might be construed as indicators of threats to internal validity. Relegation of this and similar studies to the file drawer prevents further understanding about how these (and other) intervention procedures might be adapted to clarify which elements are essential and dispensable for different learners and contexts.

Importantly, SCRD studies in which the independent variable did not appear to affect responding may still possess most design features consistent with high quality SCRD (e.g., Horner et al., 2005; Kratochwill et al., 2013). Rigorous SCRD studies that contradict previous experiments may nonetheless be greeted with a higher degree of skepticism by reviewers, editors, and the ABA community in general due to conflicting findings. Therefore, researchers who wish to publish nonreplication studies should anticipate needing to exceed typical standards of rigorous SCRD in order to justify publication. Tincani and Travers (2018) outlined 11 types of evidence associated with aspects of intervention intensity, baseline and treatment condition integrity, and a moderator analysis that should be present in a high-quality SCRD study that fails to yield robust experimental control. In particular, they suggested that studies in which the IV did not appear to affect responding may reveal a boundary of the intervention if several conditions of study design and reporting were present. These include typical standards of high-quality SCRD studies as well as evidence of (a) high procedural fidelity across all study conditions, including baseline, reported for each participant and step in the intervention protocol; (b) evidence of intervention intensity (i.e., dose, dose frequency, and dose duration) consistent with previous research, and (c) sufficient evidence that other potential confounds were not present. Tincani and Travers proposed that such a collection of evidence could reasonably support claims that an intervention boundary may have been discovered in the absence of solid experimental control. On the other hand, studies without such evidence should be considered of insufficient quality for publication.

Changing Contingencies to Address Publication Bias

Recognizing the importance of replication research in revealing scientific truth, establishing generality of ABA interventions, and establishing intervention boundary conditions, we offer three recommendations for SCRD researchers that where the IV did not produce the expected effect. These are (a) establish journal standards for publishing no effect SCRD studies; (b) encourage open sharing of research protocols and data; and (c) include gray research in systematic reviews and meta-analyses of ABA studies.

Establish Journal Standards

ABA researchers may be disinclined to submit for publication research studies that did not replicate previous results because the probability of successful publication is lower due to reviewer and/or editor bias against such studies (Mahoney, 1977; Shadish et al., 2016). Journals that publish ABA research can mitigate against publication bias by establishing explicit standards that specify conditions for publishing no effect studies (Kittelman, Gion, Horner, Levin, & Kratochwill, 2018), including features of studies that are more likely to be accepted (Kratochwill, Levin, & Horner, 2018; Tincani & Travers, 2018). Inviting researchers to submit replication studies to a particular journal (e.g., Hanley, 2017) could increase publication of replication studies of established interventions that did not produce the expected effect. However, researchers may remain hesitant to submit these papers without explicit descriptions of favorable contingencies that lead to publication. For example, Kittelman et al. (2018) reviewed the submission guidelines of 29 top-ranked peer reviewed journals in the fields of general education, special education, educational psychology, and counseling. Only one included explicit guidelines to encourage submission of studies that yielded less than optimal results. However, Kittelman et al. also surveyed those journal editors and found a large majority reported willingness to accept for publication group comparison and SCRD studies without experimental effects if rigorous methods and high intervention integrity were evident.

Studies conducted by ABA researchers that did not produce a beneficial effect tend to be investigations of nonbehavioral interventions of a questionable nature (e.g., Cox et al., 2009; Davis et al., 2013; Denton & Meindl, 2016; Losinski et al., 2017; Quigley, Peterson, Frieder, & Peterson, 2011). However, publication of these studies suggests peer reviewers and editors for behavioral journals are open to publishing studies of interventions that did not produce expected effects if a high level of methodological quality is evident. Nonetheless, reviewers may still be disinclined to recommend acceptance of such studies if they have an investment in the intervention and/or the contradictory findings conflict with their own theoretical, conceptual, or methodological orientation (i.e., behavioral interventions; Mahoney, 1977). Collectively, these factors suggest journalistic standards that describe specific favorable contingencies for publication of no effect studies, including experimental interventions rooted in ABA, as guidelines for reviewers and editors during the review process.

Encourage Open Sharing of Research Protocols, Software, and Data

A different way to facilitate dissemination of SCRD studies that do not produce expected effects is to create contingencies that encourage the open sharing of research

protocols, software, and data, including data that reflect absence of experimental control. Uploading research protocols, software, and data to open access repositories may promote visibility of studies that might otherwise be subjected to publication bias or the file drawer effect. The open science movement (OSM) aims to increase the transparency of science and reproducibility of scientific findings by encouraging free, open access to research and data (Open Science Collaboration, 2012; UNESCO, 2018). The Open Science Foundation, for instance, maintains an online registry (<https://osf.io/>) that researchers can use to freely share unpublished manuscripts (e.g., methodologically sound intervention studies that produce limited or no beneficial effects). The availability of such papers on preprint servers (e.g., PsyArXiv) increases access to studies that might otherwise be difficult to publish due to publication bias.

Researchers may share their protocols and data on open science platforms that permit access to all. For example, *Perspectives on Behavior Science* allows authors to archive online material (Hantula, 2016a, 2016b), including data protocols and software. Recent issues have included shared software (Bullock, Fisher, & Hagopian, 2017; Kaplan, Gilroy & Hursh, 2018) and video (Deochand, Costello, & Deochand, 2018). Other examples include publications that share software and protocols on public access platforms such as GitHub (e.g., Gilroy, Franck, & Hantula, 2017; Gilroy, Kaplan, Reed, Koffarnus, & Hantula, 2018). Such transparency may contribute to a culture that prioritizes the availability of protocols, software, and data, bypassing contingencies of publication bias that may otherwise prevent the dissemination of such information. The social contingencies associated with openness, transparency, and prioritizing the advancement of knowledge may effectively compete with publication metrics and related consequences in ways that undermine publication bias and replication research.

Open sharing of research protocols and data may discourage ABA researchers from engaging in questionable research practices that enhance publishability. For example, researchers may omit entire datasets of participants when absence of effect compromises the appearance of otherwise robust experimental control. Researchers also may omit data from transition or acquisition phases where undesired variability is observed, or they may favorably depict data collected during multiple, back-to-back sessions on the same day as separate data points along the x -axis as if they represent data collected over a period of time greater than one day. In addition, researchers may use interobserver agreement calculation methods that artificially inflate agreement, or they may include irrelevant steps on procedural fidelity checklists to artificially inflate the percentage of intervention steps completed correctly. Any of these manipulations may improve the likelihood a study will be published if undetected. On the other hand, if research protocols are openly shared prior to data collection, researchers may be dissuaded from engaging in these manipulative practices post hoc.

Given that scientists exist largely within specialized research communities, efforts to facilitate open sharing of research protocols and datasets must start with the leadership of their respective scholarly organizations and journals. In the case of ABA researchers, this would necessarily include their flagship organization, the Association for Behavior Analysis International. Such efforts must consider the uniqueness and complexity of ABA protocols and data as a product of ABA experiments. For instance, ABA studies typically employ a small number of participants, and the risk of breaching anonymity and confidentiality may be greater in open data sharing of SCRD studies than group

comparison studies. Although these risks are not fundamentally different than those associated with more traditional dissemination activities (e.g., conference and presentations and journal publications), they require special attention and consideration.

Include Gray Research in Systematic Reviews of ABA Studies

Interest in systematic reviews of SCRD research has increased in recent years (Jamshidi, et al., 2017). Systematic reviews are an important method for understanding the methodological rigor of the available research prior to calculating aggregate effects of an intervention across a group of studies (e.g., Maggin, Talbott, Van Acker, & Kumm, 2017). Because there is no consensus on which metrics researchers should employ to determine effect size (ES) estimates across SCRD studies, Kratochwill et al. (2013) have provided guidance on how researchers can select ES estimates, including employing multiple ES metrics for the same datasets. Systematic reviews also are useful for identifying characteristics of individuals who are more or less likely to benefit from a particular intervention strategy (i.e., boundaries; Ganz et al., 2011; Ledford et al., 2016). Two influential systematic reviews conducted by research synthesis organizations, the National Autism Center's National Standards Project Report (2015) and the National Professional Development Center on ASD's Evidence-Based Practices Report (Wong et al., 2015), identified various "evidence-based" practices for children and youth with ASD based entirely on the published experimental research literature, with a large majority of included studies having been SCRD.

Gray literature consists of unpublished studies, usually in the form of masters theses and doctoral dissertations. Systematic reviews of SCRD research do not always include gray literature in their analyses (National Autism Center, 2015; Tincani & De Mers, 2016; Wong et al., 2015). Gray SCRD studies may face barriers to publication if they do not demonstrate strong functional relations indicative of experimental control (Shadish et al., 2016). There is evidence to suggest that systematic reviews that do not incorporate gray literature may obtain inflated intervention effect sizes that likely are a function of the file drawer effect (Gage et al., 2017; Sham & Smith, 2014). We therefore believe it is critical that future systematic reviews of ABA studies include gray literature to provide the most conservative and complete estimation of effect sizes for ABA interventions. Gray literature should be included in systematic reviews and held to the same standards of methodological rigor as published research. Given that theses and dissertations may not be indexed in databases commonly searched for systematic reviews, researchers will likely need to employ specialized extraction procedures, including searching databases that index gray studies (e.g., Google Scholar, ProQuest Dissertations & Theses), and directly contact study authors for reprints when studies are not otherwise publicly available.

Conclusion

The ABA community is not immune from the pitfalls of publication bias commonly associated with group comparison research. Contingencies of reinforcement that contribute to the file drawer effect among researchers in all disciplines also operate on

applied behavior analysts. There is evidence that researcher bias may be endemic within our current scientific and publication practices, which negatively affect the scientific integrity of our field as a whole. We have offered modest and preliminary suggestions for researchers and the research community to guard against publication bias with the intention of stimulating the ABA community to adopt, improve, or expand upon them. Discounting or ignoring the problem likely will diminish the credibility of our findings among our scientific colleagues, negatively affect consumers of our services, and undermine public confidence in behavior analytic treatments. Conversely, addressing publication bias and related issues likely will advance scientific knowledge about behavior, improve consumer outcomes associated with behavioral treatments, and promote public confidence in the technology of behavior.

References

- 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.
- Barton, E. E., Reichow, B., Schnitz, A., Smith, I. C., & Sherlock, D. (2015). A systematic review of sensory-based treatments for children with disabilities. *Research in Developmental Disabilities, 37*, 64–80.
- Bloom, S. E., Iwata, B. A., Fritz, J. N., Roscoe, E. M., & Carreau, A. B. (2011). Classroom application of a trial-based functional analysis. *Journal of Applied Behavior Analysis, 44*, 19–31.
- Bullock, C. E., Fisher, W. W., & Hagopian, L. P. (2017). Description and validation of a computerized behavioral data program: “BDataPro.”. *The Behavior Analyst, 40*, 275–285. <https://doi.org/10.1007/s40614-016-0079-0>.
- Carr, E. G., & McDowell, J. J. (1980). Social control of self-injurious behavior of organic etiology. *Behavior Therapy, 11*, 402–409.
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*, 7–18. <https://doi.org/10.1037/0022-006X.66.1.7>.
- Cook, B. G., Collins, L. W., Cook, S. C., & Cook, L. (2016). A replication by any other name: A systematic review of replicative intervention studies. *Remedial & Special Education, 37*, 223–234.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.
- Cooper, L. J., Wacker, D. P., Sasso, G. M., Reimers, T. M., & Donn, L. K. (1990). Using parents as therapists to evaluate appropriate behavior of their children: Application to a tertiary diagnostic clinic. *Journal of Applied Behavior Analysis, 23*, 285–296.
- Council for Exceptional Children. (2014). *Council for Exceptional Children standards for evidence-based practices in special education*. Arlington, VA: Author.
- Cox, A. L., Gast, D. L., Luscre, D., & Ayres, K. M. (2009). The effects of weighted vests on appropriate in-seat behaviors of elementary-age students with autism and severe to profound intellectual disabilities. *Focus on Autism & Other Developmental Disabilities, 24*, 17–26.
- Davis, T. N., Dacus, S., Strickland, E., Copeland, D., Chan, J. M., Blendon, K., et al. (2013). The effects of a weighted vest on aggressive and self-injurious behavior in a child with autism. *Developmental Neurorehabilitation, 16*, 210–215.
- Denton, T. F., & Meindl, J. N. (2016). The effect of colored overlays on reading fluency in individuals with dyslexia. *Behavior Analysis in Practice, 9*, 191–198.
- Deochand, N., Costello, M. S., & Deochand, M. E. (2018). Behavioral research with planaria. *Perspectives on Behavior Science, 41*, 447–464. <https://doi.org/10.1007/s40614-018-00176-w>.
- Dixon, M. R., Reed, D. D., Smith, T., Belisle, J., & Jackson, R. E. (2015). Research rankings of behavior analytic graduate training programs and their faculty. *Behavior Analysis in Practice, 8*, 7–15.
- Doyen, S., Klein, O., Pichon, C. L., & Cleeremans, A. (2012). Behavioral priming: it's all in the mind, but whose mind? *PLoS ONE, 7*, e29081.
- Driessen, E., Hollon, S. D., Bockting, C. L., Cuijpers, P., & Turner, E. H. (2015). Does publication bias inflate the apparent efficacy of psychological treatment for major depressive disorder? A systematic review and meta-analysis of US National Institutes of Health-funded trials. *PLoS ONE, 10*, e0137864.

- Earp, B. D., & Trafimow, D. (2015). Replication, falsification, and the crisis of confidence in social psychology. *Frontiers in Psychology*, 6, 621.
- Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2009). Meta-analysis of early intensive behavioral intervention for children with autism. *Journal of Clinical Child & Adolescent Psychology*, 38, 439–450.
- Evanschitzky, H., & Armstrong, J. S. (2010). Replications of forecasting research. *International Journal of Forecasting*, 26, 4–8.
- Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. Cambridge, MA: B. F. Skinner Foundation.
- Fox, R. M., & Mulick, J. A. (Eds.). (2016). *Controversial therapies for autism and intellectual disabilities: Fad, fashion, and science in professional practice* (2nd ed.). New York, NY: Routledge.
- Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, 345, 1502–1505.
- Gage, N. A., Cook, B. G., & Reichow, B. (2017). Publication bias in special education meta-analyses. *Exceptional Children*, 83, 428–445.
- Ganz, J. B., Earles-Vollrath, T. L., Mason, R. A., Rispoli, M. J., Heath, A. K., & Parker, R. I. (2011). An aggregate study of single-case research involving aided AAC: Participant characteristics of individuals with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 5, 1500–1509.
- Gelman, A., & Fung, K. (2016). The power of the “Power Pose.” *Slate Magazine*. Retrieved January 20, 2019, from http://www.slate.com/articles/health_and_science/science/2016/01/amy_cuddy_s_power_pose_research_is_the_latest_example_of_scientific_overreach.html.
- Gilroy, S. P., Franck, C. T., & Hantula, D. A. (2017). The discounting model selector: Statistical software for delay discounting applications. *Journal of the Experimental Analysis of Behavior*, 107(3), 388–401. <https://doi.org/10.1002/jeab.257>.
- Gilroy, S. P., Kaplan, B. A., Reed, D. D., Koffarnus, M. N., & Hantula, D. A. (2018). The demand curve analyzer: Behavioral economic software for applied research. *Journal of the Experimental Analysis of Behavior*. <https://doi.org/10.1002/jeab.479>.
- Hales, A. H., Wesselmann, E. D., & Hilgard, J. (2018). Improving psychological science through transparency and openness: An overview. *Perspectives on Behavior Science*. <https://doi.org/10.1007/s40614-018-00186-8>.
- Hanley, G. P. (2017). Editor’s note. *Journal of Applied Behavior Analysis*, 50, 3–7.
- Hantula, D. A. (2016a). Editorial: A very special issue. *The Behavior Analyst*, 39, 1–5. <https://doi.org/10.1007/s40614-016-0066-5>.
- Hantula, D. A. (2016b). Expanding the scope: Beyond the familiar and beyond the page. *The Behavior Analyst*, 39, 189–196. <https://doi.org/10.1007/s40614-016-0078-1>.
- 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. *Exceptional Children*, 71, 165–179.
- Hubbard, R., & Armstrong, J. S. (1994). Replications and extensions in marketing: Rarely published but quite contrary. *International Journal of Research in Marketing*, 11, 233–248.
- Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine*, 2, e124.
- Ioannidis, J. P. (2012). Why science is not necessarily self-correcting. *Perspectives on Psychological Science*, 7, 645–654.
- Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1982). Toward a functional analysis of self-injury. *Analysis & Intervention in Developmental Disabilities*, 2, 3–20.
- Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis*, 27, 197–209.
- Jamshidi, L., Heyvaert, M., Declercq, L., Fernández-Castilla, B., Ferron, J. M., Moeyaert, M., et al. (2017). Methodological quality of meta-analyses of single-case experimental studies. *Research in Developmental Disabilities*. Advance online publication. <https://doi.org/10.1016/j.ridd.2017.12.016>.
- Jessel, J., Hanley, G. P., & Ghaemmaghami, M. (2016). Interview-informed synthesized contingency analyses: Thirty replications and reanalysis. *Journal of Applied Behavior Analysis*, 49, 576–595.
- Kaplan, B. A., Gilroy, S. P., Reed, D. D., Koffarnus, M. N., & Hursh, S. R. (2018). The R package beezdemand: Behavioral economic easy demand. *Perspectives on Behavior Science*. <https://doi.org/10.1007/s40614-018-00187-7>.
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality & Social Psychology Review*, 2, 196–217.
- Kittelman, A., Gion, C., Horner, R. H., Levin, J. R., & Kratochwill, T. R. (2018). Establishing journalistic standards for the publication of negative results. *Remedial & Special Education*, 39, 171–176.

- Koegel, R. L., & Koegel, L. K. (2006). *Pivotal response treatments for autism: Communication, social, and academic development*. Baltimore, MD: Paul H. Brookes.
- Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial & Special Education, 34*, 26–38.
- Kratochwill, T. R., Levin, J. R., & Homer, R. H. (2018). Negative results: Conceptual and methodological dimensions in single-case intervention research. *Remedial & Special Education, 39*, 67–76.
- Lane, J. (2010). Let's make science metrics more useful. *Nature, 464*, 488–489.
- Ledford, J. R., Barton, E. E., Hardy, J. K., Elam, K., Seabolt, J., Shanks, M., et al. (2016). What equivocal data from single case comparison studies reveal about evidence-based practices in early childhood special education. *Journal of Early Intervention, 38*, 79–91.
- Lilienfeld, S. O. (2017). Psychology's replication crisis and the grant culture: Righting the ship. *Perspectives on Psychological Science, 12*(4), 660–664. <https://doi.org/10.1177/1745691616687745>.
- Lonigan, C. J., Elbert, J. C., & Johnson, S. B. (1998). Empirically supported psychosocial interventions for children: An overview. *Journal of Clinical Child Psychology, 27*, 138–145. https://doi.org/10.1207/s15374424jccp2702_1.
- Losinski, M., Cook, K., Hirsch, S., & Sanders, S. (2017). The effects of deep pressure therapies and antecedent exercise on stereotypical behaviors of students with autism spectrum disorders. *Behavioral Disorders, 42*, 196–208.
- Maggin, D. M., Talbot, E., Van Acker, E. Y., & Kumm, S. (2017). Quality indicators for systematic reviews in behavioral disorders. *Behavioral Disorders, 42*, 52–64.
- Mahoney, M. J. (1977). Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy & Research, 1*, 161–175.
- Mahoney, M. J. (1994). *Scientist as subject: The psychological imperative*. Clinton Corners, NY: Percheron Press.
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher, 43*, 304–316.
- Mitchell, M., Leachman, M., & Masterson, K. (2016). *Funding down, tuition up: State Cuts to higher education threaten quality and affordability at public colleges*. Washington, DC: Center on Budget & Policy Priorities Retrieved January 20, 2019, from http://www.cbpp.org/research/state-budget-and-tax/funding-down-tuition-up#_ftn33.
- National Autism Center. (2015). *Findings and conclusions: National standards project, phase 2*. Randolph, MA: Author.
- Normand, M. P., & Knoll, M. L. (2006). The effects of a stimulus-stimulus pairing procedure on the unprompted vocalizations of a young child diagnosed with autism. *Analysis of Verbal Behavior, 22*, 81–85.
- Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*, 615–631.
- Open Science Collaboration. (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science, 7*, 657–660.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, aac4716.
- Pashler, H., & Harris, C. R. (2012). Is the replicability crisis overblown? Three arguments examined. *Perspectives on Psychological Science, 7*, 531–536.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*, 528–530.
- Perone, M. (2018). How I learned to stop worrying and love replication failures. *Perspectives on Behavior Science. Advance online publication*. <https://doi.org/10.1007/s40614-018-0153-x>.
- Pierce, W. D., & Cheyney, C. D. (2017). *Behavior analysis and learning: A biobehavioral approach*. New York, NY: Routledge.
- Quigley, S. P., Peterson, L., Frieder, J. E., & Peterson, S. (2011). Effects of a weighted vest on problem behaviors during functional analyses in children with pervasive developmental disorders. *Research in Autism Spectrum Disorders, 5*, 529–538.
- Reichow, B. (2012). Overview of meta-analyses on early intensive behavioral intervention for young children with autism spectrum disorders. *Journal of Autism & Developmental Disorders, 42*, 512–520.
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*, 638.
- Shillingsburg, M. A., Hollander, D. L., Yosick, R. N., Bowen, C., & Muskat, L. R. (2015). Stimulus-stimulus pairing to increase vocalizations in children with language delays: A review. *Analysis of Verbal Behavior, 3*, 215–235.

- Shadish, W. R., Zelinsky, N. A., Vevea, J. L., & Kratochwill, T. R. (2016). A survey of publication practices of single-case design researchers when treatments have small or large effects. *Journal of Applied Behavior Analysis, 49*, 656–673.
- Sham, E., & Smith, T. (2014). Publication bias in studies of an applied behavior-analytic intervention: An initial analysis. *Journal of Applied Behavior Analysis, 47*, 663–678.
- Sidman, M. (1960). *Tactics of scientific research*. Cambridge, MA: Cambridge Center for Behavioral Studies.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359–1366.
- Skinner, B. F. (1938). *The behavior of organisms: an experimental analysis*. Cambridge, MA: B. F. Skinner Foundation.
- Skinner, B. F. (1953). *Science and human behavior*. Cambridge, MA: B. F. Skinner Foundation.
- Slocum, T. A., Detrich, R., Wilczynski, S. M., Spencer, T. D., Lewis, T., & Wolfe, K. (2014). The evidence-based practice of applied behavior analysis. *The Behavior Analyst, 37*, 41–56.
- Tincani, M., & De Mers, M. (2016). Meta-analysis of single-case research design studies on instructional pacing. *Behavior Modification, 40*, 799–824.
- Tincani, M., & Travers, J. C. (2018). Publishing single-case experimental research studies that do not demonstrate experimental control. *Remedial & Special Education, 39*, 118–128.
- Travers, J. C., Tincani, M. J., & Lang, R. (2014). Facilitated communication denies people with disabilities their voice. *Research & Practice in Severe Disabilities, 39*, 195–202.
- UNESCO. (2018). Global open access portal. Retrieved January 20, 2019, from <http://www.unesco.org/new/en/communication-and-information/portals-and-platforms/goap/open-science-movement>
- Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., et al. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism & Developmental Disorders, 45*, 1951–1966.
- Yoon, S. Y., & Bennett, G. M. (2000). Effects of a stimulus-stimulus pairing procedure on conditioning vocal sounds as reinforcers. *Analysis of Verbal Behavior, 1*, 75–88.
- Zane, T., Davis, C., & Rosswurm, M. (2014). The cost of fad treatments in autism. *Journal of Early & Intensive Behavior Intervention, 5*, 44–51.