5-2019

Trends in Maintenance and Stability of Functional Analysis Results

Sara Snow
sesnow8@gmail.com

Trends in Maintenance and Stability of Functional Analysis Results

by

Sara Snow

A Thesis

Submitted to the Graduate Faculty of

St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree

Master of Science in

Applied Behavior Analysis

May, 2019

Thesis Committee:
Michele Traub, Chairperson
Benjamin Witts
Kimberly Schulze

Abstract

Functional analysis (FA) has become one of the most relied-upon assessments within our field and is typically used to make treatment decisions. However, relatively little is known about the longevity of these treatments or how changes in behavioral function impact treatment success or maintenance. This study examined the long-term effects of FA-informed treatments, specifically the reported follow-up data and data regarding the stability of the function during and following treatment. Studies included in this review are those that used the results of an FA to identify and implement an appropriate treatment and that reported follow up data after termination of treatment. This study contributes to the literature in two important ways: first, it details the current trends in long-term follow up and the collection of longitudinal data on treatment outcomes, and second, it provides directions for future research on the stability of function over time and the implications of functional stability on treatment maintenance.
Keywords: long-term, follow-up, maintenance, functional analysis, functional stability

Table of Contents

Chapter 1: Introduction

The impact of Iwata, Dorsey, Slifer, Bauman, and Richman's (1982/1994) article on functional analysis methods on the field of behavior analysis is well documented. Functional analysis (FA) is a form of operant procedure in which antecedents and/or consequences are manipulated to identify the environmental variables maintaining aberrant behavior. The results of an FA can then be used to create an appropriate intervention plan for the assessed individual. While the Iwata et al. (1982/1994) study helped tackle the field's limitations in treatment for self-injurious behavior (SIB), what set it apart was how its methods addressed current hypotheses on the etiology of SIB (Carr, 1977). This pivotal study and its methods exemplify the field's shift from arbitrarily implementing reinforcement and punishment procedures in its early years to the current approach in which treatment is based on the identification and manipulation of the behavior's maintaining contingencies. By treating aberrant behavior based on its function, rather than its topography, practitioners are better able to provide an individualized, effective treatment plan for their clients.

Previous literature reviews have provided extensive information on past and current trends in the use of functional analysis methods (Beavers, Iwata, & Lerman, 2013; Hanley, Iwata, & McCord, 2003). Trends assessed in these reviews include subject and setting characteristics, topographies of problem behaviors, types of functional analyses, test conditions and experimental design, duration of sessions and assessment, outcomes of assessment, and the way in which data were displayed and analyzed. Hanley et al. (2003) used this information to provide directions for future researchers and recommendations for best practice, and Beavers et al. (2013) updated the review and recommendations in light of more recent developments.

Though these reviews discuss some advancements and limitations of current research on functional analyses, neither assessed the collection of follow-up data.

One reason follow-up data are collected is to assess whether treatment effects have generalized over time. Generality, or the ability of a behavioral change to remain effective over time, across environments, or across different behaviors, was one of the seven key dimensions of the field emphasized by Baer, Wolf, and Risley (1968) in their seminal article on applied behavior analysis. Generality of behavior change over time, better known as treatment maintenance, is demonstrated through the analysis of long-term follow-up data. Stokes and Baer (1977) outlined specific procedures to establish generalization and maintenance of treatment effects, terming these procedures a "technology" of generalization, and more articles describing specific programming for maintenance have been published since (Durand & Carr, 1991; Foxx, 1999). Despite the abundance of information on programming for maintenance, there are often instances in which relapse occurs.

There are several reasons why treatment effects sometimes fail to maintain over time. One commonly reported contributor to a lack of treatment maintenance is diminished treatment integrity. Low treatment integrity may be due to implementer error, procedural drift, or deliberate changes to written programs by implementers. Interventions with low treatment integrity have been shown to have decreased effectiveness (DiGennaro & Martens, 2007; Noell, Gresham, & Gansle, 2002; Vollmer, Roane, Ringdahl, & Marcus, 1999). Related to poor integrity, another commonly cited contributor to the relapse of problem behavior is extinction-based resurgence (Lieving, Hagopian, Long, & O'Connor, 2004; Volkert, Lerman, Call, & Trosclair-Lasserre, 2009; Wacker et al., 2011; Wacker et al., 2013). Extinction-based resurgence

is a phenomenon in which a recently reinforced behavior is put on extinction, leading to the recurrence of behaviors that were previously reinforced under similar conditions (Epstein, 1983). If replacement behaviors that were established during treatment are later put on extinction, the aberrant behavior that had originally been extinguished may reemerge. Third, false-negative conclusions about multiple controlling variables may also contribute to a lack of treatment maintenance in some instances (McKerchar, Kahng, Casioppo, & Wilson, 2001). Treatment may fail soon after implementation or cease to be effective after time due to the target behavior still contacting the form(s) of reinforcement missed during assessment. The prevalence of behaviors that are maintained by multiple reinforcement contingencies may help to explain why false-negative conclusions within FAs regularly occur. Beavers et al. (2013) found that nearly one quarter (24.3%) of FA studies published between 2000 and 2012 included behaviors that were multiply controlled. They also found a significant increase in the percentage of studies that tested multiple response topographies within a single FA (75.9% between 2000 and 2012). This method of combining response topographies into one FA, although timely, may prevent the accurate functional assessment of individual response topographies.

These three maintenance-related issues are also related to the *effectiveness* of the treatment. It is possible that many of these treatments that were initially successful but fail to be maintained lacked effectiveness; the treatments did not alter behavior enough to be socially important. Baer, Wolf, & Risley (1968) explain that if a treatment does not make a change large enough for any practical value, then the application has failed. They classify effectiveness as another one of the seven key dimensions of ABA. Treatments may have proven to be *efficacious* in the clinic setting but failed to make a valuable change in the natural setting. Clients may be

relapsing after a period of time, or they may simply never have changed their behavior in the natural setting due to an ineffective treatment.

Another factor that may influence treatment effectiveness and maintenance is the stability of the target behavior's identified function. One possible cause of the recurrence of problem behavior may be that the previously extinguished behavior has contacted new maintaining contingencies. The original treatment designed based on one function will likely not be effective in continuing to eliminate the behavior if the behavior takes on a new, differing function. Lerman, Iwata, Smith, Zarcone, and Vollmer (1994) conducted a study with 4 individuals who had returned to a day-treatment program due to the recurrence of SIB after successfully treating it 2 months to 2 years prior. By conducting a second FA, the authors determined that 3 of the 4 individuals' SIB had acquired new or additional functions. Because the previously successful treatment did not address the newly acquired function, a relapse in SIB occurred. Lerman et al. (1994) explain that by reassessing function when an initially successful treatment fails, both the client and the field of behavior analysis benefit. The client benefits by having a more effective, appropriate treatment recommended based on the newly identified function, and the field benefits by gaining a better understanding of the variables that contribute to a change in behavioral function.

Despite the strong argument for reassessing function made by Lerman et al. (1994) and research demonstrating that changes in behavioral function occur (Carr & McDowell, 1980; Guess & Carr, 1991), this phenomenon has hardly been explored since 1994. Gresham, Watson, & Skinner (2001) mentioned changes in function as one of the current issues in functional behavioral assessment: "What is the stability of behavioral function over time, settings, and

assessors? …[B]ehavior may serve one function in a particular setting at one point in time and serve another function in another setting at another point in time" (pp. 169). This is an important question that warrants further research. While it is advantageous to know the contributing factors that lead to a change in function, it is also important to know how quickly and how frequently changes occur. Knowing how often the maintaining variables of an aberrant behavior change or remain the same could impact treatment decisions and help strengthen treatment effectiveness and maintenance of effects. While Lerman et al. (1994) studied the change in behavioral function through assessing cases in which previously successful treatments were not maintained, another method for evaluating functional stability is to look at functional data across days or sessions (Fox, Conroy, & Heckaman, 1998). Assessing function more frequently may reveal results of greater variability than research typically shows.

Given the lack of information on functional stability, the small sample of research demonstrating change in function, and the popularity that research related to resurgence and treatment relapse has received, it is more important than ever to examine the literature for trends in function and its role in treatment relapse. Practitioners need to know how and how often changes in function occur, so they can program effectively. The purpose of this review was to outline the current trends in functional analysis methodology regarding long-term follow up data, maintenance of treatment effects, and the stability of function. Specifically, this paper aims to identify (a) how often follow-up data are being recorded and reported; (b) whether treatment effects are maintained long-term; (c) whether the function of behavior remains stable over time; and (d) how functional stability affects the maintenance of treatment effects.

Chapter 2: General Method

A review of all published FA studies dating from 1994—the republishing of the seminal article by Iwata and colleagues—through May 2018 was conducted. The initial pool of articles was found through searching the database *PsychInfo* using the search terms: "functional analysis" and "behavior analysis". The search included studies published within the year range, in English, used human subjects, and were published within a peer-reviewed journal. The two search terms were then combined with additional search terms ("follow-up" and "maintenance") and entered into *Google Scholar* to identify any additional FA studies.

For each article this search produced, the primary purpose of the article was categorized as either (a) treatment of problem behavior, or (b) method refinement or other. Studies were all considered to fall under treatment of problem behavior unless either the abstract or the purpose section of the article discussed modification of specific methods (i.e. data collection, setting, conditions, etc.), and did not also mention treatment of problem behavior as their goal; these articles were categorized under method refinement. Articles that did not contain an empirical FA study were what constituted "other" (e.g. literature reviews; theoretical papers). Only articles with treatment of problem behavior as its primary purpose were included in this study since studies aiming to refine FA methods lack the need for the collection of follow-up data, and non-empirical studies do not collect data. Each article with the goal of treatment of problem behavior was then analyzed for the inclusion criteria.

Inclusion Criteria

Each FA study was assessed and included if it met the following criteria: (a) an FA was conducted, and its results published, (b) treatment was implemented based on the FA's results,

(c) follow-up data were collected, and (d) the time between the last treatment measure to the last follow-up measure was equal to or longer than one month. The criteria set by Hanley, Iwata, & McCord (2003) were used to determine studies that conducted an FA: "(a) a pretreatment assessment that was based on (b) direct observation and measurement of problem behavior was conducted under (c) at least two conditions involving manipulation of some environmental variable in an attempt (d) to demonstrate a relation between the environmental event and behavior," (p. 149-150). In addition to these requirements, these studies also needed to publish the results of the FA to be included. The other inclusion criteria simply required some form of data; data to show that treatment was implemented, and if/when follow-up measures were taken.

Data Collection

Additional characteristics were assessed for each FA study meeting inclusion criteria. These characteristics were used to further identify any trends within the FA literature. Each characteristic and the method of assessment is described below.

Target behavior. Target behavior was recorded based on the authors' description of behavior within the article. Some studies were more specific in their description and the behaviors listed in the subsequent tables reflect that; some studies labeled behaviors as "SIB" while others were more specific (e.g. "hair pulling"). Behaviors were also classified as either *severe* or *non-severe.* Severe behaviors were any behaviors that may have caused physical harm to the client or others, including but not limited to self-injurious behavior, aggression, property destruction, or elopement. Behavior was categorized as non-severe for behaviors that did not cause any physical harm to the client or others, such as stereotypy, vocal outbursts, or food refusal.

Function. The behavior's identified function(s) was recorded as described by the study's authors, and any differing or additional functions in follow-up or later assessment were noted. This was recorded to identify studies in which treatment effects did not maintain due to the behavior acquiring a differing or additional function.

Subject characteristics. The number of subjects, and subjects' age, gender, and diagnosis were recorded as described by the study's authors.

Treatment. In addition to recording whether treatment was conducted, the type of treatment was recorded as well. This was recorded to identify if certain treatments were more susceptible to maintenance or relapse than others.

Condition of follow-up. The conditions under which follow-up measures were collected within each study was categorized as being either: (a) identical or similar to treatment phase, or (b) no treatment.

Maintenance. The occurrence or non-occurrence of maintenance was also recorded for each article, to identify any correlations between the study's characteristics with the long-term effects of treatment. For studies that did not show maintenance, the study was further reviewed to identify: (a) what caused the lack of maintenance, and (b) if the behavior's function was re-assessed through an additional FA.

Inter-rater Reliability

Inter-rater reliability (IRR) was collected on 20% of the FA studies that were evaluated for follow-up measures (studies that conducted and published the results of an FA and implemented subsequent treatment). The second rater independently reviewed each article to identify if treatment of problem behavior was conducted, if follow-up data were collected, the

time between termination of treatment and the last follow-up measure, and whether maintenance was achieved. Articles were presented in a random order without specific information on the study's inclusion criteria. Inter-rater reliability was calculated by totaling the number of agreements for each rating divided by the number of agreements plus disagreements and multiplied by 100. Inter-rater reliability for all categories on all studies was 100%.

Chapter 3: Results

The search method yielded a total of 881 results on *PsychInfo*. Four-hundred-twenty-four of these results were excluded because they were not empirical FA studies (e.g. theoretical papers, non-behavioral assessments). Of the 457 results containing an FA, 149 of these results were excluded due to the nature of the study; the purposes of these studies included refining FA methods, training others to conduct FAs, analyses of verbal behavior, comparison of FA results to other assessment results, testing hypotheses about behavior, or otherwise no treatment implementation following the FA. Twenty-five additional studies were excluded due to not reporting treatment implementation and did not fit any of the previously described categories. From this, a total of 283 empirical FA studies were reviewed. Each was evaluated to identify which studies conducted follow-up measures, and of those that did, which conducted follow-up one month or longer after termination of treatment. In addition, the first 100 results Google Scholar yielded were reviewed to identify any additional studies. Twenty-one additional studies containing an FA were identified, 10 of which conducted follow-up, and 8 of which met all inclusion criteria. In total, 55 of the FA studies had reported conducting follow-up (18%), 48 of which conducted follow-up one month or longer after termination of treatment (16%). Figure 1 displays the data on each inclusion criterion.

The 48 articles that met inclusion criteria were further reviewed to identify target behaviors, function(s), subject characteristics, treatment type, condition of follow-up, and whether maintenance of treatment effects was demonstrated. These results are displayed in Table 1. Eighty-one percent of the articles included at least one severe behavior as a target behavior, leaving less than 20% assessing and treating only non-severe behaviors (verbalizations,

stereotypy). Children were the participants in 79% of studies, with adults participating in 25% of studies, as 3 studies included both children and adults as participants. Differential reinforcement was used in 33% of the studies, functional communication training in 27% of studies, and extinction in 19% of studies. Each of the articles also demonstrated treatment maintenance with at least one participant.

Figures 2 and 3 depict the included articles by time between treatment and the last follow-up measure and by year of publication, respectively. Figure 2 shows studies that included any duration between termination of treatment and the collection of follow-up data; those with less than one month were not excluded in this figure to provide a visual representation for comparison. Thirty-five percent of the studies that reported follow-up data collected data over 6 months after the termination of treatment (range of 7 months-4 years), 24% had collected their last follow-up measure between 4 and 6 months after termination of treatment, 27% measured between 1 and 3 months after termination of treatment, 11% measured less than 1 month after termination of treatment, and 5% did not specify the duration between termination of treatment and the collection of follow-up data. Both the number and percentage of follow-up articles is highest between 1994-1999 (20 articles; 34% of published FA articles). Between 2004-2008, only 5 published FA studies contained follow-up data (10% of published FA studies), while the lowest percentage of FA studies containing follow-up (9%) was between 2009-2013.

Chapter 4: Discussion

The results provide a few important findings regarding the FA literature. First, FA studies are rarely reporting follow-up data. Since 1994, only 16% of published FA studies have demonstrated long-term treatment effects using follow-up data. This becomes a concern regarding the effectiveness of FA-based treatments. Research shows extensively that these FA-informed treatments are *efficacious*, meaning that behavior can be altered under tightly controlled settings and conditions. However, without more data to support the longevity of these treatments, it is nearly impossible to say that these treatments are *effective.* For treatment to be deemed effective, it must demonstrate behavior change in the natural environment without the tight controls of experimentation. The field of ABA has an ethical obligation to provide clients with an effective treatment (Professional and Ethical Compliance Code for Behavior Analysts, 2014). Without stronger evidence on the effectiveness of FA-based treatments, ABA practitioners are not acting within the ethical guidelines set out by the Behavior Analyst Certification Board. This is a deeply concerning issue considering how commonplace FAs and FA-based treatments have become within the field. This also limits the credibility of ABA as a science; these practices are heavily used without strong evidence to support their long-term effectiveness. This only further confines the realms in which ABA is accepted and practiced. To demonstrate with extensive data that a certain practice (FA-based treatments) is successful in changing a client's behavior for several months or years could be an incredible turning point for ABA in relation to acceptance of the field by clients, the public, and even policymakers. The scope of ABA services has the potential to expand to a higher number of clients, clients of

differing diagnoses, and new settings. Demonstrating the long-term effectiveness of these treatments is crucial in extending the reach of the field.

In addition to the expansion of ABA, these treatments need to be connected to their long-term outcomes so that the methods leading to long-term maintenance and the variables leading to future relapse in problem behavior are identified. The practices that more often lead to long-term maintenance can be disseminated and used by the field, and the factors leading to future relapse can be prevented. One of these factors is change in behavioral function. Lerman et al. (1994) provide an example of how a change in function affected clients who had previously had an FA and FA-based treatment. These clients had a relapse in problem behavior and were re-referred to the clinic for services. After identifying a change in behavioral function, the authors were able to modify treatment appropriately and again eliminate problem behavior. This may have been a positive outcome for these clients; however, clients may not always come back when problem behavior remerges. Clients or their families may lose confidence in our services after they fail to last over time and seek help elsewhere. Without follow-up measures being taken, we must rely on the few clients that do return following relapse to get any information on why treatment failed to maintain. The more information that can be collected on the variables affecting maintenance and relapse, the better practitioners and researchers can plan for and control them. This will help prevent clients from having to come back and repeat the assessment and treatment process again and again. Consequently, this will aid in the elimination of false-positive reporting and/or exaggeration of successful results in the literature. Currently, a client may be re-referred multiple times due to several

treatments failing to maintain, while simultaneously being the subject of multiple studies publishing what appeared to be "successful" results. Although the client's history indicates several unsuccessful treatment attempts, the literature would reflect only the successful one(s) and lack any long-term outcome data. Disseminating both the failed and successful attempts can help demonstrate the variables that contribute to treatment relapse, so they are better recognized and controlled for in practice.

From the small collection of published studies that were able to successfully conduct follow-up and demonstrate treatment maintenance, a few trends are worth noting. Only about one-third of these articles collected follow-up longer than six months after treatment ended. This adds to the concern that treatment maintenance is not being demonstrated. While treatment effects were shown to last a few months after treatment ended for two-thirds of these follow-up articles, it brings into question whether these effects would maintain throughout the following months and years. Also, 80% of the articles that conducted follow-up measures longer than 6 months after treatment were behaviors considered severe (SIB; aggression) or even life-threatening (SIB). The severity of a behavior may play a part in whether follow-up measures are taken several months or years after treatment has ended.

There are other factors that affect the collection of follow-up data. Researchers often encounter several barriers when trying to conduct follow-up. A common issue is subject or client attrition (Arya, Duncana, Duncana, & Hopsa, 1999; Kanter et al., 2006). Clinicians and practitioners oftentimes have trouble getting clients or clients' families to adhere to or complete treatment. When clients do adhere throughout the entirety of treatment, they may be reluctant to have therapists or researchers return to observe after the behavior has changed due to the

intrusiveness of having someone observe them. They may be even more reluctant if the client has relapsed. On a related note, practitioners are limited to what insurance will cover. Insurance is often not willing to pay for follow-up to be collected. Thomas, Ellis, McLaurin, Daniels, & Morrissey (2007) described some of the barriers families face simply trying to access services for their children with ASD. Given the difficulty for some to get insurance to cover their services, it is not surprising that insurance companies are not willing to pay for follow-up measures. These factors may be part of the reason that most of the follow-up studies with data collected over 6 months after treatment (80%) were those containing severe or life-threatening behavior. The treatment and maintenance of these behaviors is essential for the safety of the client and those around him, so it is likely that families and even insurance companies are more willing to allow the collection of follow-up measures under these circumstances.

Another factor that hinders the collection and reporting of follow-up data is the enormous amount of professional pressure to publish. Academics are typically required to produce a high quantity of publications to maintain and advance in their career. It may be more beneficial for researchers in their career to publish their studies sooner, as opposed to waiting for follow-up to be collected first. It may also be of greater interest to researchers to produce more publications with less content, rather than fewer publications with combined content (Dupps & Randleman, 2012). This may help to explain why so many FA articles did not publish corresponding treatment results; professionally, researchers are better off using their treatment data in a subsequent article. This professional pressure occurs across all the sciences but becomes especially problematic for ABA as it seeks out the best assessments and treatments to provide

clients. Effective assessments are unable to be connected to their corresponding treatments, as well as their long-term results.

Functional Stability

Without the data showing long-term maintenance of FA-based treatments, we do not know how often clients are relapsing and how often it is due to a change in function. Given the evidence for changes in function in the literature, it is important we are reassessing function not only when a relapse in problem behavior occurs, but even just reassessing function more frequently. Valdovinos, Nelson, Kuhle, & Dierks (2009) provided an example of the usefulness in assessing function continually. Their study involved conducting multiple FAs with individuals undergoing psychotropic medication changes. The authors found that some medication changes led to differences in the target behavior's function, acting as an establishing operation for new behavioral functions. This study provides an example of how assessing function more often can show important variables effecting the outcome of treatment.

Another area in which a change in function has been reported is when an automatically reinforced behavior acquires a social-positive function. Carr & McDowell (1980) found that the scratching behavior of a child that initially began due to poison ivy had developed an attention function that persisted after the poison ivy had healed. The authors were able to successfully treat scratching behavior knowing the previous and current functions, and these results maintained in the follow-up 9 months later. Similar results were found for a client with coughing behavior that had acquired an attention function after a respiratory infection had healed. (Watson & Sterling, 1998). These are both examples of situations in which treatment was required after a new function was established.

It is important we are assessing for and responding appropriately to change in function given the evidence of it within the literature. Even if a behavior is extinguished through treatment, the behavior still remains in the individual's repertoire, meaning it could recur at any time given the appropriate contingencies are in place. If the extinguished behavior begins to receive an alternative source of reinforcement, a new function may be acquired, just as Lerman, et al. (1994) found. This change in function is a phenomenon that needs to be better understood so practitioners can control for and prevent its occurrence.

Currently, there are limited long-term follow-up data for FA-based treatments, making the knowledge about the longevity of these treatments and the stability of function limited. While these treatments can be life-changing for clients, it is important to ensure that these effects will last. To help decrease the lapse of information on maintenance and functional stability, follow-up measures need to be taken whenever possible. The collection of follow-up measures, however, is not the only barrier when it comes to publishing these data. Researchers often cannot afford to wait several months or years to submit their publication; they need to get publications out sooner rather than later. The pressure academics face to publish will not go away anytime soon. What can help provide this information without prolonging submissions is to routinely connect FAs with their effective treatments and the corresponding long-term outcomes. Researchers can reference previous articles regarding the same client(s) within their current article. For example, if an article was previously published on a client's assessment, the following article on the same client's treatment and long-term outcomes can be published with a mention to the first article (see Foxx & Faw, 1990, for an example*).* This will help connect FAs with their FA-informed treatments and to the long-term outcomes of these treatments. Another way to connect these

pieces is to add in supplementary figures or graphs into previous publications. If an article was previously published on a client's treatment and follow-up measures were taken a year later, that follow-up data can be added as a supplementary figure or graph. Many journals, including the *Journal of Applied Behavior Analysis*, allow for this type of submission as they are now offered in an online format. This way readers are able to connect the assessment to the corresponding treatment and to the long-term outcomes. Research acts as a model for what is done in practice. Consumers of the literature will see the connection and be able to imitate these best practices knowing their long-term effects.

While this review illustrated some of the needs within the field, it contains some limitations that should be noted. The biggest limitation being that none of the information obtained was systematically analyzed; articles were all categorized as the authors had described. For example, a behavior's function was not individually assessed by the reviewer given the published data, rather it was classified the same as the author had done so. With the different methods to identify function, successful treatment, and maintenance of treatment, it is possible that using a more systematic method for classifying each study would elicit slightly different results.

Second, only studies containing an FA were included. It is possible that articles containing only treatment outcomes that had previously conducted an FA were missed, skewing the data. There may be more evidence showing that FA-based treatments maintain long-term in articles that were missed in this review (i.e. those that did not publish results of an FA). Future research could modify search methods to ensure articles containing FA-based treatments are all included. This could be done by reviewing certain types of treatments and identifying their

trends in maintenance. Using articles containing FAs was the outlet chosen in this study to identify these trends. The use of different inclusion criteria or search methods may provide information that further supports, complements, or even negates what was found in this study. Although this was a limitation, it further demonstrates the difficulty in connecting FAs to their FA-based treatments and to their long-term outcomes.

While those in the field of ABA may be aware of a general lack in follow-up data collection within the literature, this study provides data to confirm that notion. Desperately needed is for researchers to collect and report their follow-up measures. A lot of important information about treatment maintenance, functional stability, and factors related to relapse are going unreported or unnoticed. Connecting our assessments with treatment results, and later with follow-up data will only help further the use of our practices and better support our clients.

References

Arya, D. V., Duncana, T. E., Duncana, S. C., & Hopsa, H. (1999). Adolescent problem behavior: The influence of parents and peers. *Behavior Research and Therapy, 37,* 217-230.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis, 1,* 91-97.

Beavers, G. A., Iwata, B. A., & Lerman, D. C. (2013). Thirty years of research on the functional analysis of problem behavior. *Journal of Applied Behavior Analysis, 46,* 1-21.

Carr, E. G. (1977). The motivation of self-injurious behavior: A review of some hypotheses. *Psychological Bulletin, 84,* 800-816.

Carr, E. G., & McDowell, J. J. (1980). Social control of self-injurious behavior of organic etiology. *Behavior Therapy, 11,* 402-409.

DiGennaro, F. D., & Martens, B. K., (2007). A comparison of performance feedback procedures on teachers' treatment implementation integrity and students' inappropriate behavior in special education classrooms. *Journal of Applied Behavior Analysis, 40,* 447-461.

Dupps, W. J., & Randleman, J. B. (2012). The perils of the least publishable unit. *Journal of Cataract & Refractive Surgery, 38,* 1517-1518.

Durand, V. M., & Carr, E. G. (1991). Functional communication training to reduce challenging behavior: Maintenance and application in new settings. *Journal of Applied Behavior Analysis, 24,* 251-264.

Foxx, R. M. (1999). Long term maintenance of language and social skills. *Behavioral Interventions, 14,* 135-146.

Foxx, R. M., & Faw, G. D. (1990). Long-term follow-up of echolalia and question answering. *Journal of Applied Behavior Analysis, 23,* 387-396.

Guess, D., & Carr, E. G. (1991). Emergence and maintenance of stereotypy and self-injury. *American Journal on Mental Retardation, 96*, 299-319.

Hanley, G. P., Iwata, B. A., & McCord, B. E. (2003). Functional analysis of problem behavior: A review. *Journal of Applied Behavior Analysis, 36*, 147-185.

Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*, 197-209. (Reprinted from *Analysis and Intervention in Developmental Disabilities, 2*, 3-20, 1982.)

Kanter, J. W., Landes, S. J., Busch, A. M., Rusch, L. C., Brown, K. R., & Baruch, D. E. (2006). The effect of contingent reinforcement on target variables in outpatient psychotherapy for depression: A successful and unsuccessful case using functional analytic psychotherapy. *Journal of Applied Behavior Analysis, 39,* 463-467.

Lerman, D. C., Iwata, B. A., Smith, R. G., Zarcone, J. R., & Vollmer, T. R. (1994). Transfer of behavioral function as a contributing factor in treatment relapse. *Journal of Applied Behavior Analysis, 27*, 357-370.

Lieving, G. A., Hagopian, L. P., Long, E. S., & O'Connor, J. (2004). Response-class hierarchies and resurgence of severe problem behavior. *Psychological Record, 54*, 621-634.

McKerchar, T. L., Kahng, S., Casioppo, E., & Wilson, D. (2001). Functional analysis of self-injury maintained by automatic reinforcement: exposing masked social functions. *Behavioral Interventions, 16*, 59-63.

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.

Professional and Ethical Compliance Code for Behavior Analysts (2014). In *Behavior Analyst Certification Board.* Retrieved from https://www.bacb.com/wp-content/uploads/2017/09/170706-compliance-code-english.pdf

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10,* 349-367.

Thomas, K. C., Ellis, A. R., McLaurin, C., Daniels, J., & Morrissey, J. P. (2007). Access to care for autism-related services. *Journal of Autism & Developmental Disorders, 37,* 1902-1912.

Valdovinos, M. G., Nelson, S. M., Kuhle, J. L., & Dierks, A. M. (2009). Using analogue functional analysis to measure variations in problem behavior rate and function after psychotropic medication changes: A clinical demonstration. *Journal of Mental Health Research in Intellectual Disabilities, 2,* 279-293.

Volkert, V. M., Lerman, D. C., Call, N. A., & Trosclair-Lasserre, N. (2009). An evaluation of resurgence during treatment with functional communication training. *Journal of Applied Behavior Analysis, 42,* 145-160.

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.

Wacker, D. P., Harding, J. W., Berg, W. K., Lee, J. F., Schieltz, K. M., Padilla, Y. C., Nevin, J. A., & Shahan, T. A. (2011). An evaluation of persistence of treatment effects during long-term treatment of destructive behavior. *Journal of the Experimental Analysis of Behavior, 96,* 261-282.

Wacker, D. P., Harding, J. W., Morgan, T. A., Berg, W. K., Schieltz, K. M., Lee, J. F., & Padilla, Y. C. (2013). An evaluation of resurgence during functional communication training. *Psychological Record, 63*, 3-20.

Watson, T. S., & Sterling, H. E. (1998). Brief functional analysis and treatment of a vocal tic. *Journal of Applied Behavior Analysis, 31,* 471-474.

Appendix A: Tables

Table 1

Summary of study characteristics not relating to inclusion criteria and function.

#	Year Published	Target Behaviors	Subject Characteristics	Treatment	Condition of follow-up
1	2017	Aggression, screaming	8-year-old male with ASD	NCR, Differential reinforcement, Response cost	Similar to treatment
2	2016	Repetitive behavior	3 children with down syndrome	DRO	Identical to Baseline
3	2016	SIB	4 males and 1 female ages 7-47 years old with ID and other additional conditions	NCR, Differential reinforcement, Response cost	Similar to baseline
4	2015	SIB, crying, task refusal, negative comments	9- and 11-year olds with ASD	Presession pairing	Similar to treatment
5	2014	Bruxism	16-year-old female with ASD	Verbal reprimand	
6	2014	Trichotillomania and skin picking	6-year-old female with ADHD	Habit reversal	
7	2012	Aggression, destruction, vocalizations	Children 6-18 years old with Angelman syndrome	FCT	
8	2010	Aggression, destruction, inappropriate sexual behavior	2 elementary students with developmental disabilities	DRA	Similar to treatment
9	2010	Delusional statements	26-year-old male with mild ID, TBI, frontal lobe syndrome, mood disorder, and mania with delusions	DRA	Similar to treatment
10	2010	Off-task behavior	6-, 8-, and 10-year-old males with ADHD	DRO, EXT	
11	2009	Destruction	2-year-old male with developmental disability and Peter's anomaly	FCT	
12	2009	SIB	26-year-old male with Prader-Willi syndrome	Time-limited bathroom visits, FCT, Differential reinforcement	Similar to treatment
13	2008	Hair pulling	8-year-old female with Cri du Chat syndrome	Differential reinforcement, response interruption, access to toys	Similar to treatment
14	2007	Aggression, destruction, noncompliance	4- and 5-year old males with developmental delays	FCT	Similar to treatment
15	2007	Stereotypy	Children 3-11 years old with ASD	Response interruption, redirection	Similar to treatment
16	2005	Destruction, aggression, disruptive behavior, elopement	4- and 9-year old males, one with developmental disability and the other moderate ID	FCT, Choice making	Similar to treatment
17	2004	Inappropriate verbal behavior	4 adult males with ABI	DRA	Similar to treatment
18	2003	Food selectivity	5-year-old male with ASD	DRA + demand fading	Similar to treatment

19	2003	Food refusal	5-year-old male	DRA, EXT	Similar to treatment
20	2003	SIB, aggression, tantrums			
21	2002	SIB, aggression, disruption	6- to 13-year-olds with various disabilities	EXT, Differential reinforcement	Similar to treatment
22	2002	Aggression	24-year-old male with profound MR	FCT	Similar to treatment
23	2001	Aggression, disrobing, elopement	11-year-old male with ASD	Noncontingent kinesthetic stimulation	
24	2001	Breath holding	16-year-old male with severe ID and cerebral palsy	Reprimand, DRO	Similar to treatment
25	2001	Hair Twirling	2-year-old female	Response prevention	Similar to treatment
26	2000	SIB, aggression, destruction	22-year-old male with fragile X and severe MR, and 9 year-old female with severe MR	NCR	Similar to treatment
27	2000	Destructive behavior	8- and 9- year old males with developmental disabilities and ASD	Sequence choice and DRO	Similar to treatment
28	1999	Aggression, tantrums, overactivity, noncompliance, poor social skills	4-year-old males both with and without developmental disabilities	Specified directions and contingent staff attention	
29	1998	Aggression	7-year-old male with severe ID and PDD	FCT, EXT, Response blocking, alternative form of stimulation	
30	1998	Vocal tic	4-year-old typically developing female	DRO	
31	1998	SIB	7-year-old female with ASD and moderate ID	NCR, EXT, warning stimuli	
32	1998	Disruptive behavior, finger picking	27-year-old female with profound ID and ASD	EXT, DRA	
33	1997	SIB, aggression, destruction	2-5-year-olds with moderate to severe disabilities	FCT	Similar to treatment
34	1997	SIB, aggression, destruction, disruptive body movements	14-year-old female with multiple disabilities	'Multicomponent positive behavior support plan'	
35	1997	SIB	Adults with profound ID	FCT	
36	1997	Aggression	9-year-old male with down syndrome and moderate ID	FCT	Similar to treatment
37	1996	SIB, stereotypy	7-year-old female with Rett-like syndrome and severe MR	Meal schedule and quantify	Similar to treatment
38	1996	SIB	2-year-old male and 7-year-old female with severe developmental delay/ID	Meal schedule	
39	1996	SIB, Aggression	8-year-old males with ASD	FCT	Similar to treatment
40	1996	Disruptive behavior	5-year-old with mild MR		
41	1995	Aggression	31-year-old male with severe MR	FCT, offering choices, rest periods, pre-task requests, backward chaining	

42	1994	SIB	7-,8-. And 12-year old children with moderate to severe MR	Extinction	Similar to treatment
43	1994	SIB, tantrums	11-year-old male student with 'severe emotional disturbance'	Curriculum modification	Similar to treatment
44	1994	SIB and hand-mouthing	3- and 4-year-olds with unspecified disabilities	Enriched environment	Similar to treatment
45	1994	SIB	2 adult women with profound intellectual disability	Non-contingent and contingent protective equipment	
46	1994	SIB, Aggression	5- to 11-year-old children with severe to profound intellectual disability	FCT	Similar to treatment
47	1994	SIB	Adults 22-46 years old with profound ID	NCR, DRO, EXT, antecedent manipulations	
48	1994	Destructive behavior	5-year-old female quadruplets with PDD and ID	NCR	

Table 1. DRA = Differential reinforcement of alternative behavior FCT = Functional communicational training
DRO = Differential reinforcement of other behavior NCR = Non-contingent reinforcement
DNR = Differential negative reinforcement
The missing information (i.e. "Condition of Follow-up") was information that was not specified within the study.

Appendix B: Figures

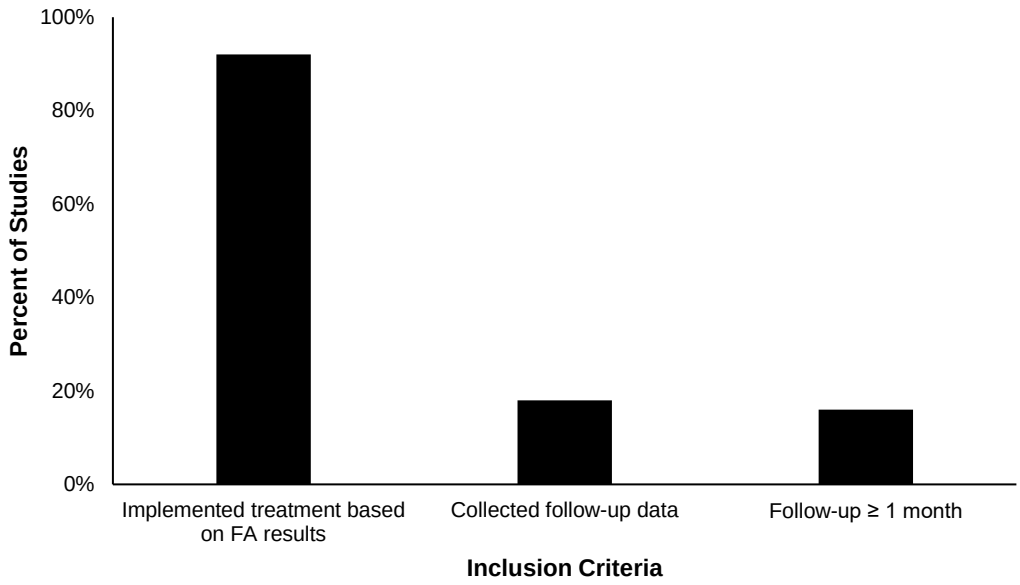

Figure 1. Percent of articles meeting each inclusion criterion.

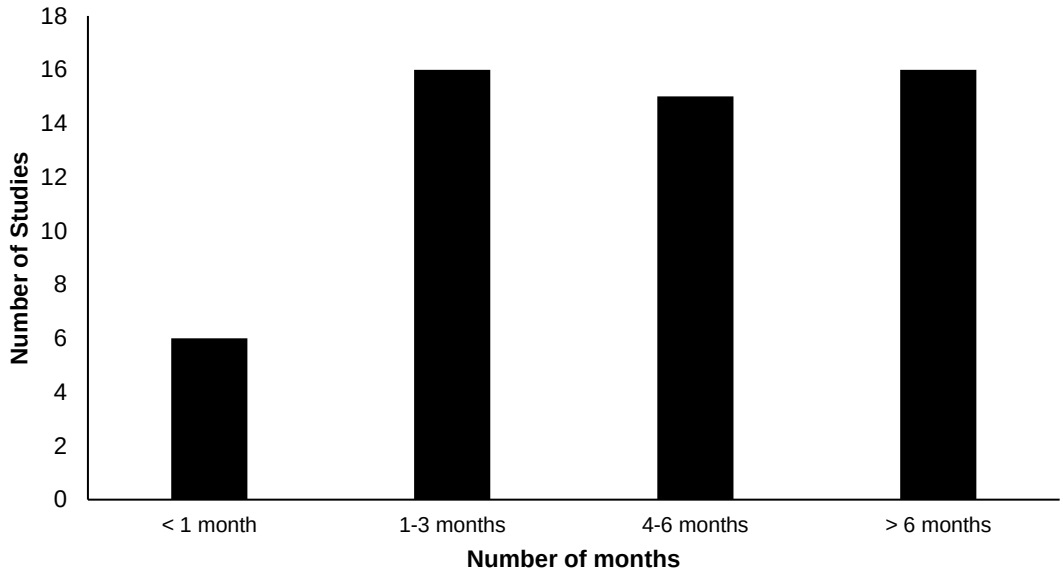

Figure 2. Studies that met inclusion criteria by time between treatment and follow-up.

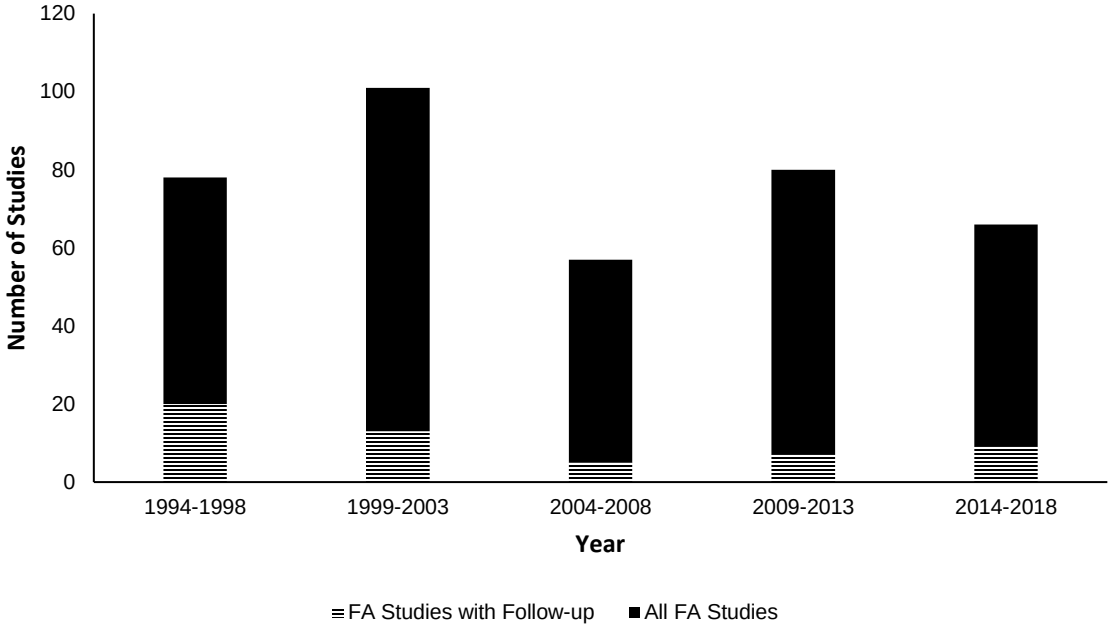

Figure 3. Studies that met inclusion criteria categorized by year published.