9-2022

Strengthening the Procedural Fidelity Research-to-Practice Loop in Animal Behavior

Tiffany Kodak
Marquette University, tiffany.kodak@marquette.edu

Samantha Bergmann
University of North Texas

Mindy Waite
University of Wisconsin-Milwaukee

Journal of the Experimental Analysis of Behavior, Vol. 118, No. 2 (2022, September): 215-236. DOI.

Strengthening the Procedural Fidelity Research-to-Practice Loop in Animal Behavior

Tiffany Kodak
Department of Psychology, Marquette University
Samantha Bergmann
Department of Behavior Analysis, University of North Texas
Mindy Waite
Department of Psychology, University of Wisconsin, Milwaukee

Abstract
Procedural fidelity is the extent to which components of an intervention are implemented as designed. Procedural fidelity is measured as a dependent variable and manipulated as an independent variable. In research and practice, procedural-fidelity data should be collected, monitored, and reported. Procedural fidelity as an independent variable has been investigated in humans using parametric analyses, and the current article summarizes some of the research conducted on the effects of procedural-fidelity errors in behavior-reduction and skill-acquisition interventions. Connections were drawn to applied animal researchers and the work of animal behavior practitioners to produce implications for practice with human and animal clients and

suggestions for future research. Further, there are multiple ways to measure procedural fidelity, and different conclusions can be drawn based on the measure and computation method. The current article describes procedural-fidelity measures that are most applicable to animal behavior researchers and professionals.

Procedural fidelity is the extent to which an intervention is consistently implemented as planned (also referred to as treatment integrity, treatment fidelity, procedural integrity, or treatment reliability; Gresham et al., **2000**). The goal of research is to ensure that changes in one or more dependent variables (DVs) are the result of manipulations of one or more independent variables (IVs; e.g., treatment components) and not extraneous variables. Therefore, because procedural fidelity can either be studied as an IV or controlled for to maximize the validity of examinations of other IVs, procedural fidelity can be described as a primary goal of scientific inquiry.

Procedural fidelity is also critical to any applied practice designed to produce meaningful improvements for human and animal clients, because interventions implemented incorrectly or inconsistently are unlikely to lead to therapeutic gains and will be limited in external validity. An assay to procedural fidelity is medication adherence, wherein medication is not taken or applied as directed. For example, owners of animals may not adhere to recommended antibiotic use for their pets, which may contribute to antibiotic-resistant infections and poor health outcomes for the animal (Wayne et al., **2011**). Although failure to implement behavioral interventions as designed may not lead to the same issues as poor medication adherence, procedural-fidelity errors can decrease or prevent the efficacy or efficiency of behavioral interventions to reduce problem behavior (e.g., St. Peter Pipkin et al., **2010**), increase compliance (e.g., Wilder et al., **2006**), and teach skills (e.g., Carroll et al., **2013**).

Procedural-fidelity Errors

Procedural-fidelity errors can be categorized into two types: errors of omission and errors of commission. Failing to implement one or more components of a protocol is an error of omission. For example, an animal behavior researcher may discover a pellet dispenser failed to deliver food pellets according to the programmed reinforcement schedule for the experiment (Pinkston et al., **2008**) or that owners did not provide programmed reinforcers to their dogs during behavioral interventions to reduce mouthing (Waite & Kodak, **2021**), human-directed aggression (Echterling-Savage et al., **2015**), or jumping (Pfaller-Sadovsky et al., **2019**). In comparison, adding components that are not part of a protocol or conducting components at incorrect times are errors of commission. For example, owners implementing an intervention to reduce mouthing in their dogs were asked to provide attention contingent on a prespecified alternative behavior; however, the most common error of comission occurred when owners provided attention (i.e., told their dogs how cute they were) in the absence of the alternative behavior (Waite & Kodak, **2021**). Errors of omission and commission can occur in isolation or in combination. Both types of procedural-fidelity errors can reduce the efficacy and efficiency of intervention (e.g., Bergmann et al., **2017**; Carroll et al., **2013**; St. Peter Pipkin et al., **2010**).

In a controlled examination of the effects of programmed interventions (IV) on behavior (DV), unplanned procedural-fidelity errors affect internal validity. Internal validity refers to the degree to which the IV has functional control over changes in the DV. Procedural fidelity measures provide data on the degree of internal validity during intervention implementation, and these data can influence the interpretation of intervention outcomes. Measuring and reporting procedural fidelity allow for an analysis of whether changes in behavior were due to the IV (high fidelity) or some other variable (low fidelity). In other words, measuring procedural fidelity during behavioral interventions allows for analysis of whether behavioral outcomes were produced by the intervention of interest (high fidelity) or a different, incorrectly implemented intervention (low fidelity).

Results of an experiment or intervention can be grouped into four types (see Table **1**). A true positive (Table **1**; first quadrant) allows a researcher or practitioner to determine that the IV (treatment protocol) produced the change in the DV (behavior); therefore, implementation of the IV should continue. A potential false positive (Table **1**; second quadrant) permits a tentative conclusion that the IV may have produced the change in the DV. The conclusion must be tentative because a change in the DV in the predicted direction does not guarantee that the IV was implemented with high fidelity. Rather, behavior change could be due to extraneous variables thereby creating a correlation, instead of a functional relation, between the IV and DV. With a false positive, the researcher or practitioner could be implementing unnecessary or ineffective behavior-change procedures, which could contribute to replication failures (Gresham, **2009**). A true negative (Table **1**; third and fourth quadrant) allows a researcher or practitioner to conclude that the IV did not produce a change in the DV. As a result, the researcher or practitioner should discontinue or modify the IV to reduce exposure to ineffective interventions (Standard 2.01; BACB, **2020**). A potential false negative (Table **1**; fourth quadrant) is possible if the IV would have been effective had it been implemented correctly. As a result, the researcher or practitioner may discontinue the IV and implement a more intrusive IV or modify the IV by adding unnecessary components (Gresham, **2009**). Given these scenarios and their potential effects on the interpretation of client outcomes, monitoring and reporting procedural fidelity remain critical issues in a science of behavior.

Table 1. *Determining Internal Validity Based on Observed Behavior Change and Monitoring and Reporting Procedural Fidelity*

		Procedural Fidelity Monitored and Reported	
		Yes	**No**
Observed Behavior Change	Yes	True positive: IV is likely effective	Potential false positive: IV may be effective
	No	True negative: IV is likely ineffective	True negative: IV may be ineffective Potential false negative: IV may be effective when implemented with high fidelity

Note. Gresham et al. (2000) and Gresham (2009) outlined these four scenarios and described their implications for applied behavior analysis.

Whereas the Behavior Analyst Certification Board requires behavior analysts to monitor procedural fidelity during service delivery (Standard 2.18 and 2.19; BACB, **2020**), certifying bodies regulating applied animal behavior professionals are not as direct in their requirements to monitor procedural fidelity. However, certifying bodies do require protocol efficacy. Specifically, the *Certification Council for Professional Dog Trainers (CCPDT)*, International Association of Animal Behavior Consultants (IAABC), and American College of Veterinary Behaviorists (ACVB) require certificants to select protocols which maximize treatment gains while using ethical interventions (*ACVB Position Statements*, **n.d.**; Least Intrusive, Minimally Aversive [LIMA] Effective Behavior Intervention Policy, **2019**; IAABC Professional Code of Ethics, **n.d.**). If procedural fidelity is not measured during service delivery, practitioners cannot determine whether a lack of treatment efficacy is due to an ineffective intervention or failure to deliver the intervention consistently and correctly. Thus, barriers to efficacious intervention cannot be adequately addressed without procedural-fidelity data.

The extant literature on procedural fidelity and intervention outcomes is primarily comprised of studies conducted with humans. Although there are several scholarly reviews on measurement and the effects of procedural fidelity on intervention outcomes, none of the reviews explicitly relate these findings to animal behavior research and training (Brand et al., **2019** for effects of reduced fidelity; Collier-Meek, Sanetti, & Fallon, **2021** for measurement; Falakfarsa et al., **2021** for reporting). Nevertheless, the behavioral principles and procedures described in these reviews of research with humans are applicable across species and practices. For example, basic research shows the frequency of behavior can be increased using reinforcement procedures, such as contingent delivery of food, and decreased using punishment procedures such as response cost, time out, and response-contingent shock, in humans and nonhuman animals (e.g., pigeons and rats; Crosbie et al., **1997**; Johnson et al., **1973**; Lattal et al., **1998**; Rawson & Leitenberg, **1973**). Therefore, the current discussion will link what is known about procedural fidelity measurement and outcomes with humans to recommendations for research and practice with animals.

Procedural Fidelity as an Independent Variable

In recent years, a burgeoning area of research manipulates the procedural fidelity of the IV in experimental studies with humans to evaluate the effects on the DV (e.g., rate of problem behavior, correct responses). Manipulating procedural fidelity as an IV has both practical and conceptual value (Vollmer et al., **1999**). Practically, descriptive assessments show procedural-fidelity errors can and do occur in the application of interventions based on the principles of behavior analysis (e.g., Brand et al., **2018**; Breeman et al., **2020**; Carroll et al., **2013**; Donnelly & Karsten, **2017**; Foreman et al., **2021**; Kodak et al., **2018**). Therefore, it is important that researchers understand the implications of procedural-fidelity errors on behavior-analytic interventions to analyze intervention efficacy and understand potential replication failures. Conceptually, evaluating how behavior changes as a function of different degrees of procedural fidelity helps us understand the boundaries of behavior change. For example, one may compare the effects of distal and recent histories of reinforcement when procedural fidelity of differential reinforcement is manipulated (Vollmer et al., **1999**). Research that has investigated the effects of reduced procedural fidelity on behavior change includes studies on behavior reduction and skill acquisition with humans. The findings gleaned from these studies can inform research and application with animals, but replications and extensions conducted by animal researchers would contribute to a better understanding of the effects of procedural-fidelity errors across species and settings.

Behavior Reduction

Programmed procedural-fidelity errors during behavior-reduction procedures have included applied, basic, and translational studies. The studies include examining the effects of reinforcing problem behavior, omitting reinforcers following appropriate behavior, and applying punishment procedures intermittently. The first experimental manipulations of procedural fidelity were treatment challenges introduced during function-based interventions designed to reduce rates of problem behavior and increase rates of appropriate behavior (Vollmer et al., **1999**). Vollmer et al. (**1999**) compared the rates of problem behavior under full and partial implementation of differential reinforcement of alternative behavior (DRA) conditions with three adolescents and children with intellectual and developmental disabilities. The full-implementation condition included continuous reinforcement of the alternative behavior and extinction of the inappropriate behavior. The partial-implementation condition included withholding reinforcers following some instances of appropriate behavior (i.e., error of omission) and providing reinforcers following some instances of inappropriate behavior (i.e., error of commission, e.g., 25% of appropriate behavior reinforced, 75% of inappropriate behavior reinforced). In the partial-implementation condition, all three participants continued to allocate more responses to the appropriate behavior even when some instances of inappropriate behavior produced reinforcers. However, as the schedule of reinforcement began to favor inappropriate behavior, the efficacy of the intervention waned.

Following the Vollmer et al. (**1999**) study, researchers investigated the effects of procedural-fidelity errors on the allocation of responses to appropriate- and inappropriate-behavior analogues in laboratory contexts. St. Peter Pipkin et al. (**2010**; Experiment 1) manipulated errors of omission and commission of reinforcers such that fidelity varied between 20% and 80% in a DRA procedure with adults of typical development (see Experiments 2 and 3 for investigations with participants with developmental disabilities). With omission errors only, overall response rates decreased as the schedule became leaner, but the participants continued to allocate most responses to the appropriate behavior. With commission errors only, participants continued to allocate most responses to the appropriate behavior until fidelity dropped below 40% upon which participants allocated more responses to the inappropriate behavior. Data suggested a potentially protective effect of previous exposure to high-fidelity implementation, as greater levels of problem behavior only occurred following baseline conditions and did not occur following 100% fidelity conditions. St. Peter et al. (**2016**) conducted a similar investigation of errors of omission in a response-cost procedure. Responding to the inappropriate-behavior analogue remained low until errors of omission increased above 40%. Once fidelity was 20%, the participants emitted higher rates of inappropriate behavior than appropriate behavior.

Parametric analyses have also been conducted with clinically relevant populations in relevant contexts. For example, Colón and Ahearn (**2019**) conducted a parametric analysis of omission errors (25%, 50%, 75%, 100% fidelity; Study 2) during response interruption and redirection with three teenagers diagnosed with autism spectrum disorder (ASD) and histories of vocal stereotypy. The results showed that fidelity at or above 50% reduced vocal stereotypy below baseline levels across participants. However, the 25% fidelity condition reduced vocal stereotypy only when 100% fidelity booster sessions were interspersed. Another parametric analysis conducted with participants in an authentic context involved omission errors during timeout from play implemented by classroom staff with five school-aged children with various diagnoses (e.g., ASD, seizure disorder, traumatic brain injury; Foreman et al., **2021**). Foreman et al. (**2021**) compared the rates of problem behavior under different levels of fidelity (i.e., 0% and 100% fidelity and the average fidelity from observations; e.g., 5%, 11%; Study 1). Compared to baseline levels, decreased rates of problem behavior occurred even when timeout was implemented with errors of omission, suggesting that punishment procedures can be effective with reduced fidelity.

From these parametric investigations on procedural-fidelity errors during differential reinforcement and punishment procedures, several outcomes emerged. In all studies, 100% fidelity was not necessary to reduce inappropriate behavior. That is, reductions in inappropriate behavior were possible with reduced fidelity. In some cases, very low fidelity levels maintained low levels of inappropriate behavior (e.g., 5%–11% in Foreman et al., **2021**; 40% in St. Peter Pipkin et al., **2010**). It is important to note that the investigations included first exposing participants to high-fidelity conditions before conducting the parametric analyses (Foreman et al., **2021**; St. Peter Pipkin et al., **2010**; St. Peter et al., **2016**; Vollmer et al., **1999**). The effects of this order of conditions needs to be investigated further. It is possible that first experiencing high-fidelity conditions may provide a potentially protective effect when subjects are subsequently exposed to reduced fidelity.

In parametric analyses of procedural-fidelity errors, reductions in inappropriate behavior maintain until fidelity drops to 20%–40% (e.g., Colón & Ahearn, **2019**; St. Peter et al., **2016**; St. Peter Pipkin et al., **2010**). At this point, the reinforcement schedules begin to favor inappropriate behavior relative to appropriate behavior. This pattern of responding may be explained by concurrent schedules of reinforcement and the matching law (Herrnstein, **1974**; Reed & Kaplan, **2011**). That is, if reinforcement favors the appropriate behavior despite errors of omission and commission, then appropriate behavior is more likely under those conditions. However, when the probability of reinforcement shifts toward the inappropriate behavior, then the appropriate behavior becomes less likely. Even with an understanding that relative reinforcement rates and concurrent schedules of reinforcement affect probability of inappropriate behavior when procedural-fidelity errors occur, more research attention should be paid to the levels of fidelity that will maintain low rates of inappropriate behavior across species, interventions, contexts, and time.

One must consider reinforcement schedules more generally when comparing error types and their effects on inappropriate behavior and appropriate behavior in differential reinforcement procedures. Reduced procedural fidelity is likely to create intermittent reinforcement schedules, which have characteristic effects on behavior (Ferster & Skinner, **1957**). Vollmer et al. (**2008**) pointed out that even very high levels of fidelity can be problematic depending on the type of errors and schedule. The authors provided an example wherein implementers correctly provide extinction following problem behavior on roughly 19 out of 20 occurrences but engage in an error of commission (i.e., reinforce problem behavior) following one occurrence. Procedural fidelity would be 95%, but problem behavior would be reinforced on a VR20, which is likely to produce moderate-to-high rates of problem behavior with little pausing (Vollmer et al., **2008**). Another variable reinforcement schedule can be created if errors of omission result in some reinforcers being withheld following appropriate behaviors. Errors of omission could lead to moderate-to-high rates of appropriate behavior. However, if errors of omission are so frequent that the conditions are more akin to extinction, then decreasing rates of appropriate behavior would be probable.

Errors of omission in punishment-based procedures lead to intermittent punishment rather than continuous punishment. Continuous punishment is recommended when using punishment-based procedures (Miltenberger, **2016**). Therefore, intermittent punishment may be especially problematic if the inappropriate behavior continues to contact reinforcement (i.e., extinction is not possible, feasible, nor consistently implemented; Azrin et al., **1963**). With intermittent punishment, an immediate decrease in inappropriate behavior is likely to occur, but the response rate will recover if punishment does not follow subsequent responses (Azrin et al., **1963**). This was not observed in the study on timeout by Foreman et al. (**2021**); however, all participants had histories of intermittent punishment prior to the study. Extending the procedures of Foreman et al. to animals would allow for further examination of the effects of intermittent punishment on behavior reduction. In addition to procedural-fidelity errors, it is important to consider how the schedule, intensity of the punishers, and concurrent motivating operations can impact responding under intermittent punishment (Azrin et al., **1963**). Animal researchers could further explore procedural-fidelity errors in behavior-reduction procedures. Conducting this research in well-controlled laboratory studies is necessary to prevent humans and animals with problem behavior from exposure to potentially ineffective punishment-based procedures that may worsen their behavior.

Suggestions for Animal Research on Behavior Reduction

Animal behavior researchers study and intervene upon problematic behavior that occurs in companion, working, and captive animals. For example, researchers have investigated function-based interventions to reduce problem behavior in shelter dogs (Winslow et al., **2018**), canine mouthing (Waite & Kodak, **2021**), and self-injurious behavior in captive animals (Dorey et al., **2012**; Morris & Slocum, **2019**). However, many studies conducted with animals either do not include data on procedural fidelity or have high levels of reported fidelity in procedures conducted by researchers (Pfaller-Sadovsky et al., **2019**). Thus, the effects of procedural-fidelity errors on intervention for problem behavior in animals remains unknown.

One error that may occur in animal and human interventions for problem behavior is inconsistent or nonexistent use of extinction. Failing to implement extinction for problem behavior is an error of commission because problem behavior produces reinforcers, such as when an owner continues to provide attention (St. Peter Pipkin et al., **2010**).

In addition, the types of errors made by animal owners and/or caretakers are unknown. Examinations of the types of fidelity errors that occur in practice will help bridge the research-to-practice gap in procedural fidelity. Animal behavior professionals and owners are likely to make the same types of errors that behavior analysts and stakeholders make when working with humans. For example, an owner may continue to provide attention when their dog mouths on them rather than conducting programmed attention extinction during intervention, similar to a teacher providing a break from the task to a student who engages in aggressive behavior rather than continuing instruction (Kodak et al., **2018**). Identifying the frequency of types of fidelity errors in practice will permit programmed examinations of the effects of these errors on behavior-reduction interventions and methods to mitigate these errors. For example, researchers seeking to evaluate ways to reduce the impact of failing to provide extinction in behavioral intervention could manipulate parameters of reinforcement by providing higher quality, immediate, and dense schedules of reinforcers for appropriate behavior, whereas naturally occurring reinforcers could be provided for problem behavior.

Based on research conducted with humans, it is expected that some procedural-fidelity errors can occur while maintaining the efficacy of behavior-reduction intervention. However, it is possible that rule-governed behavior may alter the effects of fidelity errors with humans in ways that will not occur with animals. For example, humans may engage in verbal behavior regarding hypothesized contingencies that results in behavior that does not conform to programmed contingencies in studies (e.g., Shimoff et al., **1986**), whereas the behavior of animals is likely shaped by programmed contingencies. Therefore, applied animal researchers should replicate studies examining procedural-fidelity errors to investigate the effects of fidelity errors on behavior-reduction interventions for animals. For example, animal researchers could replicate the procedures of St. Peter Pipkin et al. (**2010**; Experiment 1) by manipulating errors of omission and commission of reinforcers during a DRA procedure to examine whether animals show similar changes in behavior as the fidelity of intervention decreases.

Once the frequency and types of fidelity errors that occur in practice are identified, applied animal researchers could replicate methods by Colón and Ahearn (**2019**) during behavior-reduction interventions with animals. Specifically, examining the protective effects of 100% fidelity booster sessions on intervention delivered with reduced levels of fidelity may provide beneficial outcomes that can be useful to animal behavior professionals who work with owners. For example, the consultation model used by some animal behavior professionals working with owners involves a brief initial consultation and intervention development followed by subsequent check-ins with owners. If booster sessions with 100% fidelity offer a protective factor to the efficacy of behavior-reduction interventions, animal behavior professionals could implement the intervention briefly and with 100% fidelity during check-ins to potentially enhance the efficacy of interventions even if the owner's procedural fidelity is lower. However, the frequency of high-fidelity booster sessions needed to enhance intervention efficacy for animals requires further examination to assist in the development of guidelines for practice.

Skill Acquisition

Studies on programmed procedural-fidelity errors during skill acquisition have manipulated instructions, prompts, error correction, and reinforcer delivery—the latter garnering the most attention. Several studies examined the effects of omitting prompts on skill acquisition. For example, Noell et al. (**2002**) omitted prompts provided during math instruction implemented with 33%, 67%, and 100% fidelity with six school-aged children in need of additional instruction in math. The fewest correct responses occurred in the 33% fidelity condition, and responses were more variable in the 67% fidelity condition, which suggested that (a) errors may affect some learners more than others and (b) behavior becomes less predictable with intermediate levels of procedural fidelity. Another study that evaluated omitted prompts was conducted by Wilder et al. (**2006**). Wilder et al. omitted three-step prompting (i.e., repeat vocal instruction, model, and physically guide) within a program to teach compliance with vocal instructions with two 4-year-old children of typical development. For both participants, near-zero levels of compliance were observed in the 0% fidelity condition, compliance was observed on approximately half of the trials in the 50% fidelity condition, and consistently high levels of compliance occurred in the 100% fidelity condition.

Other studies manipulated prompts along with additional components of skill-acquisition procedures. For example, Carroll et al. (**2013**) programmed procedural-fidelity errors during the delivery of vocal instructions, prompts, and reinforcers for 33% fidelity in discrete trial instruction (DTI) with six children with ASD. Compared to high-fidelity DTI, all participants emitted lower levels of correct responses in the reduced-fidelity condition, and five of the six participants did not reach mastery-level responding. Carroll et al. also analyzed the effects of errors in delivery of vocal instructions, prompts, and reinforcers in separate conditions with 33% fidelity in an adapted alternating-treatments design. The effects of isolated errors differed across participants, but the lowest levels of correct responses occurred in the condition with omissions of prompts. A similar analysis conducted by Breeman et al. (**2020**) implemented errors in providing praise, delivering tangible items, and using an error-correction procedure for 75% fidelity in DTI with two children with ASD, and the results replicated Carroll et al. Participants acquired the skills more quickly with high-fidelity instruction than reduced-fidelity instruction.

Another analysis that included errors in prompting and other instructional components was conducted Donnelly and Karsten (**2017**). The experimenters manipulated the fidelity with which self-care behavior chains were implemented with three adolescent males with ASD. The experimenters programmed a combination of errors (i.e., prompting steps out of order, delivering the reinforcer at times other than immediately following correct completion of the target step, and omitting prompts to complete all steps within the chain) to occur on three of five teaching trials (i.e., 40% of trials implemented with fidelity; all three errors occurred during a programmed-error trial). With reduced fidelity, none of the participants emitted mastery-level responding, and the programmed procedural-fidelity errors prevented acquisition of the behavior chains. Next, the experimenters evaluated the effects of isolated errors, and participants did not emit mastery-level responding. Mastery-level responding returned with high-fidelity instruction, but correct responding decreased again when the experimenters reintroduced treatment-fidelity errors. This effect was replicated within and across participants suggesting that previous exposure to high-fidelity instruction did not provide a protective effect.

Another area of focus in parametric analyses of procedural-fidelity errors is reinforcer delivery during conditional-discrimination training with typically developing children and adults. Several studies manipulated both errors of omission and commission in the delivery of feedback or putative reinforcers in arbitrary match-to-sample (AMTS) tasks conducted with typically developing adults in a human operant arrangement (e.g., Bergmann et al., **2021**; Bottini et al., **2020**; Brand et al., **2020**; Hirst et al., **2013**; Hirst & DiGennaro Reed, **2015**). For example, Hirst et al. (**2013**) and Hirst and DiGennaro Reed (**2015**) manipulated the accuracy of statements provided following participant responses in a computerized AMTS task. Participants were randomly assigned to one of four fidelity conditions (i.e., 100%, 75%, 50%, and 25%). In reduced-fidelity conditions, a proportion of participants' responses contacted inaccurate feedback statements. If the participant's response was correct, they were told they were incorrect. On the other hand, if the participant's response was incorrect, they were told they were correct. Results demonstrated a functional relation between feedback accuracy and acquisition. Participants in the 50% and 25% fidelity conditions did not acquire the conditional discriminations to mastery, and deficits continued when the participants were subsequently exposed to 100% fidelity (i.e., carryover effects). Some participants in the 75% fidelity condition mastered the conditional discriminations but required more trials to criterion compared to participants in the 100% fidelity condition.

The findings of Hirst et al. (**2013**) and Hirst and DiGennaro Reed (**2015**) were replicated by Brand et al. (**2020**) and Bergmann et al. (**2021**). Bergmann et al. expanded this line of research by including reduced-fidelity levels not previously investigated (e.g., 85%, 90%, 95% fidelity). The results of Bergmann et al. showed acquisition with 85% fidelity was not significantly different from 100% fidelity, which supported previous findings that perfect fidelity may not be necessary to produce positive outcomes for clients. Most of the parametric analyses of reinforcer errors programmed both errors of omission and commission (Brand et al., **2019**). In a departure, Bergmann et al. (**2017**) compared the effects of errors of omission (i.e., omitting praise and a token after a correct response) and errors of commission (i.e., providing praise and a token after an incorrect response) when presented in isolation on acquisition of AMTS by two typically developing children. In Experiment 1, instruction with errors of omission or commission that occurred during 17% or 18% of trials resulted in double the number of sessions to mastery in comparison to high-fidelity instruction. In Experiment 2, both participants learned targets taught with high-fidelity instruction, whereas errors of omission during 20% of trials and errors of commission during 30% of trials hindered acquisition for each participant, respectively.

The results of programmed procedural-fidelity errors in skill acquisition have revealed several findings. Notably, smaller reductions in procedural fidelity have greater detrimental effects on target behaviors in skill-acquisition programs compared to inappropriate behavior in behavior-reduction programs. That is, the effectiveness and efficiency of skill-acquisition interventions are negatively impacted by smaller proportions of errors than differential reinforcement and punishment procedures explored in parametric analyses of procedural-fidelity errors (e.g., 75% fidelity in skill acquisition in Hirst & DiGennaro Reed, **2015**, versus 40% fidelity in DRA in St. Peter Pipkin et al. **2010**). However, additional research on degrees of reduced procedural fidelity needs to be conducted with humans and animals.

Several of the parametric analyses of procedural-fidelity errors in skill acquisition have been conducted with conditional discrimination training. When errors of commission occur in discrimination training, reinforcement of an incorrect response creates a relation other than the one arranged by the practitioner (Bergmann et al., **2021**; Davison & McCarthy, **1980**). If the participant receives a reinforcer after selecting the picture of the bird in the presence of the word “mouse,” then the participant is likely to continue selecting the picture of the bird given the word “mouse.” With repeated exposure to this contingency, the participant is likely to learn an inaccurate relation. Another example is if an animal behavior practitioner accidentally reinforces a dog's bark following the command “sit,” then the word “sit” could become discriminative for barking rather than sitting. In addition, there is emerging evidence that combinations of errors may be more detrimental than isolated errors (Bottini et al., **2020**; Carroll et al., **2013**). Researchers should continue investigating the prevalence of errors in behavior-analytic interventions for humans and nonhumans and the effects of different degrees of combined and isolated errors with different treatment components.

Suggestions for Animal Research on Skill Acquisition

Skill acquisition is likely a primary focus of training by animal behavior professionals and owners. In fact, 83% of surveyed pet owners reported purchasing treats for training or rewarding their dog (Morelli et al., **2020**). These skill-acquisition practices likely include the provision of prompts (e.g., targets, visual discriminative stimuli) and reinforcers, and research with humans shows fidelity errors in these intervention components is likely to reduce the efficacy of intervention (e.g., Carroll et al., **2013**). However, the effects of these errors on acquisition by animals remains an area in need of additional research. Studies could examine the isolated and combined effects of prompting and reinforcement errors during skill acquisition with animals, similar to methods used by Carroll et al. (**2013**). For example, animal researchers could collect data on the prevalence of prompting and reinforcement errors in animal behavior practices. Once the prevalence of these errors is known, researchers could compare combined errors in prompts and reinforcers at levels of fidelity in practice to instruction with high fidelity. Results of these comparisons can assist animal behavior practitioners in designing effective methods to address common errors in practice.

Animal researchers could also compare the effects of isolated errors (e.g., compare low-fidelity prompts to low-fidelity reinforcer delivery) on skill acquisition to identify types of fidelity errors that have the greatest impact on skill acquisition in animals. Research on the impact of isolated errors can help inform practice on specific components of intervention that must be trained to a high level of fidelity to produce skill acquisition. In addition, the outcomes of this research could be used to prevent the allocation of limited training time to low-impact intervention components that are less likely to negatively impact intervention outcomes if conducted with lower fidelity. For example, researchers could evaluate the impact of isolated errors of omission or commission on the acquisition of stationing behavior, which has been included in interventions to teach dogs to remain still for MRI scans (Strassberg et al., **2019**), chimpanzees to receive acupuncture for osteoarthritis (Magden et al., **2013**), and manatees to have their bodies measured (Colbert-Luke et al., **2001**).

Fidelity errors that occur in enrichment programs for captive animals is an important area of study. Many enrichment programs for captive animals are designed to provide opportunities to engage in species-specific behavior and are used to enhance animal well-being. Nevertheless, enrichment programs are unlikely to achieve the same improvements in psychological and health outcomes for animals if fidelity errors occur in their design and use. Programs that are not designed to match species-specific behavior or are disrupted by other practices will reduce their efficacy and could increase the likelihood of problem behavior. For example, placing scents on the edges of African wild dog enclosures can be used as behavioral enrichment and teach them about the boundaries of their enclosures, because scents are frequently used to mark the boundaries of territories in the wild (Jackson et al., **2012**). However, if scent markings are frequently removed during enclosure cleanings, this could cause distress to the animal. Further, if enrichment programs are not frequently changed to include novel stimuli (i.e., an error of omission in changing stimuli on some schedule), then the benefits of these programs will be reduced (Young et al., **2020**). Collaborations between animal researchers, animal behavior practitioners, and animal caretakers will be critical to conduct lines of research on the impact of fidelity errors on enrichment program outcomes and will help bridge the research-to-practice gap by producing outcomes that can be used to design effective enrichment programs for captive animals.

General Findings and Recommendations for Future Research

Results from parametric analyses of programmed procedural-fidelity errors in behavior-reduction and skill-acquisition procedures have several key takeaways to inform future research and practice (Brand et al., **2019**). High levels of procedural fidelity are most likely to lead to the best outcomes for clients, although fidelity does not need to be perfect to produce beneficial outcomes (e.g., Bergmann et al., **2021**; St. Peter Pipkin et al., **2010**). This is a promising finding considering that conditions in the environment may make perfect fidelity improbable (e.g., competition for resources, limited supervision; Vollmer et al., **2008**). However, there may be a difference between levels of fidelity that still result in beneficial outcomes for clients depending on whether the intervention is designed to reduce behavior or build skills. That is, problem behavior remained low when programmed fidelity levels were as low as 40% (e.g., St. Peter Pipkin et al., **2010**), but failure to acquire or lengthy acquisition periods were observed when fidelity was around 80% in skill-acquisition studies (e.g., Bergmann et al., **2021**; Brand et al., **2020**; Hirst & DiGennaro Reed, **2015**).

Another finding is that reduced fidelity can produce more variable patterns of responding across procedures and participants. Variability impairs a researcher or practitioner's ability to predict behavior. Results also support the possibility that exposure to high levels of procedural fidelity at the outset of intervention or interspersed with reduced-fidelity sessions can have protective effects for some (e.g., Colón & Ahearn, **2019**, St. Peter Pipkin et al., **2010**; Vollmer et al, **1999**); however, the effects are idiosyncratic (e.g., Donnelly & Karsten, **2017**). Therefore, training and supervision for clients implementing behavioral interventions with animals needs to be ongoing to promote continued high levels of fidelity throughout treatment. Research needs to investigate the effects of long-term exposure to poor fidelity on animal behavior (Brand et al., **2019**). Further, it is possible that caregivers performing interventions are likely to make long-term fidelity errors without support and assistance from behavior professionals. For example, an owner may fail to deliver reinforcement to their dog following the occurrence of an appropriate behavior during home-based training (e.g., Waite & Kodak, **2021**), similar to a teacher omitting praise and a reinforcer following a student's correct response on an academic task (Kodak et al., **2018**). These errors of omission may hinder the dog's or student's acquisition of the targeted skill, respectively, and may be unlikely to change without feedback from professionals.

Examinations of the short- and long-term effects of common procedural-fidelity errors in applied animal research and practice will help guide further investigation of the effects of these fidelity errors on intervention outcomes and methods to prevent these errors. Collaboration between applied animal researchers and animal behavior practitioners will enable data collection on common procedural-fidelity errors. For example, volunteers at animal shelters are responsible for implementing protocols such as dog walking and enrichment practices, and procedural-fidelity data indicate that they only implement approximately half of the protocol steps with fidelity in the absence of preservice training (Howard & Reed, **2015**; Jamison, **2003**). Once common errors are identified, applied animal researchers can examine the short- and long-term effects of these errors on intervention outcomes for animals. Further, applied animal researchers could evaluate feasible and low-cost training methods (e.g., written instructions, Howard & DiGennaro-Reed, **2014**; video-based training, Howard & Reed, **2015**) to improve procedural fidelity and compare the outcomes of those methods to more intensive training (e.g., individualized training and feedback from an animal behavior professional).

Measuring Procedural Fidelity

The IV in applied research and practice is conducted by humans. Therefore, measuring whether humans implement the IV accurately and consistently is both possible and crucial for research and practice (Collier-Meek, Sanetti et al., **2021**; Colllier-Meek, Sanetti, & Fallon, **2021**). Whereas procedural-fidelity data reported in research on humans tend to be very high (McIntyre et al., **2007**), descriptive assessments suggest practitioners and animal behavior professionals are likely to make errors when implementing behavior-analytic interventions (Brand et al., **2018**; Breeman et al., **2020**; Carroll et al., **2013**; Colón & Ahearn, **2019**; Donnelly & Karsten, **2017**; Foreman et al., **2021**; Kodak et al., **2018**; Pfaller-Sadovsky et al., **2019**). As such, behavioral interventions implemented incorrectly or inconsistently in practice may not lead to outcomes similar to published studies.

Some may argue that collecting and reporting procedural-fidelity data are unnecessary for a variety of reasons. One of these reasons is the assumption that a skilled professional will detect issues with the implementation of the IV if unexpected changes in trend, variability, or level in the DV are observed (Gresham et al., **2000**; Peterson et al., **1982**). However, identifying the reason for changes in the DV can be difficult if researchers and practitioners do not collect procedural-fidelity data. Another assumption is that if the DV remains in a steady state or behavior changes in the expected direction, then the IV must be implemented to an adequate degree (Gresham et al., **2000**). However, this rationale is not foolproof; threats to internal validity (e.g., extraneous variables) could be responsible for the change in the DV. For example, in an intervention designed to reduce separation-related problem behavior in horses, the owner may need to remove the horse's companion animal to contrive an establishing operation (Mal et al., **1991**). However, the owner may suddenly stop removing a horse's companion at the onset of intervention. If this occurs, the horse's challenging behavior may decrease due to a change in this uncontrolled variable rather than the effects of intervention. Moreover, even skilled professionals make fidelity errors (Breeman et al., **2020**; Pfaller-Sadovsky et al., **2019**). Without measuring the implementation of the IV, potential confounding variables will go undetected (Gresham et al., **2000**). This is problematic for the experiment; however, the gravest concerns are likely to emerge only when others attempt to replicate the procedures in research and practice. For replication to be possible, a researcher or practitioner must know “precisely what was done, how it was done, and how long it was done” (Gresham et al., **2000**, p. 200). Data showing the IV was implemented as prescribed increases the likelihood that the effects of a behavior-change procedure will generalize across individuals, settings, and behaviors (Baer et al., **1968**).

Procedural-fidelity errors in research and practice can hinder or prevent an improvement in behavior. The types of error made by animal researchers, animal behavior professionals, and owners remain unknown due to limited research on procedural fidelity during animal research and practice, but it is expected that their errors are similar to those made by practitioners and caregivers delivering intervention to humans. For example, an owner may provide reinforcement when the dog has not yet engaged in the targeted skill (Waite & Kodak, **2021**), similar to a teacher providing praise when the student has not yet completed an academic task. These errors of omission may reduce the dog's and student's engagement in the targeted skills, respectively. Due to the negative impact of procedural-fidelity errors on research and practice with animals, measurement of procedural fidelity in animal research and practice is an important component of intervention.

Procedural Fidelity Measurement Methods

The measurement system and data analysis method used to collect and analyze procedural-fidelity data can have important effects on the interpretation of the fidelity of intervention. Determining the purpose of the measure is necessary to select which procedural-fidelity measurement system and analysis will produce useful data. Procedural-fidelity data can be used to gauge overall fidelity or pinpoint specific errors. An overall measure of all components of intervention (i.e., global rating; Cook et al., **2015**) provides a general indication of how well intervention has been implemented and a lens through which to interpret the outcomes of intervention. Global measures of procedural fidelity may be particularly useful in practice. For example, animal behavior professionals typically teach owners to implement interventions that will occur in the professional's absence. If the owner reports that the animal's behavior has not improved, a global measure of procedural fidelity will assist in informing next steps for intervention. A determination to change the intervention would be made if procedural-fidelity data suggest the owner implemented intervention components at a level that should produce behavior change. In comparison, low levels of procedural fidelity suggest the owner requires additional training and supervision to obtain consistently high levels of accurate intervention implementation. In the absence of procedural-fidelity data, the likelihood of a successful decision-making process and treatment outcome is reduced.

Another purpose of measuring procedural fidelity is to pinpoint specific errors occurring during intervention. For example, a data collector can measure an experimenter's implementation of components of differential reinforcement of other behavior (DRO) to treat cat aggression (e.g., ceased petting once the cat met the DRO criteria, initiating petting within 3 s of the end of a break, not delivering enough pets so the cat's behavior could meet the DRO criteria; Fritz et al., **2022**). If the data identify specific errors that occur frequently, the experimenter can receive retraining on select components of the treatment protocol. Targeted retraining may reduce its duration by eliminating unnecessary practice of treatment components that are implemented correctly and consistently.

Procedural fidelity data-collection measures developed to identify overall levels of fidelity (global measures), pinpoint specific errors (component measures), or evaluate the effects of errors on outcomes must be designed carefully. Procedural fidelity measures produce different interpretations of fidelity, depending on how data are collected and analyzed. To illustrate these differences, consider an applied animal researcher studying horse target training. The researcher might program a session with five trials of target training, and each trial contains four instructional components measured by the researcher (i.e., presenting the target at the programmed distance from the horse's muzzle, saying “target,” providing a click immediately following the touch response, and providing an edible reinforcer immediately after the click, see Table **2**). If the researcher makes an error on one instructional component per trial (i.e., the experimenter fails to provide the edible reinforcer after the click) the interpretation of the researcher's procedural fidelity will vary depending on the data collection measure and analysis. Refer to Table **2** and Supporting Information (Fig. **1**) for examples of differences in obtained outcomes based on the measurement system used for procedural fidelity.

Table 2. *Comparison of Outcomes of Procedural Fidelity Measures for Horse Target Training*

Trial	Horse Behavior	Implementer Behavior	Global Scoring	All or Nothing by Trial	Component Scoring			
					Target	S^D	Click	SR+
1	Correct	✓ Target	3/4	0	1	1	1	0
		✓ S^D						
		✓ Click						
		× SR+						
2	Correct	✓ Target	3/4	0	1	1	1	0
		✓ S^D						
		✓ Click						
		× SR+						
3	Correct	× Target	2/4	0	0	1	1	0
		✓ S^D						
		✓ Click						
		× SR+						
4	Correct	✓ Target	3/4	0	1	1	1	0
		✓ S^D						
		✓ Click						
		× SR+						
5	Correct	✓ Target	3/4	0	1	1	1	0
		✓ S^D						
		✓ Click						
		× SR+						
Total			**14/20**	**0/5**	**4/5**	**5/5**	**5/5**	**0/5**
			70%	**0%**	**80%**	**100%**	**100%**	**0%**

Note. Procedural fidelity scores for a target training session with a horse across different procedural fidelity measures. Numbers indicate the scores across each trial and then the total score for the session. Target = target on stick placed in correct location; S^D = discriminative stimulus “Target”; click = click delivered; SR+ = edible/positive reinforcer delivered.

If the applied animal behavior researcher analyzes global procedural fidelity by calculating the number of instructional components implemented correctly during the teaching session, the researcher will sum the number of correctly implemented components and divide by the total number of possible components (Table **2**, fourth column; Supporting Information Fig. **1**). For example, the researcher will sum up to four components per trial for five trials (i.e., maximum of 20 components implemented correctly) and divide by 20, which is the overall number of treatment components per teaching session. This proportion is multiplied by 100 to obtain a percentage of global procedural fidelity. In Table **2**, 14 components were implemented correctly for a global fidelity score of 70% (i.e., 14/20 = 0.70, multiplied by 100 = 70). A fidelity estimate of 70% may lead the researcher to retrain the implementer or collect procedural fidelity data for an additional session before retraining.

Global-fidelity scores provide a measure of overall fidelity, and they are frequently used when calculating caregivers' fidelity during interventions for human participants (e.g., Catania et al., **2009**; Lerman et al., **2008**). Global measures also may be used by animal behavior professionals to determine when owner training has resulted in a sufficient level of fidelity to discontinue training and instruct owners to implement intervention independently. Nevertheless, global measures can hinder identification of procedural-fidelity errors (Cook et al., **2015**). If the owner continuously makes errors to a critical component of intervention (e.g., failing to provide a reinforcer following target behavior), the overall global measures may still be relatively high. If only the global measure is reported, adequate fidelity may be assumed, and the owner would not receive retraining on the incorrectly implemented component.

In comparison, if the researcher uses a more stringent analysis of procedural fidelity, such as calculating all-or-nothing fidelity, a different outcome will be obtained (Table **2**, fifth column). All-or-nothing by trial fidelity is calculated by scoring a trial as implemented with fidelity (a score of 1) only when all components are conducted correctly. If the implementer makes an error on one or more components of the trial, then the trial is scored a 0. Applying all-or-nothing fidelity to the same five-trial instructional opportunity shown in Table **2** will require the researcher to score each of the five trials as a 0, because the implementer erred on one instructional component in each trial. Thus, after summing the scores for all trials, dividing that sum by 5 (for the number of total trials), and multiplying by 100 to obtain a percentage, procedural fidelity would be 0% for the instructional session. Based on a fidelity estimate of 0%, the researcher would likely conduct retraining to correct fidelity errors.

Researchers may analyze fidelity of the implementation of individual intervention components. In these analyses, each component of intervention is analyzed separately to determine whether specific components are implemented consistently and correctly (Table **2**, last columns, "component scoring"). Component scoring can be all-or-nothing or occurrence and nonoccurrence by individual component. For all-or-nothing by component, the implementer must conduct the component correctly on all opportunities for the component to be scored as a 1. If the implementer makes one or more error when conducting a component, then the component is scored a 0. In our example of the five-trial instructional session, the implementer erred when conducting the target and reinforcer components. These components would be scored 0; whereas the S^D and click components would be scored 1. Thus, after summing the components implemented correctly, dividing by 4 (i.e., the total number of components), and multiplying by 100 (i.e., 2/4 = 0.5 x 100 = 50), the procedural fidelity estimate for all-or-nothing by component would be 50%. A researcher would likely retrain the incorrect components based on this fidelity estimate.

The researcher can also analyze fidelity by scoring the occurrence and nonoccurrence of each individual component (Table **2**, last columns, "component scoring"). In our example of the five-trial instructional session, the implementer did not deliver the edible reinforcer following touch responses in any of the five trials. An analysis of the occurrence of individual components will require the researcher to sum the number of correct implementations of a component (e.g., providing the edible reinforcer), divide by the number of trials or opportunities to implement the component, and multiply the proportion by 100 to obtain a percentage. The researcher would score 0% fidelity for delivering an edible reinforcer because that component was implemented incorrectly on every trial, 80% fidelity for presenting the target, and 100% fidelity on all other individual intervention components. A researcher would likely retrain the implementer on the delivery of an edible reinforcer following target behavior.

All-or-nothing by trial/component fidelity and individual component fidelity offer more stringent methods of analysis compared to global scores. Unlike global scores, which can produce moderate or high levels of procedural fidelity when components are routinely implemented with errors, all-or-nothing and component scores will not yield high fidelity estimates. Using all-or-nothing by trial/component and individual component analyses will permit identification of the need for retraining and may prevent the continuation of poorly implemented intervention. Further, all-or-nothing and component fidelity could be used to establish mastery criteria for experimenter and owner training. All-or-nothing by component or individual component fidelity permits identification of specific components that would benefit from retraining, which could be more efficient than retraining all components of an intervention. Component fidelity is less frequently included as the primary dependent variable for training, despite the benefits of this measure (Cook et al., **2015**). In contrast, all-or-nothing fidelity is more frequently used as a measure of procedural fidelity in studies with humans (e.g., Halbur et al., **2021**; Kisamore et al., **2016**), particularly for studies that manipulate procedural fidelity as an IV (e.g., Bergmann et al., **2017**; Carroll et al., **2013**; **2016**).

Collecting Procedural Fidelity Data

Although different fidelity measurement methods may be more informative or valid than others (e.g., global vs. component measurements; Cook et al., **2015**), measures differ in relation to time/effort, training, staff, and technical skills required to collect and analyze data. The ability to collect component-fidelity scores requires considering the complexity of the intervention and resources needed for training, observation, and data collection. For example, if a dedicated observer has been highly trained to detect procedural-fidelity errors in an intervention, then collecting occurrence and nonoccurrence data for every component across every opportunity may be feasible. However, if an animal behavior professional is assessing their own implementation of an intervention while actively working with an animal, then another measure of procedural fidelity might be more appropriate.

Researchers have recently evaluated other fidelity measures that have been described as more feasible for use in practice. A Likert scale may be used by animal behavior researchers and professionals to measure overall fidelity following the implementation of intervention (see Supporting Information for an example of a rating scale). Suhrheinrich et al. (**2020**) compared the reliability of several measures of procedural fidelity. Five-point and 3-point Likert scales were evaluated, with each point on the scale anchored to specific behavioral definitions and percentages of accurate implementation (e.g., a score of 4 on the 5-point Likert scale indicated intervention components were implemented correctly 80% to 99% of the intervention session). The authors analyzed interobserver agreement for each Likert scale in comparison to trial-by-trial coding (i.e., occurrence or nonoccurrence and summed for global fidelity measure). Suhrheinrich et al. reported high exact agreement between the 5-point Likert scale and trial-by-trial coding for intervention components, whereas the 3-point Likert scale had low to moderate agreement with the 5-point Likert scale and trial-by-trial coding. The results suggest the 5-point Likert scale was sufficient for measuring procedural fidelity and may be a more feasible measure for use in practice once additional research validates the consistent and accurate use of the 5-point Likert scale in practice.

To maximize efficiency, Cook et al. (**2015**) suggested first assessing global fidelity scores as an overall indicator of procedural fidelity and requiring a stringent criterion (e.g., fidelity at 95%-100%) for adequate fidelity levels. However, this recommendation may not be feasible for many animal behavior professionals. As described above, calculating global scores requires collecting data on the occurrence and nonoccurrence of every component on every opportunity. This type of data collection may be arduous for many applied settings. In addition, global scores can hide repeated errors on specific components (Cook et al., **2015**). Therefore, it may be more feasible for animal behavior professionals to evaluate procedural fidelity by all-or-nothing by component with a checklist (see Supporting Information Fig. **2** for an example). A checklist can be created with rows of the required components. The observer or implementer (see self-monitoring below) would record whether each component was implemented perfectly (i.e., recording Yes) or had at least one error (i.e., recording No). The animal behavior professional sums the number of components implemented perfectly and calculates the fidelity estimate. These data can be used to determine whether specific components need to be retrained. Errors in components for prompting or reinforcing target behaviors may be especially impactful for animal behavior researchers and professionals focused on teaching skills, and these specific intervention components could be the focus of some component analyses (Carroll et al., **2013**). For example, the animal behavior professional could examine component fidelity for reinforcer delivery by dividing the number of intervention sessions with correct reinforcer delivery by the total number of intervention sessions and multiplying by 100 to obtain a percentage.

Furthermore, focusing on component measures can save time and money, which is critical when working directly with owners who typically pay for intervention out of pocket (Bisco & Fier, **2020**). However, from a research perspective, programming more stringent measures of fidelity is critical to testing new applied animal behavior interventions, especially with owners as experimenters, as these measures will allow for assessment of intervention efficacy and feasibility.

Self-monitoring

Whenever possible, an observer should collect data on an implementer's procedural fidelity (Weston et al., **2020**). However, some settings and conditions may not permit frequent observations, such as when an animal behavior professional is working with an owner and has a limited number of training sessions. As such, self-monitoring procedural fidelity, an emerging area in research and practice, may be a desirable option. Self-monitoring (also called self-recording, self-observations, and self-evaluation) involves a person observing their own behavior and recording the occurrence or nonoccurrence of a target behavior (Cooper et al., **2020**). Several studies in the behavior-analytic and behavioral-consultation literature have used self-monitoring to assess procedural fidelity. A review of self-monitoring integrity by teachers (Rispoli et al., **2017**) revealed that most studies included in-vivo monitoring rather than video observations, and teachers used procedural-fidelity checklists, rating scales, and event recording to collect data on their implementation (see Supporting Information for sample data sheets). Event recording using a variety of instruments (e.g., tallies on paper, clickers/counters) was the method used in most of the studies (i.e., 10 of 17 studies; Rispoli et al., **2017**). For example, Sanetti and Fallon (**2011**) evaluated whether teachers' implementation of the Good Behavior Game improved with self-monitoring. The data showed teachers' fidelity increased when self-monitoring was used compared to weekly meetings with a consultant who asked questions about fidelity.

Currently, some research supports the use of self-monitoring with videos. That is, implementers watch a video of themselves conducting the intervention and complete a direct-observation form, such as an individual-component checklist or occurrence/nonoccurrence for a global score. Belfiore et al. (**2008**) taught direct-line staff to collect data on their own implementation of a behavior intervention plan with a client with ASD. After teaching staff to use the checklist and requiring 90% correct data collection on one video, the staff monitored their own procedural fidelity from recorded videos. The self-monitoring intervention increased procedural fidelity of direct-line staff. Similarly, Pelletier et al. (**2010**) taught direct-line staff to use a checklist to monitor their fidelity from videos. The self-monitoring intervention increased the procedural fidelity of all staff implementation, and high levels of fidelity were observed in maintenance probes. However, there are variations in the degree to which self-monitoring increases procedural fidelity across participants, and additional feedback on performance may be necessary (e.g., Belfiore et al., **2008**; Bishop et al., **2015**; Mouzakitis et al., **2015**).

Although a variety of individuals (e.g., teachers, direct-line staff) have engaged in self-monitoring to assess their procedural fidelity, self-report of procedural fidelity is probably more commonly used by many animal behavior professionals to assess procedural fidelity of clients. Self-report is an indirect assessment wherein an animal behavior professional asks the owner how well they implemented the intervention on their own. Self-reports can be convenient, but data gathered via self-report should be interpreted with caution. Self-reports may not accurately reflect adherence as implementers are likely to rate their implementation as artificially high (Gresham et al., **2000**; Noell, **2014**). Data gathered from self-reports may show high agreement across time, but there is little empirical support for using self-report data as valid measures of procedural fidelity (Gresham, **2014**; Gresham et al., **2017**). In addition, self-reported procedural fidelity in interviews with a consultant did not improve teachers' procedural fidelity when implementing behavior intervention plans (Noell et al., **2005**); therefore, the utility of self-reported fidelity is questionable.

As an alternative to self-report, self-monitoring may be useful for caregivers or owners of animals because they will implement the majority of the protocol alone and fidelity may wane following training (Echterling-Savage et al., **2015**). With some support from family members or arranging the environment with a recording device and tripod, it may be feasible for many owners and caregivers to record themselves implementing procedures. The applied animal professional could teach the owner or caregiver how to complete the self-monitoring form (e.g., component checklist, occurrence/nonoccurrence data, rating scales; Belfiore et al., **2008**; Mouzakitis et al., **2015**; Pelletier et al., **2010**).

When using self-monitoring for procedural fidelity, animal behavior professionals should include strategies to increase the reliability of data. Cooper et al. (**2020**) provides guidelines and recommended procedures for using self-monitoring to change behavior (pp. 696–699). These recommendations include: (a) designing self-monitoring data collection systems that are easy to use, (b) providing supplementary sources of control (e.g., audio or tactile prompts to self-monitor), (c) focusing measurement on the most critical dimensions of the target behavior(s), (d) beginning self-monitoring early on and continuing self-monitoring on a frequent basis, and (e) reinforcing accurate self-monitoring (e.g., Martella et al., **1993**). However, interobserver agreement and feedback on the self-monitoring results may be necessary for longer-term maintenance (Mouzakitis et al., **2015**). For example, an applied animal researcher could design a self-monitoring system and teach owners to self-monitor their implementation. Thereafter, the researcher could send the owner a text to prompt daily collection of self-monitoring data, request the owner complete a shared online self-monitoring form, and submit data on the animal's behavior following intervention each day. The applied animal researcher could provide inexpensive tangible reinforcers (e.g., a bag of dog treats based on their pet's preference) to owners for consistent data collection and self-monitoring as well as provide feedback on and brief retraining for consistent errors in data collection or implementation (as indicated in the self-monitoring forms). Applied animal researchers could examine the reliability and validity of self-monitoring procedural fidelity measures across interventions and settings (laboratory, animal's home, training center), including schedules for self-monitoring observations, to maintain accurate data collection.

Timing of Data Collection on Procedural Fidelity

In research studies, procedural-fidelity data should be collected across all study phases (e.g., baseline, treatment, follow-up) to evaluate whether procedures were implemented as planned. Ongoing monitoring of procedural fidelity permits rapid detection of errors that can be corrected via retraining to promote true positives and negatives (Table **1**). If procedural-fidelity data are not monitored throughout the study, researchers may later identify that the study procedures were not implemented with sufficient fidelity to warrant publication. Thus, monitoring procedural fidelity throughout the study can prevent allocating extensive time and effort to collecting data that will ultimately never be published because of questionable internal validity.

In practice, it is most important to collect procedural-fidelity data during intervention, particularly in the early portion of intervention implementation. High procedural fidelity at the outset can produce faster skill acquisition (Bergmann et al., **2021**; Carroll et al., **2013**; DiGennaro Reed et al., **2011**; Jenkins et al., **2015**) and supports greater maintenance of behavior change even if fidelity reduces in later sessions (Echterling-Savage et al., **2015**; St Peter Pipkin et al., **2010**). For example, an animal behavior professional could monitor an owner's procedural fidelity for the first 2 days of intervention. If the owner implements intervention with high levels of fidelity (e.g., fidelity is at 90% or above), then the animal behavior professional could schedule less frequent fidelity checks as intervention progresses, subsequent fidelity checks show consistently high fidelity, and the animal's behavior is improving. In contrast, if frequent fidelity errors occur in the first few sessions of intervention, the animal behavior professional could immediately provide feedback to the owner to correct the errors or conduct behavioral skills training for specific intervention components to improve fidelity.

Careful consideration is necessary regarding how to fade fidelity checks over time. The behavioral consultation literature suggests teacher-delivered intervention may be at 100% fidelity immediately after training and in the first few days of intervention but decrease to low levels within 10 days after training (Noell et al., **1997**; Sanetti et al., **2007**; Witt et al., **1997**). Similarly, studies on dog owners suggest that procedural fidelity may not maintain at high levels (Echterling-Savage et al., **2015**). Noell et al. (**2005**) thinned performance feedback based on the teachers' procedural fidelity. In their rapid thinning procedure, feedback on procedural fidelity was provided following the first day of implementation and every subsequent day until fidelity was 100%. Then, the researchers monitored procedural fidelity and provided feedback every other day until fidelity was 100% for 2 days. After meeting this criterion, fidelity checks and performance feedback decreased to once per week. Thus, more frequent fidelity checks may be necessary for at least the first 10 days of intervention, followed by ongoing monitoring and criteria for fading fidelity checks contingent on stable and high levels of fidelity.

Suggestions for Future Research on Measurement

Whereas there have been multiple studies investigating the extent to which researchers collect and report procedural-fidelity data on human interventions (e.g., Falakfarsa et al., **2021**; McIntyre et al., **2007**; Peterson et al., **1982**), there is only one review which examined additional measures of procedural fidelity like the data-collection method (Collier-Meek et al., **2018**). Future reviews should describe not only *if* authors of behavior-analytic literature are collecting and reporting procedural fidelity data but *how* these data are collected and computed. Because procedural-fidelity methods and measures can produce different estimates of procedural fidelity, reviews of the literature and empirical studies that directly compare different methods with the same samples of behavior would help elucidate important variables to consider when assessing procedural fidelity and may produce valuable recommendations to guide research and practice. For example, applied animal researchers could collaborate with animal behavior professionals to compare the accuracy and feasibility of component fidelity measures to rating scales used in practice. Although rating scales will not provide the depth of analysis offered by collecting and analyzing procedural-fidelity data on each component of intervention, rating scales may be identified as more feasible and, thus, more likely to result in the collection of procedural-fidelity data by animal behavior professionals. In addition to whether procedural fidelity data collection methods are valid, reliable, and feasible in practice, one should also consider whether stakeholders find the measures and estimates socially acceptable. That is, stakeholders may prefer to collect and receive data from certain measures of procedural fidelity over others (e.g., rating scale versus occurrence/nonoccurrence data), which could impact the manner in which the feedback is received. If the stakeholder reacts poorly to the feedback provided, it is possible that the animal behavior professional's assessment of procedural fidelity may become less reliable (Matey et al., **2021**).

Generally, including more intervention components in the observation system will lead to a more representative estimate of procedural fidelity. However, there is not an agreed upon standard for the level of detail of a measurement system, the number of observations that must be conducted, nor the frequency of observations to best estimate the fidelity of intervention (Gresham, **2014**). A study that examined the number of observations needed to provide a reliable estimate suggested that at least four observations conducted over a 2-week period produced reliable estimates of an intervention implemented by teachers (Gresham et al., **2017**). Evaluating the number of observations needed to reliably estimate procedural fidelity in practice is highly applicable to applied animal researchers and animal behavior professionals who may work with owners to implement intervention outside of the laboratory. Thus, this area of research needs to be replicated and extended by applied animal researchers with many animal-behavior interventions and across settings.

Conclusions

The effects of procedural fidelity on intervention outcomes have been examined as both a DV and IV. Basic, translational, and applied research findings support functional relations between procedural fidelity and participant and client outcomes and maintenance of behavior change. The measurement system used to collect procedural-fidelity data can have important effects on interpretation of the fidelity of intervention. Determining the purpose and feasibility of the measure is necessary to select a procedural-fidelity measure that will produce useful data. The focus of the current article was on adherence to intervention components, which is important for quantifying procedural fidelity. However, there are other dimensions of procedural fidelity such as the quality or dosage of the intervention that should be the topic of further investigation (Sanetti & Collier-Meek, **2014**).

Studies that report procedural fidelity most often measure and report the researcher's or practitioner's accurate implementation of intervention. Existing descriptive assessments of interventions for humans suggest that procedural fidelity during real-world implementation is lower than in published research studies (e.g., Carroll et al., **2013**; Kodak et al., **2018**). This may be especially pertinent to animal behavior researchers and practitioners because behavioral interventions are frequently implemented by stakeholders (e.g., shelter volunteers, owners of companion animals, zookeepers) who require training to conduct interventions with fidelity. Additional studies on animal interventions conducted by stakeholders will be critical to determine the levels of procedural fidelity necessary for initial treatment effectiveness as well as long-term maintenance and generalization. Specifically, applied animal researchers should perform descriptive assessments and experimental analyses on procedural fidelity to identify the types, frequency, and intervention impacts of common fidelity errors.

Whereas research has begun to investigate the effects of procedural-fidelity errors on outcomes of behavior interventions to reduce problem behavior and build new skills in humans, the current knowledge about the effects of procedural fidelity on intervention outcomes for animals is limited. For example, sufficient levels of component fidelity for behavior-reduction and skill-acquisition programs to retain efficacy and efficiency of interventions remains an area in need of further investigation. Similarly, there is a paucity of research manipulating fidelity as an IV for applied animal behavior. However, there are substantial challenges to evaluating procedural fidelity as an IV when working with human and animal clients. The long-term effects of treatment-integrity errors on learning remain unknown, and clients in need of services have the right to effective intervention (BACB, **2020**). In addition, the consultation model used during service provision to owners of companion animals may not provide extensive time for training and direct observation of procedural fidelity, making it difficult to manipulate and maintain prespecified levels of procedural fidelity. As such, continuing lines of procedural-fidelity research in laboratory settings and with analogue preparations may be ideal. Research on the effects of component-fidelity errors during skill acquisition could be continued in laboratory investigations to minimize any negative effects on learning in practice (Brand et al., **2019**). Similar studies could be performed with shelter dogs, which are consistently accessible during their time in the shelter and could benefit from additional opportunities for human engagement and behavioral intervention (Winslow et al., **2018**). In addition, animal behavior professionals could implement behavioral interventions to obtain data on the efficacy and efficiency of function-based interventions when implemented with initially high levels of procedural fidelity. Following further demonstrations of the conditions necessary to produce sufficient fidelity across interventions and populations (humans and animals) in laboratory settings, translation of these findings into practice will improve the efficacy and efficiency of behavioral interventions, thus providing better outcomes for clients.

Understanding that procedural-fidelity errors occur and can affect the outcomes of behavioral interventions should lead animal behavior professionals to retrain stakeholders when procedural fidelity degrades. Individuals interested in assessing and promoting accurate implementation of interventions for animal behavior may find guidance in the literature on behavioral consultation with teachers, because teachers are also likely to implement behavioral interventions following brief training with feedback from consultants. Strategies to promote fidelity in a consultation model like performance feedback (e.g., Noell et al., **1997**), graphical feedback of performance (Sanetti et al., **2007**), and self-monitoring (Rispoli et al., **2017**) may be generalizable to animal behavior professionals working with owners and volunteers. Further, given the importance of procedural fidelity for behavioral interventions, organizations overseeing certification for animal behavior practitioners should consider including procedural-fidelity requirements in their code of ethics.

References

ACVB position statements (n.d.). American College of Veterinary Behaviorists. Retrieved September 26, 2021, from https://www.dacvb.org/page/PositionStatement

Azrin, N. H., Holz, W. C., & Hake, D. F. (1963). Fixed-ratio punishment. *Journal of the Experimental Analysis of Behavior*, **6**(2), 141– 148. https://doi.org/10.1901/jeab.1963.6-141

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, **1**(1), 91– 97. https://doi.org/10.1901/jaba.1968.1-91

Belfiore, P. J., Fritts, K. M., & Herman, B. C. (2008). The role of procedural integrity: Using self-monitoring to enhance Discrete Trial Instruction (DTI). *Focus on Autism and Other Developmental Disabilities*, **23**(2), 95– 102. https://doi.org/10.1177/1088357607311445

Behavior Analyst Certification Board (2020). Ethics code for behavior analysts. Author.

Bergmann, S., Kodak, T., & Harman, M. J. (2021). When do errors in reinforcer delivery affect learning? A parametric analysis of treatment integrity. *Journal of the Experimental Analysis of Behavior*, **115**(2), 561– 577. https://doi.org/10.1002/jeab.670

Bergmann, S., Kodak, T., & LeBlanc, B. (2017). Effects of programmed errors of omission and commission during auditory-visual conditional discrimination training with typically developing children. *Psychological Record*, **67**(1), 109– 119. https://doi.org/10.1007/s40732-016-0211-2

Bisco, J. M., & Fier, S. G. (2020). Licensing and reporting in the U.S. pet insurance market. *Journal of Insurance Regulation*, **39**(2), 1– 19.

Bishop, C. D., Snyder, P. A., & Crow, R. E. (2015). Impact of video self-monitoring with graduated training on Implementation of embedded instructional learning trials. *Topics in Early Childhood Special Education*, **35**(3), 170– 182. https://doi.org/10.1177/0271121415594797

Bottini, S, Morton, H, Gillis, J, & Romanczyk, R. (2020). The use of mixed modeling to evaluate the impact of treatment integrity on learning. *Behavioral Interventions*, **35**, 372– 391. https://doi.org/10.1002/bin.1718

Brand, D., Elliffe, D., & DiGennaro Reed, F. D. (2018). Using sequential analysis to assess component integrity of discrete-trial teaching programs. *European Journal of Behavior Analysis*, **19**(1), 30– 47. https://doi.org/10.1080/15021149.2017.1404392

Brand, D., Henley, A. J., DiGennaro Reed, F. D., Gray, E., & Crabbs, B. (2019). A review of published studies involving parametric manipulations of treatment integrity. *Journal of Behavioral Education*, **28**, 1– 26. https://doi.org/10.1007/s10864-018-09311-8

Brand, D., Novak, M. D., DiGennaro Reed, F. D., & Tortolero, S. A. (2020). Examining the effects of feedback accuracy and timing on skill acquisition. *Journal of Organizational Behavior Management*, **40**(1–2), 3– 18. https://doi.org/10.1080/01608061.2020.1715319

Breeman, S., Vladescu, J. C., DeBar, R. M., Grow, L. L., & Marano, K. E. (2020). The effects of procedural integrity errors during auditory–visual conditional discrimination training: A preliminary investigation. *Behavioral Interventions*, **35**(2), 203– 216. https://doi.org/10.1002/bin.1710

Carroll, R. A., Kodak, T., & Adolf, K. J. (2016). Effect of delayed reinforcement on skill acquisition during discrete-trial instruction: Implications for treatment-integrity errors in academic settings. *Journal of Applied Behavior Analysis*, **49**(1), 176– 181. https://doi.org/10.1002/jaba.268

Carroll, R., Kodak, T., & Fisher, W. (2013). An examination of treatment integrity errors on skill acquisition during discrete trial instruction. *Journal of Applied Behavior Analysis*, **46**(2), 379– 394. http://0-dx.doi.org./10.1002/jaba.49

Catania, C. N., Almeida, D., Liu-Constant, B., & Reed, F. D. D. (2009). Video modeling to train staff to implement discrete-trial instruction. *Journal of Applied Behavior Analysis*, **42**, 387- 392. https://doi.org/10.1901/jaba.2009.42-387

Colbert-Luke, D. E., Fellner, W., Bauer, G. B., Manire, C. A., & Rhinehart, H. L. (2001). Husbandry and research training of two Florida manatees (*Trichechus manatus latirostris*). *Aquatic Mammals*, **27**, 16– 23.

Collier-Meek, M. A., Fallon, L. M., & Gould, K. (2018). How are treatment integrity data assessed? Reviewing the performance feedback literature. *School Psychology Quarterly*, **33**, 517– 526. https://doi.org/10.1037/spq0000239

Collier-Meek, M. A., Sanetti, L. M., & Fallon, L. (2021). Exploring the influences of assessment method, intervention steps, intervention sessions, and observation timing on treatment fidelity estimates. *Assessment for Effective Intervention*, **46**(1), 3– 13. https://doi.org/10.1177/1534508419857228

Collier-Meek, M. A., Sanetti, L. M., Gould, K., & Pereira, B. (2021). An exploratory comparison of three treatment fidelity assessment methods: Time sampling, event recording, and post-observation checklist. *Journal of Educational and Psychological Consultation*, **31**(3), 334– 359. https://doi.org/10.1080/10474412.2020.1777874

Colón, C. L. & Ahearn, W. H. (2019). An analysis of treatment integrity of response interruption and redirection. *Journal of Applied Behavior Analysis*, **52**, 337– 354. https://doi.org/10.1002/jaba.537

Cook, J. E., Subramaniam, S., Brunson, L. Y., Larson, N. A., Poe, S. G., & St. Peter, C. C. (2015). Global measures of treatment integrity may mask important errors in discrete-trial training. *Behavior Analysis in Practice*, **8**(1), 37– 47. https://doi.org/10.1007/s40617-014-0039-7

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (3rd ed.). Pearson.

Crosbie, J., Williams, A. M., Lattal, K. A., Anderson, M. M., & Brown, S. M. (1997). Schedule interactions involving punishment with pigeons and humans. *Journal of the Experimental Analysis of Behavior*, **68**(2), 161– 175. https://0-doi-org. /10.1901/jeab.1997.68-161

Davison, M., & McCarthy, D. (1980). Reinforcement for errors in a signal-detection procedure. *Journal of Experimental Analysis of Behavior*, **34**(1), 35– 47. https://doi.org/10.1901/jeab.1980.34-35

DiGennaro Reed, F. D., Reed, D. D., Baez, C. N., & Maguire, H. (2011). A parametric analysis of errors of commission during discrete-trial training. *Journal of Applied Behavior Analysis*, **44**(3), 611– 615. https://doi.org/10.1901/jaba.2011.44-611

Donnelly, M. G., & Karsten, A. M. (2017). Effects of programmed teaching errors on acquisition and durability of self-care skills. *Journal of Applied Behavior Analysis*, **50**(3), 511– 528. https://doi.org/10.1002/jaba.390

Dorey, N. R., Tobias, J. S., Udell, M. A. R., & Wynne, C. D. L. (2012). Decreasing dog problem behavior with functional analysis: Linking diagnoses to treatment. *Journal of Veterinary Behavior*, **7**(5), 276- 282. https://doi.org/10.1016/j.jveb.2011.10.002

Echterling-Savage, K., DiGennaro Reed, F. D., Miller, L. K., & Savage, S. (2015). Effects of caregiver-implemented aggression reduction procedure on problem behavior of dogs. *Journal of Applied Animal Welfare Science*, **18**(2), 181- 197.

Falakfarsa, G., Brand, D., Jones, L., Godinez, E. S., Richardson, D. C., Hanson, R. L., Velazquez, S. D., & Wills, C. (2021). Treatment integrity reporting in *Behavior Analysis in Practice* 2008-2019. *Behavior Analysis in Practice*, **15**, 443- 453. https://doi.org/10.1007/s40617-021-00573-9

Ferster, C. B., & Skinner, B. F. (1957). Schedules of reinforcement. *Appleton-Century-Crofts.* https://doi.org/10.1037/10627-000

Foreman, A. P., St. Peter, C. C., Mesches, G. A., Robinson, N., & Romano, L. M. (2021). Treatment integrity failures during timeout from play. *Behavior Modification*, **45**(6), 988– 1010. https://doi.org/10.1177/0145445520935392

Fritz, J. N., Fletcher, V. L., Dyer, S. P., Carpenter, K. A., Skrbec, M. L., & Yankelevitz, R. L. (2022). Functional analysis and treatment of aggression exhibited by cats toward humans during petting. *Journal of Applied Behavior Analysis*, **55**(1), 169– 179. https://doi.org/10.1002/jaba.877

Gresham, F. M. (2009). Evolution of the treatment integrity concept: Current status and future directions. *School Psychology Review*, **38**(4), 533– 540.

Gresham, F. M. (2014). Measuring and analyzing treatment integrity data in research. In L. M. H. Sanetti & T. R. Kratochwill (Eds.), *Treatment integrity: Conceptual, methodological, and applied considerations for practitioners* (pp. 109– 130). American Psychological Association. https://doi.org/10.1037/14275-007

Gresham, F. M., Dart, E. H., & Collins, T. A. (2017). Generalizability of multiple measures of treatment integrity: Comparisons among direct observation, permanent products, and self-report. *School Psychology Review*, **46**, 108– 121. https://doi.org/10.17105/SPR46-1.108-121

Gresham, F. M., MacMillan, D. L., Beebe-Frankenberger, M. E., & Bocian, K. M. (2000). Treatment integrity in learning disabilities intervention research: Do we really know how treatments are implemented? *Learning Disabilities Research & Practice*, **15**(4), 198– 205. https://doi.org/10.1207/SLDRP1504_4

Halbur, M., Kodak, T., Williams, X., Reidy, J., & Halbur, C. (2021). Comparison of sounds and words as sample stimuli for discrimination training. *Journal of Applied Behavior Analysis*, **54**, 1126- 1138. https://doi.org/10.1002/jaba.830

Herrnstein, R. J. (1974). Formal properties of the matching law. *Journal of the Experimental Analysis of Behavior*, **21**(1), 159– 164. https://doi.org/10.1901/jeab.1974.21-159

Hirst, J., & DiGennaro Reed, F. (2015). An examination of the effects of feedback accuracy on academic task acquisition in analogue settings. *Psychological Record*, **65**(1), 49– 65. https://doi.org/10.1007/s40732-014-0087-y

Hirst, J. M., DiGennaro Reed, F. D., & Reed, D. D. (2013). Effects of varying feedback accuracy on task acquisition: A computerized translational study. *Journal of Behavioral Education*, **22**, 1– 15. https://doi.org/10.1007/s10864-012-9162-0

Howard, V. J., & DiGennaro Reed, F. D. (2014). Training shelter volunteers to teach dog compliance. *Journal of Applied Behavior Analysis*, **47**(2), 344– 359. https://doi.org/10.1002/jaba.120

Howard, V. J., & Reed, F. D. D. (2015). An evaluation of training procedures for animal shelter volunteers. *Journal of Organizational Behavior Management*, **35**(3–4), 296– 320. https://doi.org/10.1080/01608061.2015.1093052

IAABC professional code of ethics. Retrieved September 26, 2021, from https://m.iaabc.org/about/ethics/

Jackson, C. R., McNutt, J. W., & Apps, P. J. (2012). Managing the ranging behaviour of African wild dogs (*Lycaon pictus*) using translocated scent marks. *Wildlife Research*, **39**, 31– 34. https://doi.org/10.1071/WR11070

Jamison, I. B. (2003). Turnover and retention among volunteers in human service agencies. *Review of Public Personnel Administration*, **23**(2), 114– 132. https://doi.org/10.1177/0734371X03023002003

Jenkins, S. R., Hirst, J. M., & DiGennaro Reed, F. D. (2015). The effects of discrete-trial training commission errors on learner outcomes: An extension. *Journal of Behavioral Education*, **24**(2), 196– 209. https://doi.org/10.1007/s10864-014-9215-7

Johnson, D. L., McGlynn, F. D., & Topping, J. S. (1973). The relative efficiency of four response elimination techniques following variable-ratio reinforcement training. *The Psychological Record*, **23**, 203– 208. https://doi.org/10.1007/BF03394156

Kisamore, A. N., Karsten, A. M., & Mann, C. C. (2016). Teaching multiply controlled intraverbals to children and adolescents with autism spectrum disorders. *Journal of Applied Behavior Analysis*, **49**, 826- 847. https://doi.org/10.1002/jaba.344

Kodak, T., Cariveau, T., LeBlanc, B., Mahon, J., & Carroll, R. A. (2018). Selection and implementation of skill acquisition programs by special education teachers and staff for students with autism spectrum disorder. *Behavior Modification*, **42**(1), 58– 83 https://doi.org/10.1177/0145445517692081

Lattal, K. A., Reilly, M. P. & Kohn, J. P. (1998). Response persistence under ratio and interval reinforcement schedules. *Journal of the Experimental Analysis of Behavior*, **70**(2), 165- 183. https://doi.org/10.1901/jeab.1998.70-165

Least intrusive, minimally aversive (LIMA) effective behavior intervention policy (2019). https://www.ccpdt.org/wp-content/uploads/2015/01/LIMA-Policy-2019.pdf

Lerman, D. C., Tetreault, A., Hovanetz, A., Strobel, M., & Garro, J. (2008). Further evaluation of a brief, intensive teacher-training model. *Journal of Applied Behavior Analysis*, **41**, 243- 248. https://doi.org/10.1901/jaba.2008.41-243

Magden, E. R., Haller, R. L., Thiele, E. J., Buchl, S. J., Lambeth, S. P., & Schapiro, S. J. (2013). Acupuncture as an effective adjunct therapy for osteoarthritis in chimpanzees (*Pan troglodytes*). *Journal of the American Association for Laboratory Animal Science*, **52**(4), 475– 480. https://pubmed.ncbi.nlm.nih.gov/23849446

Mal, M. E., Friend, T. H., Lay, D. C., Vogelsang, S. G., & Jenkins, O. C. (1991). Behavioural responses of mares to short term confinement and social isolation. *Applied Animal Behaviour Science*, **31**(1-2), 13- 24. https://doi.org/10.1016/0168-1591(91)90149-R

Martella, R., Leonard, I. J., Marchand-Martella, N. E., & Agran, M. (1993). Self-monitoring negative statements. *Journal of Behavioral Education*, **3**, 77– 86. https://doi.org/10.1007/BF00947146

Matey, N., Sleiman, A., Nastasi, J., Richard, E., & Gravina, N. (2021). Varying reactions to feedback and their effects on observer accuracy and feedback omission. *Journal of Applied Behavior Analysis*, **54**(3), 1188– 1198. https://doi.org/10.1002/jaba.840

McIntyre, L. L., Gresham, F. M., DiGennaro, F. D. & Reed, D. D. (2007). Treatment integrity of school-based interventions with children in the Journal of Applied Behavior Analysis 1991–2005. *Journal of Applied Behavior Analysis*, **40**, 659- 672. https://doi.org/10.1901/jaba.2007.659-672

Miltenberger, R. G. (2016). *Behavior modification: Principles and procedures* (6th ed.). Cengage Learning.

Morelli, G., Marchesini, G., Contiero, B., Fusi, E., Diez, M., & Ricci, R. (2020). A survey of dog owners' attitudes toward treats. *Journal of Applied Animal Welfare Science*, **23**(1), 1- 9. https://doi.org/10.1080/10888705.2019.1579095

Morris, K. L., & Slocum, S. K. (2019). Functional analysis and treatment of self-injurious feather plucking in a black vulture (*Coragyps attratus*). *Journal of Applied Behavior Analysis*, **52**(4), 918- 927. https://doi.org/10.1002/jaba.639

Mouzakitis, A., Codding, R. S., & Tryon, G. (2015). The effects of self-monitoring and performance feedback on the treatment integrity of behavior intervention plan implementation and generalization. *Journal of Positive Behavior Interventions*, **17**(4), 223– 234. https://doi.org/10.1177/1098300715573629

Noell, G. H., & Gansle, K. A. (2014). The use of performance feedback to improve intervention implementation in schools. In L. M. Hagermoser Sanetti & T. R. Kratochwill (Eds.), *Treatment integrity: A foundation for evidence-based practice in applied psychology* (pp. 161- 183). American Psychological Association. https://doi.org/10.1037/14275-009.

Noell, G. H., Gresham, F. M., & Gansle, K. A. (2002). Does treatment integrity matter? A preliminary investigation of instructional implementation and mathematics performance. *Journal of Behavioral Education*, **11**, 51– 67. https://doi.org/10.1023/A:1014385321849

Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly*, **12**(1), 77– 88. https://doi.org/10.1037/h0088949

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., Koenig, J. L., Reseter, J. L., & Duhon, G. J. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review*, **34**(1), 87– 106. https://doi.org/10.1080/02796015.2005.12086277

Pelletier, K., McNamara, B., Braga-Kenyon, P., & Ahearn, W. H. (2010). Effect of video self-monitoring on procedural integrity. *Behavioral Interventions*, **25**(4), 261– 274. https://doi.org/10.1002/bin.316

Peterson, L., Homer, A., & Wonderlich, S. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, **15**(4), 477– 492. https://doi.org/10.1901%2Fjaba.1982.15-477

Pfaller-Sadovsky, N., Arnott, G., & Hurtado-Parrado, C. (2019). Using principles from applied behaviour analysis to address an undesired behaviour: Functional analysis and treatment of jumping up in companion dogs. *Animals*, **9**(12), 1091. https://doi.org/10.3390/ani9121091

Pinkston, J. W., Ratzlaff, K. L., Madden, G. J., & Fowler, S. C. (2008). An inexpensive infrared detector to verify the delivery of food pellets. *Journal of the Experimental Analysis of Behavior*, **90**(2), 249– 255. https://doi.org/10.1901/jeab.2008.90-249

Rawson, R. A., & Leitenberg, H. (1973). Reinforced alternative behavior during punishment and extinction with rats. *Journal of Comparative and Physiological Psychology*, **85**, 593– 600. https://psycnet.apa.org/doi/10.1037/h0035312

Reed, D. D., & Kaplan, B. A. (2011). The matching law: A tutorial for practitioners. *Behavior Analysis in Practice*, **4**(2), 15– 24. https://doi.org/10.1007/BF03391780

Rispoli, M., Zaini, S., Mason, R., Brodhead, M., Burke, M. D., & Gregori, E. (2017). A systematic review of teacher self-monitoring on implementation of behavioral practices. *Teaching and Teacher Education*, **63**, 58– 72. https://doi.org/10.1016/j.tate.2016.12.007

Sanetti, L. M. H., & Collier-Meek, M. A. (2014). Increasing the rigor of treatment integrity assessment: A comparison of direct observation and permanent product methods. *Journal of Behavioral Education*, **23**, 60– 88. https://doi.org/10.1007/s10864-013-9179-z

Sanetti, L. M. H., & Fallon, L. M. (2011). Treatment integrity assessment: How estimates of adherence, quality, and exposure influence interpretation of implementation. *Journal of Educational & Psychological Consultation*, **21**, 209– 232. https://doi.org/10.1080/10474412.2011.595163

Sanetti, L. M. H., Luiselli, J. K., & Handler, M. W. (2007). Effects of verbal and graphic performance feedback on behavior support plan implementation in a public elementary school. *Behavior Modification*, **31**(4), 454– 465. https://doi.org/10.1177/0145445506297583

Shimoff, E., Matthews, B. A., &Catania, A. C. (1986). Human operant performance: Sensitivity and pseudosensitivity to contingencies. *Journal of the Experimental Analysis of Behavior*, **46**, 149- 157. https://doi.org/10.1901/jeab.1986.46-149

St. Peter, C. C., Byrd, J. D., Pence, S. T., & Foreman, A. P. (2016). Effects of treatment-integrity failures on response-cost procedure. *Journal of Applied Behavior Analysis*, **49**(2), 308– 328. https://doi.org/10.1002/jaba.291

St. Peter Pipkin, C. C., Vollmer, T. R., & Sloman, K. N. (2010). Effects of treatment integrity failures during differential reinforcement of alternative behavior: A translational model. *Journal of Applied Behavior Analysis*, **43**(1), 47– 70. https://doi.org/10.1901/jaba.2010.43-47

Strassberg, L. R., Waggoner, L. P., Deshpande, G., & Katz, J. S. (2019). Training dogs for awake, unrestrained functional magnetic resonance imaging. *Journal of Visualized Experiments*, **152**, e60192. https://doi.org/10.3791/60192

Suhrheinrich, J., Dickson, K. S., Chan, N., Chan, J. C., Wang, T., & Stahmer, A. C. (2020). Fidelity assessment in community programs: An approach to validating simplified methodology. *Behavior Analysis in Practice*, **13**, 29– 39. https://doi.org/10.1007/s40617-019-00337-6

Vollmer, T. R., Roane, H. S., Ringdahl, J. E. & Marcus, B. A. (1999). Evaluating treatment challenges with differential reinforcement of alternative behavior. *Journal of Applied Behavior Analysis*, **32**, 9– 23. https://doi.org/10.1901/jaba.1999.32-9

Vollmer, T. R., Sloman, K. N., & St. Peter Pipkin, C. C. (2008). Practical implications of data reliability and treatment integrity monitoring. *Behavior Analysis in Practice*, **1**(2), 4– 11. https://doi.org/10.1007/BF03391722

Waite, M., & Kodak, T. (2021). Owner-implemented functional analyses and reinforcement-based treatments for mouthing in dogs. *Behavior Analysis in Practice*, **15**, 269- 283. https://doi.org/10.1007/s40617-021-00554-y

Wayne, A., McCarthy, R., & Lindenmayer, J. (2011). Therapeutic antibiotic use patterns in dogs: Observations from a veterinary teaching hospital. *The Journal of Small Animal Practice*, **52**(6), 310– 318. https://doi.org/10.1111/j.1748-5827.2011.01072.x

Weston, R., Davis, T. N., Radhakrishnan, S., O'Guinn, N., & Rivera, G. (2020). Comparing performance feedback and video self-monitoring within a BST package to train pre-service behavior analysts to conduct

preference assessments. *Journal of Behavioral Education*, **29**(3), 571– 583. https://doi.org/10.1007/s10864-019-09333-w
Wilder, D. A., Atwell, J., & Wine, B. (2006). The effects of varying levels of treatment integrity on child compliance during treatment with a three-step prompting procedure. *Journal of Applied Behavior Analysis*, **39**(3), 369– 373. https://0-doi-org. /10.1901/jaba.2006.144-05
Winslow, T., Payne, S. W., & Massoudi, K. A. (2018). Functional analysis and treatment of problem behavior in 3 animal shelter dogs. *Journal of Veterinary Behavior*, **26**, 27– 37. https://doi.org/10.1016/j.jveb.2018.04.004
Witt, J. C., Noell, G. H., LaFleur, L. H., & Mortenson, B. P. (1997). Teacher use of interventions in general education settings: Measurement and analysis of the independent variable. *Journal of Applied Behavior Analysis*, **30**(4), 693– 696. https://doi.org/10.1901/jaba.1997.30-693
Young, R. J., Schetini de Azevedo, C., & Cipreste, C. F. (2020). Environmental enrichment. In V. A. Melfi, N. R. Dorey, & S. J. Ward (Eds.), *Zoo Animal Learning and Training* (2nd ed., pp. 101- 118). John Wiley & Sons Ltd.