The Use of Functional Assessment to Promote Maintenance and Institutionalization of a Performance Management Intervention for Behavior Technicians

Daniel John Cymbal

Follow this and additional works at: https://repository.fit.edu/etd
The Use of Functional Assessment to Promote Maintenance and Institutionalization of a Performance Management Intervention for Behavior Technicians

by

Daniel John Cymbal

A dissertation submitted to the College of Psychology and Liberal Arts of Florida Institute of Technology in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Behavior Analysis

Melbourne, Florida
July, 2023
We the undersigned committee hereby approve the attached dissertation, “The Use of Functional Assessment to Promote Maintenance and Institutionalization of a Performance Management Intervention in Behavior Technicians.”

by

Daniel John Cymbal

David A. Wilder, Ph.D., BCBA-D
Professor
School of Behavior Analysis
Major Advisor

Elbert Blakely, Ph.D., BCBA-D
Assistant Professor
School of Behavior Analysis

Gary N. Burns, Ph.D.
Professor
School of Psychology

Bryon Neff, Ph.D., BCBA-D
Associate Professor
School of Behavior Analysis

Robert A Taylor, Ph.D.
Professor and Dean
College of Psychology and Liberal Arts
Abstract

Title: The Use of Functional Assessment to Promote Maintenance and Institutionalization of a Performance Management Intervention in Behavior Technicians

Author: Daniel John Cymbal

Advisor: David A. Wilder, Ph.D., BCBA-D

In Organizational Behavior Management (OBM), any intervention is judged by its immediate effects, the long-term impact on the targeted performance, and the likelihood that the organization and/or its constituents will adopt the intervention (Sigurðsson & Austin, 2006). However, maintenance and adoption of procedures is rarely measured (Redmon, 1991). A core tenant of behavior analysis is the role of environmental stimuli in maintaining behavior, but most tactics to promote durability are not examined in this fashion (e.g., Conard et al., 2016; Johnston, 1979). A logical, but under investigated, avenue is to examine the role of functional assessment in producing durable change. The current study examined the predictive ability of a common OBM functional assessment tool, the Performance Diagnostic Checklist- Human Services (PDC-HS; Carr et al., 2013; Carr et al., 2016) to influence maintenance and institutionalization of a prescribed intervention to increase learning opportunities provided to clients in session. Two multiple baseline designs across participants were utilized with six behavior technicians. One group received a PDC-HS indicated intervention, a refined process to provide accurate and readily available program stimuli. The other group received the same intervention, in addition to their supervisors receiving their own PDC-HS indicated intervention to support the behavior technician performance with performance feedback and a job aid/checklist. All behavior technicians increased their provided learning opportunities per hour following intervention in both single-case designs. A mixed repeated measures analysis of variance (ANOVA) revealed that the multi-level intervention produced statistically significant
change over baseline compared to the control group. Limitations and future research are discussed.
# Table of Contents

Abstract ................................................................................................................................ iii

List of Figures ....................................................................................................................... ix

List of Tables ...................................................................................................................... xi

Acknowledgement ............................................................................................................... xii

Dedication .......................................................................................................................... xiii

Chapter 1 Introduction ........................................................................................................... 1
   Conceptual Foundations of Response Maintenance and Related Concepts in ABA/OBM .......................................................................................................................................... 3
   Applied Definitions of Maintenance ................................................................................. 5
   Generalization and Generality .......................................................................................... 6
   Institutionalization ............................................................................................................ 8

Chapter 2 Evaluation of Maintenance in Behavior Analytic Literature: Methods, Data, and Derived Technologies .......................................................................................... 10
   General Strategies to Promote Generalization of Interventions ...................................... 11
   Strategies to Promote Maintenance ................................................................................ 16
   In Behavioral Safety ....................................................................................................... 18
   In Human Service Settings ............................................................................................. 23
   General Strategies to Promote Institutionalization of Interventions ............................... 26
   Proposal of an Omnibus Term ........................................................................................ 29

Chapter 3 Barriers in the Study of Maintenance and Institutionalization ............................ 33
   Potential Broad Sources of Conceptual and Technological Drift ................................... 35
   ABA and OBM Differences ............................................................................................. 37
List of Figures

Figure 1: Bar graphs depicting the results of the PDC-HS, administered to clinical directors, regarding variables impeding frequency of learning opportunities provided and recorded during session by behavior technicians.............................................................177

Figure 2: Bar graphs depicting the results of the PDC-HS, administered to clinical directors, regarding variables impeding supervisors from ensuring that behavior technicians had adequate materials to provide frequent learning opportunities...........178

Figure 3: Graph depicting the performance of Group A participants providing learning opportunities in consecutive daily sessions when provided the resource and material intervention. .......................................................................................................................179

Figure 4: Graph depicting the weekly average performance of Group A participants providing learning opportunities when provided the resource and material intervention. ........................................................................................................................................... 180

Figure 5: Graph depicting the isolated performance of Participant 1 in Group A providing learning opportunities in consecutive daily sessions when provided the resource and material intervention. .......................................................................................................................181

Figure 6: Graph depicting the isolated performance of Participant 2 in Group A providing learning opportunities in consecutive daily sessions when provided the resource and material intervention. .......................................................................................................................182

Figure 7: Graph depicting the isolated performance of Participant 3 in Group A providing learning opportunities in consecutive daily sessions when provided the resource and material intervention. .......................................................................................................................183

Figure 8: Graph depicting the performance of Group B participants providing learning opportunities in consecutive daily sessions when provided the resource/material intervention and when their supervisors received bi-weekly feedback on maintaining the intervention. ........................................................................................................................................... 184

Figure 9: Graph depicting the weekly average performance of Group B participants providing learning opportunities when provided the resource/material intervention and when their supervisors received bi-weekly feedback on maintaining the intervention.
Figure 10: Graph depicting the isolated performance of Participant 1 in Group B providing learning opportunities in consecutive daily sessions when provided the resource/material intervention and when their supervisor received bi-weekly feedback on maintaining the intervention.

Figure 11: Graph depicting the isolated performance of Participant 2 in Group B providing learning opportunities in consecutive daily sessions when provided the resource/material intervention and when their supervisor received bi-weekly feedback on maintaining the intervention.

Figure 12: Graph depicting the isolated performance of Participant 3 in Group B providing learning opportunities in consecutive daily sessions when provided the resource/material intervention and when their supervisor received bi-weekly feedback on maintaining the intervention.

Figure 13: Graph depicting the average performance of each group’s supervisors’ fidelity in maintaining the resource/material intervention for behavior technicians. This includes the baseline average and weeks 2-5 of the intervention phase. Week 1 is not pictured as this was when the researcher arranged the resources and materials for the behavior technicians (i.e., fidelity was 100%).

Figure 14: Graph depicting the effect of intervention group on mean changes in learning opportunities per hour.

Figure 15: Graph depicting the results of the procedural acceptability surveys for behavior technicians participating in the single-case designs (i.e., Group A and B) and their average responses to each item.
List of Tables

Table 1: Mixed Design ANOVA Results ................................................................. 185
Acknowledgement

It would be inappropriate to begin this acknowledgement without first thanking my advisor, Dr. David Wilder. Your support and guidance throughout the program has been invaluable and I hope that the end of this chapter merely marks the beginning of a longstanding connection as colleagues. I also owe a great deal to my committee. Dr. Eb Blakley has been there from the beginning of my time at Florida Tech and always made sure I never forgot the fundamentals. Dr. Bryon Neff, thank you for your encouragement, kind feedback, and making this process far less aversive than it might have been. Finally, Dr. Gary Burns, thank you for always letting me hound you through text or email and taking your time to answer my novice questions about stats.

I also need to thank the faculty, past and present, at Florida Tech, for all serving as mentors in some aspect of my life throughout this journey. I’d like to mention a few in particular. First, Drs. Jose Martinez-Diaz and Josh Pritchard got me started on this track what seems like an eternity ago. Drs. Gravina, Ferguson (née Tilka), and Weatherly all nursed my interest and passion for OBM. Dr. Nicholson not only taught me a great deal about behavior analysis, but also parenting while finishing a doctoral program. In addition, I cannot forget all the colleagues and friends I’ve made along the way. I could fill another page talking about you all, but I will abridge it by expressing my sincere excitement to continue growing and collaborating with you all.

Last, I have to issue to an apology to my family and friends. Although it seemed for awhile I had given up on higher education, I appreciate your patience and understanding as I dove into this headfirst. Specifically, to Mom and Bob, Dad and Cindy, and Chris, thanks for your encouragement and support along the way (and in life). Last, but not least, thanks to Aaron, for always being a better friend to me than I am to you.
Dedication

I dedicate this to Sam, Ander, and Kieran. Sam, you are always my biggest cheerleader and have been here through the (many) ups and downs of this process. I couldn’t have done it without you and wouldn’t have wanted to. I’m looking forward to the next chapter immensely; sorry it took so long. I love you all.
Chapter 1
Introduction

Applied Behavior Analysis (ABA) is often associated with its therapeutic applications (i.e., treatment of autism and other developmental delays). However, the broader science of behavior analysis also subsumes a variety of sub-specialties, including Organizational Behavior Management (OBM). OBM practitioners utilize behavior analytic principles to influence performance of employees and organizations (Wilder et al., 2009). OBM remains a seemingly less common application of behavior analysis, at least in terms of formal certification. OBM practitioners are estimated to comprise less than one percent of behavior analysts’ common practice areas (Behavior Analyst Certification Board, n.d.). However, in 2022, approximately 60% of the United States population was employed (Bureau of Labor Statistics, 2022). Therefore, while there is not a 1:1 need of practitioners to employees, there seems to be ample opportunity for growth. The promulgation of behaviorism-based practices, especially in non-traditional areas, will likely be reliant on production of effective, practical technologies and cultivation of consumer support, over and above demonstration of scientific rigor (Alligood & Gravina, 2020; Conard et al., 2016).

An important indicator of effectiveness in troubleshooting problems and influencing associated behavior change is that effects are persistent once the formal intervention or consultation ceases. In fact, this is noted as a critical measure of behavioral programs’ success, subordinate only to the initial robustness of the intervention’s effects.
Generality of behavior changes across settings, people, and time are part and parcel of ABA’s oft-referenced parameters (i.e., Baer et al., 1968) and techniques to promote the likelihood that therapeutic intervention effects sustain or extend to other settings have been discussed frequently in behavior analytic literature (e.g., Osnes & Lieblein, 2003; Stokes & Baer, 1977; Stokes & Osnes, 1989). Durable behavior change across time, often referred to as response maintenance (e.g., Foxx, 2013), is critically important to social validation and widespread adoption of behavior analytic procedures. The importance of durability in OBM intervention is evident when considering commonly treated problems (e.g., employee safety and health concerns, productivity and quality issues, and dissatisfied consumers; VanStelle et al., 2012). Unchecked, these problems might lead to a variety of adverse outcomes. Failed initiatives, in addition to reduced confidence in behavior analytic procedures, could lead to the dissolution of businesses (i.e., weak results of quality and productivity improvements) or result in workplace accidents and death (e.g., a short-lived safety intervention). Thus, OBM literature mirrors therapeutic ABA literature in this regard, stressing the importance of designing intervention to facilitate lasting change (McGee & Crowley-Koch, 2021; Sigurdsson and Austin, 2006).

Unfortunately, one of the historical foci of most behavioral analytic research is on consistency of intervention implementation (i.e., procedures pertinent to research), with less analysis and planning given to the adoption of technologies and maintenance of intervention effects. Redmon (1991) informally reviewed research involving behavioral consultation and identified a tendency to focus primarily on evaluation of implemented interventions while generally reporting less on pre-intervention and follow-up, a trend that
has persisted more recently in OBM literature. Empirical evaluation of maintenance continues to be ephemeral and infrequent. In the Journal of Organizational Behavior Management (JOBM), the flagship OBM publication, follow-up to formal intervention typically occurred in close proximity to intervention termination (i.e., within 3 months) and was measured in fewer than one-third of empirical studies published (VanStelle et al., 2012). While production of response maintenance is a concern germane to all OBM practitioners, it remains a relatively under investigated topic. Sasson et al. (2006) suggest that when gaps in the research are observed, that there are two potential explanations: the specified phenomenon is of little pragmatic use or an adequate research protocol has yet to be extrapolated from theory. Thus, the characterization of response maintenance as an ABA mandate (i.e., Baer et al., 1968) juxtaposed with relative dearth of evaluation (e.g., Conard et al., 2016; Sigurdsson & Austin, 2006) suggests the latter. There is likely significant room to expound on the concept within key areas of contribution in OBM (i.e., concepts and theory, behavior analytic methodology, empirical data, and replicable technologies; Riley & Frederikson, 1984).

Conceptual Foundations of Response Maintenance and Related Concepts in ABA/OBM

The term maintenance has existed from the earliest iterations of the operant conditioning paradigm, though with a different connotation. Skinner (1953, p. 98) described measurement of maintenance effects as a key contrast to traditional approaches to learning wherein topographical outcomes (i.e., skill acquisition) were the primary variables of interest. Outcome measures which merely assert whether a skill is present or
not, lack sensitivity and resultant explanatory power. That is, when a particular behavior was first emitted at strength does not explain why it occurred, recurs, or ceases. In contrast, Skinner suggests that an operant conditioning perspective, with its characteristic rate of response measurement, yields insight into the past and current reinforcement contingencies, providing an ongoing and causal account. Therefore, most early mentions of maintenance referred more to a probabilistic response rate as a function of environmental contingencies. Description of sustained responding comprised response patterns produced under various reinforcement schedules and operant extinction (e.g., Skinner, 1953, p. 99-104; Skinner, 1969).

However, as early scientific endeavors in behavior analysis gave rise to the promise of a wealth of applied behavioral technologies, subsequent applications to behaviors of social significance prompted a focus on the qualitative nature of outcomes. In their seminal article, Baer et al. (1968) elaborated on potential societal implications of an applied behavioral science and circumscribed the boundaries of ABA. Whereas basic research needs to foremost fulfill standards of scientific rigor, appraising an ABA study that influenced the lives of its participants would, presumably, require additional and different metrics. Generality across settings, behaviors, or time was one of seven necessary characteristics described by the authors and became the basis for many applied formulations of describing response maintenance. In most ABA and OBM literature, this
discussion tends to center around a few key terms: maintenance, generalization, generality, and institutionalization.

**Applied Definitions of Maintenance**

Most modern definitions of maintenance are oriented toward applied research and service delivery contexts (e.g., post-intervention continuity of behavior; Foxx, 2013). Response maintenance can potentially be asserted as soon as the behavior reoccurs following initial establishment (e.g., Malott & Suarez, 2003) or when levels of behavior sustain once intervention conditions have ceased or formal implementers have left (e.g., Boyce & Geller, 2001; Cooper et al., 2007, p. 305; Foxx, 2013). Variations of the latter remain among the most common definitions. Thus, maintenance is most frequently described in the applied literature and graduate-level textbooks temporally (i.e., post-treatment) and topographically (i.e., a stable pattern of responding). Yet, despite this implicit temporal criterion (i.e., that the effects must persist for some time past intervention), there is no formally agreed upon duration for what constitutes demonstration of durability, though some authors have set arbitrary criteria to delineate the boundaries of categories in reviews (e.g., 2 weeks or greater; Boyce & Geller, 2001). Evaluations vary greatly in ABA and OBM literature. Measurement following intervention can be as short as a single session to years following the cessation of a consultant’s services (e.g., Osnes & Lieblein, 2003; VanStelle et al., 2012). The duration of follow-up phases is almost always
variable, and the only definitive component is that the consultant or researcher relationship will, invariably, end (Sigurdsson & Austin, 2006).

Generalization and Generality

Another concern in producing the behavior change across time is ensuring that the behavior of interest proliferates outside of any artificial training or intervention conditions. Historically, this has led to discussion of response maintenance in much of the applied literature pertaining to generalization. In other words, training or intervention conditions are either implicitly embedded with or explicitly programmed for in such a manner that the behavior occurs in other similar contexts or environments. The targeted behavior is then said to “generalize,” meaning transfer of control to non-training environments and naturally occurring consequences that maintain ongoing responding.

Generalization, as a term, originated in the description of basic research phenomena. Skinner (1953) described generalization specifically in relation to discrimination procedures. An organism does not generalize learned stimuli, rather this refers to the transfer of stimulus control to other stimuli with similar formal properties. This comprises the basis of what is commonly referred to as stimulus generalization. Generalization was typically measured in the lab with intense precision where the gradient of stimulus control could be clearly specified (Stokes & Osnes, 1989). Other common forms of generalization include the emergence of other functionally equivalent topographies (i.e., response generalization; response induction). Stokes and Baer (1977) noted that, historically, behavior technology was generally built around stringent teaching procedures which do not promote stimulus generalization but, rather, preclude its
occurrence. The absence of the phenomena was a sign of strong usage of differential reinforcement procedures. In this sense, discrimination was most often identified as the explicit planned occurrence, while generalization occurred tacitly or as an undesired byproduct of intervention. Yet, demonstration of this phenomena need not be observed as a “passive” effect of intervention, nor should it be considered a function of poor teaching procedures.

Instead, generalization, Stokes and Baer (1977) contend, should be utilized as a tool to assure the benefit of therapeutic change across settings, subject, and time. In other words, generalization should be actively harnessed through direct programming. Stokes and Baer reviewed research literature and devised nine techniques a posteriori that promote generalization. However, the spread of training effects into the natural environment (i.e., across time, people, and settings), while a prerequisite to successful behavioral intervention, is also not directly equivalent to the initial conceptual formulation of generalization. Rather, it is a “pragmatic” (Stokes & Baer, 1977, p. 350) and topographical description, with less regard to the functional relations that produce maintenance and generalization (Stokes & Osnes, 1989). Taking the latter concern into account, Stokes and Osnes refined and grouped the original taxonomy of generalization strategies according to various behavioral principles (e.g., varying elements of stimulus control, reinforcement, and schedules of reinforcement). Even though the term “generalization”, in this sense, was used non-technically (Stokes & Baer, 1977, p. 350) and, perhaps, inaccurately describes
the phenomenon (Johnston, 1979), technologies to promote generalization nonetheless comprise multiple tactics also purported to promote response maintenance.

The term generality, sometimes used interchangeably with generalization, may be the more apt term to describe the continuation of intervention effects across different contexts. Generality, described as the “universality or replicability” (Johnston, 1979, p. 3) of an observed effect, can be elaborated in multiple ways. Investigation of generality can include evaluation of the universality of functional relationships (e.g., variables, methods, and processes) as well as examination of the functional relationships across various dimensions (e.g., species, subjects, responses, settings). As defined, generality of a behavioral technology is then of key importance to the extension or maintenance of behavior interventions. That is, if research has carefully evaluated an intervention, distilled the essential functional components of its implementation across various dimensions, then it is much more likely to be successful without reliance on additional, probabilistic, and, sometimes, artificial strategies.

Institutionalization

One commonly identified variable critical to the generality of any ongoing OBM intervention is ensuring that the management of the intervention or behavior transfers from researcher to workplace managers and other stakeholders (Sigurdsson & Austin, 2006). If the organization continues to maintain some portion of the intervention following the cessation of formal researcher or consultant involvement, this is commonly referred to as institutionalization (Boyce & Geller, 2001; Conard et al., 2016; Sigurdsson & Austin, 2006). Maintenance and institutionalization are naturally related concepts, in that the
former specifies a pattern of responding after intervention, while the latter denotes a likely predictor of that pattern (e.g., Sigurdsson & Austin, 2006). However, the specifics of that relationship are not clearly defined. Boyce and Roman (2002) suggest that the relationship may function bidirectionally, as patterns of maintenance could also result in organizational adoption of interventions without formal planning of institutionalization. In other words, if a successful intervention is chosen (i.e., the effects maintain), then the organization might be more likely to adopt and continue utilizing its components. Others too have argued that distinguishing between the two terms in not necessarily useful. Kessler (2001) posits that because both maintenance and institutionalization provide lasting behavioral change, then the distinction is unclear or, perhaps, unnecessary. However, a delineation between the two can be extrapolated. Conard et al. (2016) note that there are instances wherein maintenance can be observed without institutionalization, specifically when the intervention is antecedent based (e.g., training). An employee could presumably contact already existing environmental consequences in the workplace once they acquire and reliably emit an absent skill or repertoire. In such a case, environmental variables (e.g., managers) need not institute any change. Maintenance then only requires consistent change of a behavior or behaviors of interest, whereas institutionalization, definitionally, requires change in multiple behaviors and behavers. So, while maintenance, generalization, and institutionalization, terms typically referenced regarding the generality of ABA interventions, have been commonly debated conceptually, consensus dictates that each should be explicitly planned for and programmed (e.g., Boyce & Roman, 2002; Conard et al., 2016; Sigurdsson and Austin, 2006; Stokes & Baer, 1977).
Chapter 2
Evaluation of Maintenance in Behavior Analytic Literature: Methods, Data, and Derived Technologies

Derivation of tactics to promote maintenance in ABA literature typically comprises post-hoc evaluation of studies’ methods (e.g., implementation of the independent variable) and outcomes (i.e., measurement of behavior post-intervention) to identify patterns. From those patterns, various formulations of strategies and technologies have been proposed. Thus, there are two notable features regarding the methods used to devise procedural promotion of maintenance. First, there is a relative lack of empirical evaluation as to which derived tactics, either utilized alone or in combinatory fashion, promote generalized and/or sustained behavior change most efficiently and effectively. For example, in JOBM, some commonly referenced tactics (e.g., introduce to natural contingencies) have been sparsely utilized (Conard et al., 2016). Second, there is a frequent reliance on citing multiple strategies that were, at least initially, explicitly intended to promote generalization of applied intervention outcomes (e.g., Stokes & Baer, 1977). In fact, many of the strategies typically cited to produce maintenance are identical, perhaps due to the role that response maintenance plays in evaluating the success of generalization training (Foxx, 2013). Therefore, identification of the variables historically ascribed to promote response
maintenance requires examination of the technology developed to promote generalization, much of which has been drawn from the wells of therapeutic ABA intervention.

General Strategies to Promote Generalization of Interventions

Stokes and Baer (1977) reviewed over one-hundred studies and derived nine basic categories under which intervention planning fell to promote generalized outcomes. Those categories were: train and hope, sequential modification, introducing natural contingencies, train sufficient exemplars, train loosely, using indiscriminable contingencies, programming common stimuli, mediate generalization, and train to generalize. The first two tactics, train and hope and sequential modification, were both the most frequently observed and what the authors deem the “least analytical,” in that the researchers do not actively program for generalization (Stokes & Baer, 1977, p. 363). Train and hope refers to a non-strategy of refraining from planning for generalization, a tactic apparent in nearly half of the studies they reviewed. Procedurally, this was characterized as researchers recording a series of post-treatment probes, hopeful to demonstrate that generalization had, somehow, occurred without any formal treatment components or subsequent modifications to produce generalized outcomes. Sequential modification refers to, in the absence of observed and measured generalization following treatment, systematically varying intervention components in non-generalized conditions. In other words, if the behavior change was not observed within certain settings, then the researcher would introduce additional intervention and support in the non-generalized conditions and re-test. The remaining
seven strategies imply active programming of teaching procedures to produce
generalization.

Introduction to natural reinforcement contingencies, described as the “most
dependable” of the strategies (Stokes & Baer, 1977, p. 353) suggests introducing the
intervention in some way that already present environmental contingencies will assume
control over the behavior rather than formal intervention. Introducing natural contingencies
may include both selecting behaviors that are likely to contact reinforcement (e.g., social
behavior of children) as well as altering the environment to modify the rate of naturally
occurring contingencies (e.g., teaching subjects to solicit social praise or teaching subjects
to deter maladaptive behavior). Training sufficient exemplars refers to teaching as many
instances of a typical training condition as possible to produce generalization. Researchers
using this tactic typically trained enough members of a response or stimulus class that
generalization was demonstrated with non-training conditions or stimuli. Training loosely
refers to modifying the discrimination paradigm in such a way that multiple stimuli and
response variations are permitted so that stimulus control may transfer more readily to a
fuller range of relevant situations. Using indiscriminable contingencies refers to the use of
reinforcement schedules wherein discriminative stimuli or response requirements are not
necessarily salient to the subject. Functionally, this technique can produce results by
capitalizing on the use of intermittent schedules to increase resistance to extinction of the
programed behavior or make use of obfuscated or delayed reinforcement contingencies to
maintain other desirable behaviors or behaviors occurring outside of treatment conditions.
Programming common stimuli means that training should occur using stimuli present in
the environment that is the target of generalization. Studies that incorporated this technique sometimes used stimuli with other forms of functional control in the non-training environment (e.g., peers) as well as stimuli to ensure the setting resembled the target environment (e.g., making a treatment room appear like a regular classroom, using common academic material). Studies that mediate generalization are those that teach responses that can be utilized in a variety of conditions. Generally, this refers to capitalizing on human language (e.g., rules) in concert with intervention to increase the likelihood that behaviors are utilized in multiple conditions or settings. Stokes and Baer note that such mediation using language has been most prevalent in research that examined self-management and self-control techniques. Training to generalize essentially means treating generalization as though it were a response to be reinforced. Essentially, training should provide opportunities for the subject to contact programmed consequences for generalization (e.g., teaching a general example of a concept and asking a student for another example).

Stokes and Osnes (1989) later extended the initial analysis of Stokes and Baer (1977) noting two significant weaknesses of the original article were failing to specify functional relationships responsible for generalization and creating categories of tactics that were reflective of practice rather than conceptual principles. Their restructured and refinement of the taxonomy of generalization strategies omitted tactics that were essentially not active modifications (i.e., train and hope) or represented action typically taken within single subject experimental designs (i.e., sequential modification). Stokes and Osnes further explicated some of the original categories and grouped techniques under
three distinct principles: exploiting current functional contingencies, training diversely, and incorporate functional mediators. Each principle contained four different strategies either derived or elaborated from Stokes and Baer (1977) and supported each with references to research. Exploiting current functional contingencies denoted three strategies (i.e., contact natural consequences, recruit natural consequences, and modify maladaptive consequences) that were elaborated from introduction to natural contingencies. The fourth tactic in this category, reinforce occurrences of generalization, was an adaptation of “train to generalize.” In short, these four tactics suggest training conditions that promote responding likely to contact consequences already present in the environment or modification of the environment to increase the probability of reinforcing the targeted response or generalized responding. The second principle, training diversely, combines three distinct categories from Stokes and Baer (1977): use of indiscriminable contingencies, training loosely, and training sufficient exemplars. Stokes and Osnes (1989) suggest four specific tactics to train diversely, including use of an adequate amount of exemplar stimuli and training an adequate amount of response forms (i.e., training multiple exemplars) as well as varying evocative stimuli used in training to promote loose discriminative control (i.e., training loosely), and utilizing intermittent or unclear (i.e., indiscriminable) reinforcement contingencies. The third and final principle, incorporating functional mediators, elaborates on two categories from Stokes and Baer (1977), program common stimuli and mediate generalization. Program common stimuli is split into two tactics, the use of common environmental stimuli during training (e.g., physical items) and the use of common social stimuli (e.g., peers, teachers) in the training setting. Mediating generalization is split similarly by the use of self-mediated physical stimuli (e.g.,
notebooks or commitment cards) and the utilization of self-mediated verbal or covert stimuli (e.g., goal setting, rules).

Osnes and Lieblein (2003) conducted a follow up literature review to Stokes & Baer (1977) and Stokes and Osnes (1989) to assess the degree to which generalization and maintenance had become foci for the research literature. The authors reviewed ninety-three articles from behavior analytic journals that alluded to or implemented generalization and maintenance procedures. Within their search, they identified four review articles that investigated generalization and maintenance programming, all focused on interventions to promote social behavior. Two of the four reviews (i.e., Chandler et al., 1992; Landrum & Lloyd, 1992) incorporated some appraisal of the generalization techniques promoted in Stokes and Baer (1977) as well as Stokes and Osnes (1989). Specifically, Chandler et al. (1992) found that studies they reviewed predominantly adopted a combination of the three generalization principles (e.g., exploiting current functional contingencies, training diversely, and incorporating functional mediators; Stokes & Osnes, 1989). Of the eighty-eight research articles identified, Osnes and Lieblein (2003) reported that thirty articles specifically addressed generalization and eight focused on maintenance. The authors mainly concluded that, while use of the strategies outlined in Stokes and Baer (1997) and Stokes and Osnes (1989) appear to be of increasing interest and utility to the research community, most studies continue to err on the side of experimental rigor, most often
demonstrating control of the dependent variable using highly discriminable contingencies and avoiding conditions that might promote durability and generalization.

**Strategies to Promote Maintenance**

Interestingly, Osnes and Lieblein (2003) found in their review that most articles wherein generalization was a desired outcome contained some facet of explicit programming for generalization. Conversely, most articles addressing maintenance merely assessed the continuation of intervention effects following withdrawal. Of the four articles that actively designed interventions to promote maintenance, all utilized strategies related to one generalization principle identified by Stokes and Osnes (1989): exploit natural functional contingencies. While the assertion that maintenance is not typically programmed for was unsurprising to the authors, the suggestion that only one of the proposed Stokes and Osnes’s principles for generalization may be pertinent to maintenance programming is important. Many strategies to promote maintenance have limited empirical support (Foxx, 2013) and the term natural functional contingencies is broad. In any given environment, functional contingencies are potentially innumerable and can entail a variety of stimuli with that influence behavior. Therefore, there are likely a variety of potential idiosyncratic factors that influence the maintenance of behavioral interventions.

Foxx (2013) attempted to identify some of these factors by reviewing previous research efforts that promoted therapeutic behavior change and conducted long term follow-up measurement. In a review of studies that addressed addictive behaviors of participants (e.g., nicotine dependence and excess caffeine consumption) and followed up
progress at various stages (i.e., 18, 30, and 40 months), the critical components of maintaining change appeared to be resultant of multiple functional contingencies, including positive feedback, antecedents specifying deleterious delayed outcomes (e.g., information regarding bodily harm caused by excess consumption), aversive contingencies (i.e., social pressure), and covert contingencies (e.g., self-management). In four studies addressing social behavior (i.e., vocational, situational, and sexual social skills), Foxx again found that a variety of potential factors may have contributed to maintenance. Participants’ behavior that maintained after eighteen months seemed most maintained by extant social contingencies in their environment, moderated by variables such as individual skills, maturation, and the nature of the skills taught (i.e., some vocational skills seemed to have more naturally occurring opportunities than other types of skills). Three studies comprising six participants designed to evaluate naturalistic teaching procedures for language were also revisited several years following training. Effects from these studies were hypothesized to have endured due to a focus on teaching attention to cues in the natural environment, use of test conditions with novel personnel, and systematic fading of training as well as thinning of reinforcement schedules. However, the relative contribution of each component was debatable, due to the packaged nature of interventions.

In a ten-year follow-up of treatment of maladaptive behavior, evaluation of reduction interventions was conducted on eight previously institutionalized clients, demonstrating mixed maintenance outcomes. Generally, Foxx observed that treatment seemed to endure best with clients that demonstrated the most expressive language skills. This effect was likely due to the increased access to social reinforcers following reduction
of problem behavior (e.g., being able to communicate wants and needs, attend social functions and activities). Additional idiosyncratic variables also contributed to the ongoing success of treatment, including topography of the behavior and the time elapsed between treatment and reoccurrence. In one instance, functional assessment and ongoing development of treatment had occurred in the intervening decade between treatment and follow-up measurement. In sum, Foxx suggested that the best way to promote maintenance of therapeutic interventions was through ongoing assessment, systematic analysis of contingencies, and continuous measurement of outcomes.

In Behavioral Safety

The same components are arguably more critical in OBM interventions, wherein the behaviors of interest are typically complex, verbally mediated, and measured, not only individually, but as an aggregate (i.e., group and organizational performance; Alvero et al., 2001). A variety of contextual variables, including those proximate to the performance (e.g., antecedent stimuli, reinforcers), the nature of the response (e.g., topography, fluency), and the functional components (i.e., contingencies within the system) are critical to consider when cultivating lasting effects (Sulzer-Azaroff, 1990). Thus, OBM literature evaluating the production of sustained intervention outcomes has highlighted the importance of analyzing the sum of variables relevant to the targeted behaviors or performances. The value of behavior changes of repertoires at multiple organizational levels is no more apparent in studies involving Behavior-Based Safety (BBS; Agnew & Daniels, 2011), wherein safety outcomes (i.e., an increase safe behavior and an absence of at-risk behaviors and workplace accidents) are most valid when consistently managed by
leadership and the benefits experienced by a significant portion of the front-line employees.

Alavosius and Sulzer-Azaroff (1990) demonstrated maintenance effects following a feedback intervention designed to increase appropriate and safe patient movement by staff in a residential care setting. Feedback improved staff performance in all conditions, with fastest acquisition shown for response classes when feedback was delivered on a near continuous schedule. Following intervention, maintenance was assessed for over seven months absent feedback and performances maintained above baseline for all participants, regardless of previous feedback modality and absent ongoing researcher involvement. The authors suggested that these improvements were likely maintained by natural consequences resultant of the safer procedures (e.g., better patient cooperation, lessened risk of physical injury to staff with proper form, and faster, less effortful completion of job duties). In addition, staff also reported that they had begun adopting the practice of giving feedback to others on their performances. Thus, the behavior generated potential reinforcing contingencies and staff internalized practices of the intervention.

Such findings are characteristic of studies that measure maintenance in behavioral safety. Grindle et al. (2000) reviewed eighteen behavioral safety studies in manufacturing settings, identifying effects associated with common interventions and impacts on injury rates as well as other benefits (e.g., cost). Of the initial studies, six noted information about maintenance of the safety programs, with four of those six mentioning that safety programs were retained and continued by their respective organizations. The authors suggest that these four programs were maintained due to a pre-intervention systems analysis to ensure
that the components of the safety program functioned efficiently within the current organizational structure. In addition, despite the discontinuation of the safety programs in the other two studies, safety (i.e., safe behaviors and results absent incidents) remained consistent during follow-up measurement in all six studies, though the evaluation period varied and, in one study, went unreported. Grindle et al. suggest that the ongoing maintenance of safety, regardless of formal intervention program, introduced new organizational norms. Because policies were established that supervisors had to enforce, this led to employees contacting a variety of natural contingencies (e.g., earplug use preserves hearing) and the creation of conditioned reinforcers (e.g., permanent products resulting from targeted behaviors).

Boyce and Geller (2001) reviewed a variety of occupational and traffic safety studies that assessed the maintenance of intervention effects. The authors discuss four of the generalization techniques as described by Stokes and Baer (1977) that they suggest are likely to promote maintenance of intervention effects: training loosely (e.g., programming multiple evocative stimuli), using indiscriminable contingencies (i.e., intermittent or delayed reinforcement of target behavior), using common stimuli (e.g., training using stimuli present in the common environment), and mediated generalization (e.g., rule-governed or indirect-acting contingencies). The identified studies were diverse in methodology, evaluated through a variety of single-subject experiments, most commonly reversal and multiple baseline designs, with multiple distinct interventions utilized, including written commitment of behavior change, token economies, incentives/rewards, and performance feedback. In each group of interventions, the authors attempted to distill...
the essential procedures that produced maintenance. For example, in studies that utilized written commitment strategies, those that demonstrated maintenance of intervention typically included a signed commitment, customization of the program to the organizational culture, employee participation in intervention, and a targeted response that required relatively low effort. Moreover, most commitment programs were implemented visibly within their respective organizations. Boyce and Geller inferred that these strategies likely were effective in the long-term predominantly due to mediated generalization. Essentially, these strategies led to the development of rules that are likely to translate into powerful social or cultural contingencies. Similar mechanisms were also posited as a key feature in maintaining a token economy intervention, in which tokens were delivered by managers contingent on safe behavior and the daily value of tokens was determined by the percentage of employees engaging in the specified topography (i.e., earplug use).

Incentive reward programs and performance feedback as safety interventions were also scrutinized. Boyce and Geller (2001) observed that most successful incentive/reward programs reviewed were short in intervention length, highly visible to employees in the organizations, utilized relatively thin schedules of reinforcement (i.e., odds of reward were low), and incorporated willing employees into the intervention. Similar procedures were also suggested as the main determinants of the successfully maintained feedback interventions. Lasting feedback interventions included in the review generally targeted multiple response topographies while researchers provided rationale for the intervention, used a mixture of non-specific positive and corrective feedback, utilized employees to deliver feedback, and instituted other intervention components (e.g., training and goal
setting). Based on the lean schedules of reinforcement and participative intervention processes, the authors hypothesized that these programs were maintained because of indiscriminable contingencies and the use of common stimuli (i.e., onsite implementers). Low probability of contacting rewards and the non-specific feedback potentially made specific contingencies less discriminable, while on-site implementers led to transfer of stimulus control for safe behaviors. In sum, Boyce and Geller (2001) suggest that the most durable interventions are those that include consequences that are minimally effective as reinforcers (e.g., intermittent or delayed reinforcement schedules), use feedback that is relatively non-specific by either targeting multiple behaviors, or are designed to generate rules that specify effective contingencies (e.g., providing rationale for an intervention strategy).

Myers et al. (2010) further elaborated on the variables that might maintain a behavioral safety program, with an evaluation of a safety process that was observed for fifteen years post-intervention in an oil refinery. Intervention comprised a safety process with multiple intervention components. Researchers conducted initial interviews with all employees as well as a comprehensive assessment to examine current safety initiatives, solicited input from employees, identified major areas for risk, and facilitated managerial support for intervention. This was followed with employee involvement in the creation of safety values, pinpointing of value related behaviors, and identifying contexts where espousal of those values was critical. Additional intervention components included training, safety observations, positive feedback and rewards, with ongoing meetings and measurement of safety data. While these interventions were initially piloted in the assessed
high-risk area (i.e., power thermal), eventually the process was rolled out across other areas of operation. Overall, the rate of incidents and their severity remained well below industry average in the years following formal intervention and translated into significant cost savings on worker compensation for the refinery. While the main conceit of the study was to evaluate the use of a values-based safety process, Myers et al. noted that the process likely maintained for so long due to the involvement of staff in the design of intervention, the ongoing measurement and data collection, and establishment of consequences maintained by refinery employees. Due to the participation and collaboration with refinery employees, intervention components (i.e., safety observations) were trained to all personnel which led to training and ongoing dissemination of practices. Further studies have corroborated and quantified the impact of employee participation on maintenance of safety programs. For example, Hagge et al. (2017), while not explicitly investigating maintenance, found that only thirty percent of employees participating in a BBS program led to a 50% reduction in accident rates in a coal mine for seven years following intervention.

In Human Service Settings

However, industries that operate in hazardous environments have not been the only settings to demonstrate long term intervention effects. Human services settings, which includes ABA service providers, are notorious for issues with staff attrition and burnout (e.g., Cymbal et al., 2021; Kazemi et al., 2015). As a result, they are a common setting for OBM intervention (Gravina et al., 2018). Christian (1984) provides a case study of one such example, conducted at a large non-profit program for children with autism and other
developmental disorders. The project was implemented over the course of five years and, similar to the large-scale safety studies (e.g., Myers et al., 2010), comprised multiple components including: creation of a two-year plan of goals and objectives in conjunction with organizational leadership, clarification and establishment of organizational structure, development of work roles and process development, recruitment and training of staff and other personnel, seeking out financial resources and support, as well as ongoing evaluation of the program, assurance of compliance with legal standards, and dissemination of information. Christian reported progress across sixteen different dependent variables, noting significant increases to staff, quality of services, and public perception of the organization as well as decreased rates of turnover, need for psychotropic prescriptions, and operations costs. While the sheer scope of the project precludes determination of causal relationships, the study highlights the typical scope of intervention required to craft lasting organizational change noted in some of the behavioral safety research.

Reid et al. (2017) conducted another long-term investigation of maintenance of staff performance in a school and resource center that served primarily adults and adolescents with severe disabilities. At the inception of the intervention, some thirty years preceding the follow-up paper, researchers treated a problem wherein adolescents with disabilities were frequently provided non-functional classroom tasks (i.e., they were unlikely to be useful during daily life or failed to contribute to independent living repertoires). To enhance the use of functional tasks (i.e., socially valid either in comparison to typically developing peers or of social or vocational value) in the classroom, researchers operationally defined tasks under their relevant domain (e.g., leisure, social, domestic) and
according to a functional or non-functional dichotomy. The researchers devised a multi-component training package which included an instructional handout, meeting with staff, provision of vocal feedback, prompting, and on-going monitoring of classrooms to ensure students were engaged with functional tasks. Success of the intervention, a near doubling of time spent on functional tasks by students as well as positive social validity reports, prompted an extension to additional classrooms. This resulted in training of new implementers by staff and collaboration with further staff members. Results observed were similar, with an average of on-task behavior increase to 92% during observations, maintained across two years. Follow-up measurement on the maintenance of the program was conducted at twenty and thirty years, with averages of on-task functional behavior occurring in 100% of observations during both evaluations, despite some turnover and movement of staff involved in the initial intervention. Compared to other classrooms measured as a normative standard, on-task functional behavior was similar in the baseline condition but remained above two normative measurements taken around the time of each long-term follow-up. Reid et al. cited the collaborative approach with teachers and educators as a key factor in the maintenance of the intervention components specifically because staff were more likely to adopt the intervention and that the training and dissemination of the intervention was then passed on to new staff. Furthermore, there were likely some contingencies established by linking staff performance (i.e., the assignment of functional tasks) to the positive outcomes in relation to client welfare (i.e., students
engaging in the functional tasks). Such links likely supplemented the monitoring and feedback intervention in exerting control over ongoing maintenance of the program.

**General Strategies to Promote Institutionalization of Interventions**

Thus, staff involvement with planning and implementation of the intervention appears to be a critical component of ensuring ongoing maintenance, as the organization’s constituents are the ones likely to be perpetuating the intervention (e.g., Alavosius and Sulzer-Azaroff, 1990; Reid et al., 2017). Arguably, the adoption or the institutionalization of interventions by the organization might be more important to producing maintenance than any other procedure or component (McSween & Matthews, 2001). Recognizing a need for operationalizing variables that predict institutionalization, Sigurdsson and Austin (2006) crafted a taxonomy comprising four key areas that interventions can incorporate to promote institutionalization, derived from the recommendations of Grindle et al. (2000) and McSween and Matthews (2001).

The first institutionalization variable was involvement in design, meaning that some employees or staff were either partially or primarily responsible for development of the intervention. This could include crafting or identifying organizational goals, devising dependent or independent variables, participating in meetings, and assisting in the development or selection of any assessment or rewards. The second institutionalization variable was training of internal staff, meaning that staff were directly trained to implement some component of the intervention. Training could include management or front-line
personnel and could be conducted by either the researcher or another staff member. Training content could involve learning to observe or monitor behavior of themselves or others, complete checklists, deliver consequences, manage or monitor data systems, or further develop the intervention process. A formal data collection system was the third variable, meaning that internal employees were tasked with the collection of performance data. The final variable was a formal system of dispensing consequences, in which employees, either managers or non-managers, were required to deliver rewards or praise. Sigurdsson and Austin identified and reviewed thirty-one empirical JOBM studies from 1991-2002 that included some form of maintenance measure. Involvement of internal staff was found to be the most common institutionalization variable, present in 48% of studies reviewed. Training of internal staff and the presence of a formal data collection system were present in 42% of all studies, while a formal system of dispensing consequences was observed in 29% of studies reviewed.

Overall, 78% of the reviewed studies included at least one of the four institutionalization variables with a median of two variables per study. Though internal staff involvement in planning intervention was split fairly evenly between managers and non-managers, managers were noted as the most likely to be trained in the intervention as well as tasked with data collection and delivery of consequences. While only 4 of the 31 studies reported incorporating all four variables, a regression analysis indicated that the addition of an additional institutionalization variable would account for a statistically significant increase in the effect size of the formal intervention condition. However, this relationship was not necessarily predictive of maintenance. Though the addition of an
institutionalization variable also accounted for an increase in effect size between baseline and maintenance, this was not measured as a statistically significant relationship.

Sigurdsson and Austin (2006) recommend the incorporation of institutionalization variables into intervention where possible due to the strong relationship on the impact of the initial intervention, with the strongest effects noted where all four are utilized. However, the impact of these variables on maintenance is more tenuous. One potential reason is the small number (i.e., 7) of studies incorporated into the analysis, but there remains additional empirical evidence that institutionalization or, more specifically, variables described in the taxonomy, do not totally account for probabilistic maintenance of interventions. Gravina and Austin (2018) utilized a training program, called the “Consultant Workshop Model,” to teach thirteen senior therapists that oversaw behavioral service delivery for individuals with autism to independently implement performance improvement projects. The goal of the project was to demonstrate the utility of the consultant workshop program in training managers at human service organizations as well as to evaluate maintenance and generalization of the skills trained. Four trainings, comprising discussion, lecture, and activities were conducted and participants were evaluated on their ability to conduct their projects while six of the thirteen participants also self-reported follow-up data on maintenance, generalization, and social validity. The follow-up data indicated that, while four of the six projects maintained some intervention components, maintenance of performance improvement was mixed with two of the six reporting anecdotal maintenance and one providing data supporting maintenance.
Therefore, while the Consultant Workshop Model contained, at a minimum, two of
the institutionalization variables (i.e., involvement in design and training of internal staff)
with the possibility of more variables amongst the individual performance improvement
projects designed by participants, maintenance and institutionalization of procedures was
not assured. Gravina and Austin suggested that the original taxonomy proposed by
Sigurdsson and Austin (2006) might have some room to improve by integrating additional
components, such as alignment with current organizational systems or soliciting
managerial support. Therefore, while institutionalization undoubtedly remains a
conceptually consistent and important variable to cultivate maintenance, namely by
producing more robust intervention effects, it falls shy of explaining the precise variables
and components that predict lasting effects.

Proposal of an Omnibus Term

While maintenance and institutionalization are distinct terms with corresponding
promotion strategies, both terms provide some index of durability through time. Conard et
al. (2016) proposed these terms could be described with resulting tactics subsequently
grouped under a single term, temporal generality. Thus, there are two potential measures of
temporal generality. First, there is data collection or probes following termination of an
intervention or consultant’s involvement (i.e., maintenance). Second, there is data
describing employee and manager utilization or management of an intervention (i.e.,
institutionalization). Conard et al. crafted their own taxonomy of eleven temporal
generality tactics incorporating five tactics specified in the literature regarding
generalization: intervene and hope (i.e., train and hope), sequential modification, introduce
to natural maintaining contingencies, use of indiscriminable contingencies, and mediate generalization (Stokes & Baer, 1977). The authors omitted those tactics that were primarily identified to target generality across settings or behaviors, including training of sufficient exemplars, training loosely, programming common stimuli, and training to generalize.

Four of the tactics were those identified by Sigurdsson & Austin (2006) to promote institutionalization: involvement in intervention design, training internal staff, a formal data collection system(s), and a formal system of dispensing consequences.

Unique to the temporal generality taxonomy, Conard et al. (2016) also synthesized additional tactics from practices identified in the OBM literature to promote generality. Foremost amongst these is the use of Behavioral Systems Analysis (BSA; Diener et al., 2009). BSA refers to a complex framework through which consultants analyze the entirety of an organizational system at multiple levels to identify improvement opportunities. Typically comprising a suite of different tools, interventions derived from BSA are designed to account for all the variables that impact organizational and individual performance within the system. Another tactic added by Conard et al. is the promotion of social validity. The authors argue that measuring social validity, which refers to the acceptability of procedures, goals, and results (e.g., Kazdin, 1977; Wolf, 1978), might provide important information in determining what unique contexts might promote ongoing behavior change. The last technique identified by Conard et al. is the use of instructional design factors. Using teaching or assessment techniques derived from empirical literature should improve the likelihood that a trainee will learn to emit the
targeted behavior proficiently, and thus make the reoccurrence of that behavior more likely.

The temporal generality taxonomy was then applied to the experimental publication history of JOBM between 1997 and 2014. Conard et al. (2016) reviewed sixty-three different experiments that recorded follow-up data, coding them for inclusion of temporal generality tactics, the types of interventions used (e.g., feedback, goal setting), frequency and duration of follow-up data collection, the presence of social validity components, and the overall change in behavior between intervention and follow-up phases. Though most reported behavior changes were robust in follow up, the most commonly utilized tactic was intervene and hope, similar to previous research reviewing the use of generalization tactics (e.g., Osnes & Lieblein, 2003). In other words, most studies merely assessed temporal generality after the intervention had concluded or was withdrawn. Moreover, there was a distinct disparity in the types of tactics used. Multiple tactics (i.e., BSA, formal data collection system(s), and formal system of dispensing consequences) were implemented sparsely in the literature while introduction to natural maintaining contingencies was not used in any studies, perhaps due to the relative difficulty of implementing them (e.g., BSA is a multi-level framework of assessment and intervention).

Conard et al. were also unable to come to firm conclusions regarding the efficacy of multiple tactics as 45 of the 63 studies utilized only one temporal generality tactic. Use of multiple tactics was comparatively rare, with 15 studies employing 2 tactics, 2 using 3 tactics, and 1 incorporating 4 tactics. Generally, these data suggest that OBM literature
rarely investigates the use of the temporal generality tactics and infrequently evaluates the social validity of its practices during relatively short (e.g., 1-2 week) follow-up phases. Relatedly, Conard et al. posit that there are a multitude of unexplored research possibilities to evaluate efficacy of the tactics, identify social contexts conducive to behavior change, and assess the long-term impacts of interventions. As a result, the list of tactics is likely not comprehensive. The authors concede that the temporal generality taxonomy may not account for the entire breadth of mechanisms that might produce change. There may be some strategies that could improve temporal generality that are not commonly evaluated within the literature, such as proprietary, consultant-developed strategies, the role of break managerial support, and the effort required to implement the intervention.
Chapter 3
Barriers in the Study of Maintenance and Institutionalization

Many commentaries have attempted to explain the lack of empirical research in maintenance and commonly associated variables (e.g., institutionalization and generalization). Foxx (2013) hypothesized that the dearth of insight into determining factors influencing maintenance, was, in part, due to the lack of measurement of long-term maintenance and generalization in studies. Potential methodological issues often threaten to undermine these efforts and prevent accurate evaluation of post-treatment intervention effects, including confounding variables (e.g., practice effects generated by subsequent measurement and evaluation), limitations of measurement (i.e., long term recording methods), and common experimental designs. Some single-subject designs, may, in fact, preclude formal investigations of maintenance (e.g., reversal/withdrawal designs; Boyce & Geller, 2001). In a withdrawal design, experimental control is asserted from reverting environmental conditions back to baseline. When a behavior fails to revert following withdrawal of an intervention, experimental control is not demonstrated. If experimental control of robust effects is not demonstrated, then a study will likely not be published. These variables suggest that minimal empirical investigation of maintenance is resultant of contingencies present in a typical researcher’s environment (e.g., Conard et al., 2016; Foxx 2013). Redmon (1991) notes that most behavior analytic training and publications are likely biased toward technological precision of procedures and methodology, while ignoring potential contextual variables that would influence changes post-intervention.
There have been many potential solutions suggested to provide more incentive for researchers to measure generality during the publication process. For example, journal editors could require researchers to include reports of monitoring the extant intervention effects for longer periods (i.e., 2 years or more; Foxx, 2013) or dedicating some of the manuscript to description of contextual variables that gave rise to and promote the longevity of the intervention. More research could also be encouraged from practitioners in the field for whom publication pressures are lessened and where consulting relationships often exceed the duration of the initial intervention evaluation (Conard et al., 2016). Such changes could, hypothetically, contribute to a more consistent literature base describing conditions that promote durability and, thus, allow for further derivation of related technologies and methodologies.

As a counterpoint, literature calling for the long-term continuation of interventions and evaluation of maintenance has clearly persisted, despite the absence of these explicit consequences for researchers. As evidenced by multiple articles written through various decades calling for explication of the variables that reliably produce maintenance, generalization, and institutionalization (e.g., Austin, 2008; Baer et al., 1968; Conard et al., 2016; Sigurdsson & Austin, 2006; Stokes and Baer, 1977), academic interest in promoting long term change clearly remains. Thus, despite the wealth of reviews and discussion as well as a not entirely insignificant portion of empirical JOBM articles measuring post-intervention effects, there has been a decided lack of widespread use and evaluation of the techniques prescribed in the literature (e.g., Conard et al., 2016; Osnes & Lieblein, 2003). A potential alternative explanation is simply that the generality of these procedures is
inadequate. The lack of applied utility was, in fact, predicted: “A catalog of inadequately replicated techniques (the components of which have not been analyzed) each of which worked at least once…may hold the appearance of an interwoven and established literature, but it will prove to be a disappointing façade which does not live up to its seeming utility” (Johnston, 1979, p. 5). In other words, the present conceptualization of temporal generality components (i.e., maintenance and institutionalization) and techniques either misguide or outright preclude empirical evaluation and validation.

Potential Broad Sources of Conceptual and Technological Drift

Despite these expressed concerns, the presumed relationship between generalization technology and maintenance has been perpetuated in contemporary OBM literature, albeit with some terminological refinement and expansion (e.g., temporal generality; Conard et al., 2016). One potential reason is a preoccupation with certain types of performance problems and maintaining variables. Anecdotally, this may be, in part, due to broader trends in ABA practice and common consumer populations. For example, most BACB certified practitioners currently deliver or supervise therapeutic services to treat autism and other developmental disabilities (BACB®, n.d.). Common consumers of therapeutic ABA intervention are likely assessed for and receive treatment to remediate individual performance relative to peers. Many of these performance problems are pure skill deficits that could be broadly described as training and knowledge deficits. In other words, issues wherein the generalization taxonomy would be more appropriate to use
Moreover, applications of the generalization taxonomy have been criticized for overlooking the influence of language (e.g., Boyce and Geller, 2001). Absent or deficient verbal repertoires are a commonly identified symptom of autism and other developmental delays (American Psychiatric Association, 2013). The robust treatment effects of ABA treatment and the subsequent demand for services have also resulted in a great deal of research focused primarily on those representing a small subset of the population distribution (Friman, 2010; Friman, 2021). If a not-insignificant proportion of contemporary ABA findings come from data collected from populations and settings wherein there is reduced likelihood of influence of certain variables (e.g., verbal behavior and social mediation), then it is unsurprising that derived technology would, to some degree, overlook these phenomena. VanStelle et al. (2012) noted that most studies published in JOBM through 2009 were interventions focused on the individual performance of workers. Most OBM research in JOBM would then be classified then as Performance Management (PM), which specifies the use of behavior principles to influence the performance of employees (Wilder et al., 2009). Thus, a potential bias in OBM research, which often frames analysis through the operant paradigm at the individual level, similar to ABA, is not unexpected.

A related problem is also observed when assessing outcomes of intervention. Social validity of ABA interventions has traditionally focused on norm-referenced or informant-based methods, meaning that treatment goals are oriented toward improving individual consumers’ abilities relative to others in areas such as communication and/or participation in society or based on reports of satisfaction and acceptability (Kazdin, 1977).
An intervention then could be called socially valid if a person has been taught a repertoire that is sufficient to contact natural contingencies. Therefore, another tacit implication is that some potent consequences or potential maintaining variables are already available in common environments that receive ABA intervention. Skill interventions can be selected based on commonly taught repertoires of demographic peers (e.g., compared to normalization criteria) or because the participants were satisfied with some facet of the treatment (i.e., the intervention produced adequate, observable change under control of variables in the natural environment). Finding and harnessing these contingencies is critical to developing therapeutic treatment for individuals. For example, Functional Communication Training (FCT; Durand & Carr, 1991; Durand & Carr, 1992) is used to differentially reinforce alternative communicative responses to problem behavior. However, their primary concern with maintenance in these studies was that the taught language was evoked under different conditions. Presumably, they were less concerned that communicative language, once taught in some form, would ever contact environmental contingencies. The supporting macrosystem (e.g., culture and language) already exists and, presumably with some planning, can sustain the behavioral target (i.e., a general mand) following training.

ABA and OBM Differences

Yet supportive environmental features are not always present or of sufficient strength to produce behavior reliably by default. For example, workplace interventions represent a notable deviation from therapeutic intervention, both in the number of different variables (e.g., systems and processes comprising multiple performers) and in focus. The
primary concern is not isolating environmental variables to remediate a deficient individual repertoire, but in diagnosing and correcting problematic systems (e.g., Rummler & Brache, 1990; Malott, 2003) and their component contingencies (e.g., Performance Management; Wilder et al., 2009) to produce better aggregate performance. Performance problems resultant of the complex interconnected behavior of groups of people or outright lack of controlling variables are often the target of behavior analysts specializing in OBM. OBM consultation and intervention has similar criteria to determine social validity as therapeutic ABA services (e.g., norm-referenced performance or procedural acceptability measures) but dissimilar approaches to the problem. Therapeutic service is generally funded to allow an individual to function effectively within the existing broader environment. OBM services, by contrast, seek to ensure the environment is optimally arranged to allow for the desired performances of all constituents. While this is a somewhat semantic distinction (i.e., all interventions designed from a behavior-analytic perspective involves manipulation of environmental determinants of behavior), there is a practical implication.

If environmental arrangement is at the forefront, rather than the specific skills targeted, designing and maintaining an OBM intervention must entail more than ensuring transfer of training effects to the natural environment. In this regard, terms such as maintenance, which have distinct meanings in both basic research paradigms (e.g., resistance to extinction; Kazdin, 1977) and applied contexts (i.e., sustained change following intervention removal; Boyce & Geller, 2001) have limited utility in OBM. The former refers to the programmed and controlled withholding of the maintaining consequences. The latter merely suggests that some variables unique to the intervention are
withheld, removed, or suspended. While superficially similar, this is suggestive of two different intervention strategies, neither of which are pertinent if the goal of OBM is to design environments conducive to consistent and probabilistic performance. Austin et al. (1999) noted that most OBM interventions are focused on building and increasing weak behavior rather than replacing aberrant responding. Therefore, identifying, harnessing, or, in some cases, creating those mechanisms in the target environment is likely more often indicated than planning for explicit extinction or removal of intervening variables. Ensuring the environment is receptive prior to the intervention (i.e., the various variables to maintain the performance are present) at the outset is then a necessary pre-intervention concern when seeking to produce maintenance of targeted performances.

While this is not to diminish the importance of the natural environment when treating therapeutic concerns, pre-existing environmental receptivity to intervention can easily be taken for granted when social structure and norms presently exist for common intervention targets like language and daily living skills. Adjusting a non-receptive environment to subsume the intervention contingencies (i.e., institutionalization) or produce the requisite rate of performance (i.e., maintenance) then necessitates a wider range of techniques to account for existing, alternative, or required sources of control as well as to harness mediating factors (e.g., rules and verbal behavior). In this sense, the empirical pursuit of maintenance effects seems well-suited for OBM.

For example, Sulzer-Azaroff et al. (1990) achieved robust effects implementing a large-scale safety intervention over the course of several months, crediting the use of a thorough systems analysis prior to intervention. The assessment was conducted across the
organization including identification of key performances, stakeholders, and assessing existing contingencies (e.g., policy, rules, roles, structure, etc.). In contrast to the generalization taxonomy, consideration of a greater variety and depth of variables to produce maintenance further supports the clear pragmatic limit to the use of generalization technology. In fact, many of the components of the systems analysis that contributed to maintenance (e.g., assessing important roles and contingencies) could be attributed to a single category of the Stokes and Osnes (1989) taxonomy, exploit natural functional contingencies, which is also one of the least commonly investigated techniques in JOBM (Conard et al., 2016). In such instances, transfer of training outcomes to non-training conditions only represents a fraction of the concerns of assessment and intervention. Suggesting that this is the key challenge of sustaining most interventions, especially in complex verbal organizations, is then reductive. Even distinctions between the training/non-training settings represent a relatively superficial dichotomy. Continuing to ascribe control to these procedures likely misrepresents or misjudges the overall contribution of generalization phenomena to behavior change procedures and minimizes critical extra-intervention effects of human language, extant organizational structure, or other contextual variables. So, despite overlap between ABA and OBM, borrowing a strictly behavior-oriented approach to research and practice, akin to therapeutic applications of ABA, may also potentially perpetuate incomplete examination of variables that contribute to sustained intervention effects. This was historically recognized in OBM, as the discipline was not singularly influenced by Skinner’s work, but also general systems theory (Brethower et al., 2021). The term “performance” comprises not only the behavior, but the results of the behavior as well (Hyten, 2009). Changing a pinpointed target
behavior should not be blind to myriad contextual variables that surround the performance as well as the important resultant outcomes for businesses (Hyten, 2009; Redmon, 1991). Therefore, while such concerns were accounted for at OBM’s outset, robust successes in applied operant conditioning procedures may have inadvertently swayed the course of empirical investigation.

Discussion of Response Maintenance in JOBM

The limitations in theory and utility of generalization techniques were succinctly discussed in multiple commentaries evoked by Boyce and Geller (2001), in their review of occupational and traffic studies. The authors used relatively thorough procedures in describing long term patterns and assessing the variables likely to produce maintenance, which resulted in equally pointed discussion. For example, the authors specify clear duration and level criteria for response maintenance (i.e., the persistence of target behavior(s) above initial levels for at least two weeks following the formal cessation of intervention) to distinguish maintenance from institutionalization. Furthermore, the authors provide a highly specific conceptualization of maintenance. First, maintenance, much like generalization, must be actively programmed to attain control from naturally occurring maintaining variables. Second, they assert the generality of procedures to promote generalization and/or training transfer procedures when applied to concerns of durability. In this regard, they analyze maintenance as a form of generalized behavior change across time. Third, much like narrow stimulus control suggests an effect opposite to generalization, the mechanisms (i.e., contingencies) that underlie maintenance are distinct or in opposition from those that produce initial acquisition. Whereas acquisition of
behavior is acquired under clear, highly probable consequences, maintenance of behavior
is produced by contingencies that are difficult to discriminate or under intermittent
schedules of reinforcement. So, while many of the ensuing commentaries praised the
authors’ thorough contribution to the discussion of maintenance (e.g., Baer, 2001; Malott,
2001), criticism was frequent regarding the specific characterization of variables in the
analysis and subsequent methodological faults.

Commentaries on Boyce and Geller (2001)

One primary area of concern was the concept of maintenance as defined by Boyce
and Geller (2001). Describing maintenance in relation to the removal of the intervention
invokes the experimental roots of the term (i.e., resistance to extinction; McSween &
Matthews, 2001). From a basic research perspective, the maintenance of acquired skills is
typically ascribed to intermittent reinforcement schedules (e.g., Miltenberger, 2016, p. 82;
Skinner, 1957, p. 99). In OBM and other applied settings with typically developed
individuals, interventions rarely rely upon simple operant conditioning paradigms for their
effectiveness, meaning that the pragmatic relevance of this conceptualization of
maintenance is questionable. Moreover, this parameter also implies that the primary target
outcome of behavioral safety initiatives is to, eventually, remove the intervention. OBM
interventions, in contrast to therapeutic treatments, are not necessarily designed to fade or
reduce support. Often, consultants tasked with improving organizational performance
design behavioral programs to remain as a permanent organizational fixture (e.g.,
feedback, incentives; Fleming, 2001). Withdrawal of the intervention should instead be
contingent on the contextual variables surrounding the problem. Some organizational
problems, such as those described by the Boyce and Geller review (i.e., safe behavior in industries with dangerous environments), necessitate ongoing intervention. Furthermore, some organizational contexts support different types of interventions, mediated by factors such as employee skills or the long-term needs of the organization. Thus, the choice and management of an intervention is dependent on the needs of the studied environment. Otherwise, researchers and practitioners risk limiting the scope of potential approaches and interventions. Malott (2001) notes that this is a faulty rule common in other applications of behavior analysis, citing the generally accepted relationship between intermittent reinforcement and resistance to extinction as justification to thin and eventually remove intervention contingencies. Malott suggests that this means some inexperienced practitioners, seeking to increase behavioral durability, might wrongly conclude that behavior can, eventually, sustain itself absent contingencies (i.e., the “myth of intermittent reinforcement”; p., 88-89).

Of course, Malott’s (2001) suggestion that any behavior analysts might espouse the belief that behavior is self-sustaining is a reductio ad absurdum. A more conceptually sound alternative explanation proposed is that, past the offset of the intervention, control has transferred to other variables present in the targeted environment. Malott notes that maintenance by these natural contingencies is often upheld as the optimal target for behavioral interventions, whereas ongoing reliance on contrived contingencies (i.e., those designed and imposed by intervention) are often regarded as indicators that the intervention has failed. This perceived bias toward contrived contingencies could, perhaps, be partly ascribed to a broader cultural resistance to overt incentives and rewards (e.g.,
Kohn, 1993). In terms of behavior analytic intervention specifically, less contrivance is also procedurally sensible. Utilizing already occurring contingencies is far more time and cost efficient. So, while a reasonable approach, as a presumption of behavior analysts prior to intervention, it is problematic and limiting. If any given environment generally features sufficient contingencies to support desired behavior, then the implied explanation is that most people are simply unaware of the natural consequences to their actions (Malott, 2001). Yet, Malott counters that this belies a large amount of evidence to the contrary, including the societal prevalence of problematic repertoires (e.g., smoking, drug abuse). Moreover, were effective natural reinforcement contingencies already readily available in any given environment, then there would likely be minimal reason to formally intervene at all. The desired performance would already be likely to occur. While environments or organizations with effective contingencies and performance issues presumably exist, this circumscribes the range of problems that interventions identified by Boyce and Geller (2001) are effective to treat. Performers unable to emit required behaviors necessary to contact existing reinforcement contingencies suggest a very narrow range of deficits. Such problems would be described as skill and knowledge deficits (McSween & Matthews, 2001). Thus, the restricted scope of performance problems suggests appropriate functional interventions are restricted to the educational (e.g., training and brief supportive interventions) while problems of motivation and related solutions are omitted.

Therefore, the express purpose of the Boyce and Geller (2001) review (i.e., identifying techniques that best promote response maintenance) is hobbled by relying on a narrow criterion of maintenance, explicitly excluding a variety of successful PM
interventions that remained in place following termination of formal intervention. The reasoning for omitting such studies was also the focus of some criticism. Boyce and Geller stated that the primary reason to bypass review of persistent intervention was to avoid encompassing instances of institutionalization, something they considered a different phenomenon. However, it is arguable there is little discernable difference in patterns of performers’ behaviors (Fleming, 2001). Both maintenance and institutionalization presume some form of the maintaining contingencies in the environment following termination of formal intervention. Both then lead to lasting change and thus, there is little pragmatic difference. Maintenance and institutionalization then may be less conceptually distinct and, instead, might represent a gradient along which intervention sustains (Kessler, 2001).

Moreover, McSween and Matthews (2001) argued that the only viable solution to promoting sustained behavior was to encourage strategies that are most likely to be utilized by the organization targeted. That is, withdrawing intervention and seeking to stave off the effects of extinction is not only philosophically misguided but also potentially dangerous, especially in the interventions examined by the Boyce and Geller review (i.e., safety interventions).

In addition to mischaracterizing the variables, the maintenance of said performances is likely not related to the variables or procedural choices analyzed. Rather, there were other behaviors indirectly intervened upon observed in the study, such as employee participation, that maintained the intervention (Fleming, 2001; Kessler, 2001). Participation would likely lead to the establishment of other social contingencies (e.g., discussion or more interactions about safety) that might comprise more accurate
maintaining variables. This remains a problematic confound throughout their analysis because many of the reviewed interventions explicitly discuss the participation of onsite workers in the intervention (e.g., delivering performance feedback). McSween and Matthews (2001) also contend that Boyce and Geller misrepresented the underlying mechanisms responsible in the studies reviewed. Like Fleming (2001) and Kessler (2001), they believed that verbal behavior of employees played a larger role. In their opinion, stability in responding attributed to features of programmed generalization (e.g., mediated generalization, training loosely, programming common stimuli) all point to underlying generation of, and control by, rules. If these other mechanisms (i.e., verbal rules and social contingencies) were, in fact, the actual determinant of long-term success, then generalization techniques add little explanatory benefit. Moreover, the development of rules in these interventions was not necessarily formally programmed by researchers and, technically, then reliant on the least recommended strategy to promote generality (i.e., train and hope; Stokes & Baer, 1977). So, while it was fortuitous that Boyce and Geller (2001) observed a variety of successful safety programs even with minimal intervention components (e.g., weak probability of rewards, non-specific feedback), the success of the interventions was left largely to chance. The Boyce and Geller review, choosing to use only four of seven proposed generalization techniques, ignored more potentially relevant strategies to social contingencies (e.g., introduce to natural maintaining contingencies; Kessler, 2001).

Alternatively, if the verbal and social behavior of participants is largely the responsible mechanism, then the generalization taxonomy proposed by Stokes and Baer
(1977) could be dismissed entirely from the analysis. Because OBM intervention comprises typically developing, adult participants, then there are several methodological issues inherent to the Boyce and Geller (2001) analysis of techniques to promote maintenance. By analyzing these studies under this lens, the commentary is over-reliant on findings derived from basic research literature. Malott (2001) points out that obscuring contingencies and thinning reinforcement schedules is another way in which basic laboratory findings are over generalized to verbal human beings. That is, it is likely not the confusion or unpredictability surrounding consequences that produces durable behavior change, it is mediation by rules and other verbal behavior. All such approaches (e.g., training loosely) in these classification systems are then merely advocating for weak and ineffective training to build new repertoires while understating the role of verbal behavior in the repertoire of typically developing adults. Hoping that intervention will contact maintaining variables in the natural environment or produce effective rules is almost never sufficient. There are many circumstances in which the development of the natural contingencies or control by rules is extremely unlikely. Often, the naturally occurring consequence for a given response class is extremely weak (e.g., seatbelt use and traffic accidents) or improbable (e.g., smoking and lung cancer). Malott adds that, in these situations, contingencies are nearly always required, because their failure to engage in desired target behaviors is likely the result of low probabilities of negative outcomes. Simply applying a rule may be initially successful, but control may be susceptible to transfer to other adventitious reinforcement or automatic avoidance contingencies. Instead, ideal training techniques (i.e., highly probable and frequent consequences) should be used and maintained throughout.
In short, maintenance, should be cautiously described as responding in the absence of intervention contingencies (i.e., extinction). Durable responding occurs with supportive environmental contingencies and, if not initially present, interventions in some settings (e.g., workplaces) should never cease to provide these supportive changes. In this regard, a truer goal for OBM is not behavior in the absence of intervention, but instead institutionalization of components (Malott, 2001). In this regard, these interventions were not inherently made to be durable. Rather, they created new naturally encouraging contingencies in the workplace. In either case, whether by maintaining intervention contingencies or creating others, extinction is a generally irrelevant term and removal of the intervention is merely a topographical label. Baer (2001) notes that Boyce and Geller (2001) are accurate in acknowledging that maintenance of effects is an ongoing and potentially technological process, over and above just choosing an adequate environment after training. Yet, their analysis is merely speculative. The small range of studies is vulnerable to multiple threats to validity (e.g., analyzing studies post-hoc), the outcomes specified likely could have been due to chance, and the results often point to tenuous advice for practitioners. Limiting analysis to Stokes and Baer (1977) generalization taxonomy does not clarify the mechanisms responsible for change underlying the techniques (e.g., the precise role of rules when strategies are used) and also excluded other idiosyncratic variables which could be contributing to sustained performance (e.g., stereotyped responding patterns or the visibility of the interventions) or could moderate the impact of those variables (e.g., turnover and other organizational problems; Kessler, 2001). Additionally, some of the key features in maintaining safety performance in the review (i.e., written commitment strategies, safety awards, and group feedback) could not
necessarily be described as resounding successes, sometimes failing to produce performance above necessary levels for modern industry (McSween & Matthews, 2001).

In sum, there are several areas in this discourse indicative of broader trends within the literature. At the forefront, there is an overall lack of consensus surrounding the definitions and applicability of terms, namely generalization. A fundamental concern is deciding whether phenomena derived from studies on non-human animals and basic research function as true models of typically verbal human behavior or are simply convenient and useful analogues (e.g., Malott, 2001, McSween & Matthews, 2001). Behavior analytic accounts of language and cognition imply something closer to the latter and suggest limits for which certain phenomena, such as stimulus generalization, can explain complex human behavior (e.g., Relational Frame Theory; Hayes et al., 2001, p. 45-46). Occupational repertoires, the typical focus of OBM intervention, exemplify these issues, often intertwined in intricate, dynamic contingencies, and subject to indirect control by verbal behavior in addition to direct reinforcers and punishers (O’Hora et al., 2013). While a small degree of conceptual drift was acknowledged at the inception of generalization technology (i.e., Stokes & Baer, 1977) the continued prevalence in applied literature, despite advances in the sophistication of behavior analytic science and in spite of a narrow range of applicable contexts, remains worrisome. This is not to say that techniques to promote occurrence of behavior outside of training or intervention conditions are unimportant or that the tactics listed herein are without any merit. However, this is a misnomer in labeling these observed patterns uniformly (e.g., Malott, 2001; Sigurdsson & Austin, 2006) and continued misuse of the term potentially deters investigation and
elaboration of important behavioral processes (Johnston, 1979). Such disparities likely have contributed to the observation that OBM practice has long surpassed its theory (McSween & Matthews, 2001). Maintenance has been assumed as a default component of an intervention outcome but not demonstrated with experimental rigor sufficient to generate findings which inform the creation of general and robustly researched applied techniques. Arguably, this absence has resulted in the continued reliance on temporal generality tactics and techniques (e.g., those outlined in Conard et al., 2016), prescribed to be applied probabilistically to improve the durability and generality of interventions. While the overall nature of most behavior analysis is probabilistic, success of such tactics relies too much on chance, more analogous to the behavior modification dependence on default technologies rather than on the, now common, reliance on functional analysis techniques in ABA (Mace, 1994). Of course, to feasibly study determinants of lasting change, the parameters of what constitutes maintenance require due consideration.

Definitions of Maintenance and Dimensions of Maintenance for Study

Without clear definitions of terms, examination of generality patterns in previously published applied literature can be difficult to draw conclusions from (e.g., Pennington et al., 2019). To produce an adequate definition of maintenance to study empirically, multiple inconsistencies should be addressed. The literature base hints at three main definitional concerns regarding maintenance: the variables present and/or absent that delineate maintenance from other intervention phases, the duration of sustained responding that
constitutes “successful” maintenance, and the performance that accurately encapsulates maintenance (i.e., sustained levels of the dependent variable or continued implementation by non-researchers). For the first concern, much like generalization, the term invokes multiple potential definitions.

In many applied interventions maintenance is defined topographically both in phase change decisions (i.e., maintenance is typically assessed when the intervention or researcher is no longer present; Conard et al., 2016) as well as patterning criteria for success (e.g., sustained levels or trend similar to intervention phases; Foxx, 2013). In other cases, maintenance is simply defined as the continuation of the targeted performance after its initial establishment (McSween & Matthews, 2001). Any of the prescribed parameters have worked well enough to produce isolated evaluation of maintained behavior change in the literature but such wide variation is likely to lead to inconsistent explanation of the variables accounting for response maintenance, as well as any methods used to enhance evaluation. The latter definition, while compact, is likely the most difficult for experimental evaluation. Any occurrence after the initial establishment of performance (e.g., a steady state in a single-subject research design; Johnston & Pennypacker, 2020) could suggest control from experimental variables (i.e., an ongoing implementation by a researcher) as well as any active natural environmental contingencies and other supports. Defining maintenance specifically as an end to the intervention phase could, alternatively, suggest a full transition to natural maintaining variables. However, the two instances where this would be true are either with complete withdrawal of the formal intervention or if the intervention becomes “natural” (i.e., maintained by in-house implementers). The former
case, removal of the intervention, as noted in some commentaries (e.g., Fleming, 2001), is not necessarily a desirable outcome in many instances. The latter case is problematic by subsuming a separate performance of interest (i.e., institutionalization). Arguably, in most cases, maintenance phases represent something in between a complete return to baseline conditions and repeated measures of the precise intervention conditions ad infinitum.

An alternative is that the transition from intervention to maintenance is neither response dependent (e.g., initial performance criteria are met) nor dependent on changes to the independent variable (i.e., removal or reduction of intervention contingencies). Rather, the maintenance phase is initiated once the researcher or consultant has ceased some dimension of formal involvement. Sigurdsson and Austin (2006) noted that this is likely a more ubiquitous outcome. Yet this definition is applicable only in cases wherein the researcher or consultant is external to the organization or operates outside of the target performance. In this regard, while maintenance may not easily have an absolute and concise definition outside of basic research paradigms (i.e., as a response pattern under altered reinforcement or extinction conditions), there are two potential criteria on which to assert maintenance. First, a maintenance evaluation phase should not be defined narrowly. Rather, it should be delineated broadly by some aspect of change in the intervention formality (e.g., experimenter involvement, intervention, implementers, etc.) in a manner predetermined by the researcher, according to the targeted environment and population (e.g., the intervention may never truly cease, but the implementer will not maintain continuous measurement). Second, there is a performance criterion to assert maintenance.
The measured dependent variable should remain at or improve upon formal intervention levels.

**Duration to Assert Maintenance and Long Term Follow-up in Maintenance Research**

In addition, determining where maintenance begins does not necessarily specify where maintenance ends. One common refrain in maintenance related literature is the request for longer term investigations (e.g., Boyce & Geller, 2001; Conard et al., 2016; Foxx, 2013). Robust demonstrations of results over time are certainly necessary to help describe the long-term effects of OBM (and ABA) applications. However, these results should be approached cautiously in evaluating an intervention and the specific determinants of maintenance.

Critically, long-term duration of maintenance evaluation may not necessarily serve as a definitive representation of intervention effectiveness. McSween and Matthews (2001) noted that the Boyce and Geller (2001) review reported intervention outcomes as successful maintenance that included performances below industry safety standards. Foxx (1990) also observed that, in some cases, failure to deviate from treatment scenarios might also produce counterproductive effects on performance. Inconsistency of effects should also be expected. Few behaviors are likely to remain similar in rate or occurrence ad infinitum, especially in OBM contexts. Organizations are dynamic systems with adaptive qualities (Malott, 2003) and workplace performances are then unlikely to comprise specific response classes that will be eternally useful to the organization and, by extension, the
performer’s repertoire. Situations wherein maintenance of the response has occurred absent continuation of a specific intervention (e.g., Boyce & Geller, 2001; Grindle et al., 2000) or situations where intervention continues but produces diminishing results can both be feasible outcomes over time, regardless of the initial validity of the intervention.

Such evaluation should be also approached with care from an empirical standpoint. Experimental control is dictated when, in the absence of other variables, the independent variable produces reliable changes to the dependent variable. As time following formal intervention grows there is a probable inverse relationship between confidence that the original independent variable is still producing the change. In clinical cases, behavioral function has been observed to be dynamic, necessitating updated functional assessment (Lerman et al., 1994). Furthermore, checks and balances of experimental methodology (i.e., reliability and integrity measures), which assure the relationship between variables remains intact, are likely no longer measured when researchers or consultants are formally absent. Thus, short term maintenance evaluation, where the experimenter can remain relatively confident of internal validity, can be useful in providing accurate observations about the corresponding variables that are supportive of the intervention effects. Assessing performance change over a span of years, conversely, may ascribe too much explanatory power to the potency of the original intervention. In addition, studies that have examined long-term maintenance note that maintenance of the performance (i.e., the dependent variable) did not necessarily entail the maintenance of the intervention with high fidelity. For example, Grindle et al. (2000) noted that some safety targets persisted despite discontinuation of the formal safety program due to the creation of new social norms.
Therefore, in conducting long-term evaluation, it can be difficult to suggest which aspects of the intervention, if any, were systematically responsible for maintaining continuing change in behavior.

Furthermore, as length of measurement time increases, the chances that the intervention participants remain consistent decreases. Turnover, attrition, or other loss of participants does not necessarily preclude measurement of maintenance. In fact, evidence of behavioral intervention persisting despite turnover arguably bolsters assertions of durability. Yet, similar to changes over time in functional variables, such as when control is transferred to social contingencies, differing participants also undermine confidence that the initial intervention variables or features thereof are responsible for ongoing change. Rather, this suggests that other mechanisms are responsible for persistent behavioral patterns. For example, Reid et al. (2017) observed that their staff management program persisted despite turnover of some initial participants in the intervention. The authors suggested that the consistency of intervention over multiple decades was likely due to the dissemination and training of new implementers resultant of planning that included participation from key stakeholders. In this instance, the dependent variable (i.e., the implementation of functional tasks in the classroom) maintained over time, but it is difficult to ascribe these extant effects to components of the initial intervention because of differing maintaining variables (i.e., social contingencies acted as supplements to the original intervention package) and participants. A more conservative determination is that the intervention, some thirty years later, targeted a socially valid performance, because it resulted in contact with functional variables in the natural environment. Moreover, the
observed consistency in the use of functional tasks in the classroom in follow-up measurement was not explicitly produced by the intervention, but rather in the maintenance of some of its constituent components. This distinction between maintenance of the performance and maintenance of the intervention is, presumably, what causes difficulty in separating maintenance from institutionalization (e.g., Kessler, 2001).

Thus, the question of the requisite duration to assert maintenance of intervention effects is specific to the intervention and setting, and, from an empirical perspective, does not rule out evaluation over shorter durations (i.e., to demonstrate experimental control of performances when certain intervention features are changed or removed). Previous long-term evaluations of interventions have suggested situations in which confidence in the integrity of procedures may have waned, whether through attrition of original participants, leadership changes, or, simply, lack of measurement. Asserting that the intervention contributed to long-term maintenance is, in some ways, logical but does not inherently demonstrate what about the intervention was important leading to speculative outcomes (e.g., Boyce & Geller, 2001). In the case of studies where the intervention was removed successfully, it is feasible that potentially none of the components were definitively necessary, other than specifying and highlighting the required behavioral changes of performers. That is, if socially mediated behavior was the necessary controlling variable to produce safe performance in the long term, as hypothesized by some other commentaries (e.g., McSween & Matthews, 2001), then some of the other strategies removed were arguably superfluous. Long term evaluations of program maintenance, at least without rigorous and continuous measurement, are then probably better described as a quantitative
measure of social validity (i.e., the length of time the behavior is useful for the consumer, performer, or organization) rather than as a description of specific intervention variables that successfully contributed to maintained performance. Response maintenance as a function of the planned intervention components is then likely best measured following the initial demonstration of the behavior at strength or after providing evidence of steady responding in the presence of the intervention.

In sum, many current applied definitions of maintenance are not general enough to account for all possible cases (e.g., removal of intervention, removal of experimenters or consultants, transfer of implementation to others, etc.) and are sometimes inconsiderate of the maintaining variables (i.e., the performance recurrence after establishment). Concerns over the latter also tend to persist when reexamining interventions months or years following the initial experimental phases. Definitional inconsistency, too, likely contributes to non-significant empirical findings when trying to examine predictors of maintenance (i.e., Sigurdsson & Austin, 2006). Thus, for the purposes of this paper, response maintenance will be defined as the following: the continuation of the targeted performance in the desired direction following its initial establishment (i.e., achievement of a steady state) after cessation of some component of experimental contrivance (e.g., researcher presence, continuous measurement and observation, etc.). Under this definition, maintenance can be described as a singular behavioral pattern (i.e., separate from institutionalization), pragmatic for research purposes in delineating a difference in variables from intervention conditions, while encompassing the widest variety of post-study outcomes.
Determinants of Maintenance

Overall, reviews of previous studies suggest two general findings when assessing the variables most likely to produce maintenance. First, taxonomies of tactics are generally underutilized in most research studies (Conard et al., 2016; Osnes & Lieblein, 2003) and represent incomplete explanations for the occurrence of maintenance in applied organizational settings (e.g., Fleming, 2001; Kessler, 2001; Malott, 2001; McSween & Matthews, 2001). A second trend is noting that specific variables responsible for the perpetuation of interventions tend to be idiosyncratic to the sorts of problems and populations intervened upon (see Foxx, 2013; Myers et al., 2010; Reid et al., 2015), but the variables consistently comprise contact with pre-existing (i.e., “natural”) or institutionalized contingencies. Previous research also suggests that the precise variables evoking and maintaining ongoing performance will likely evolve over initial intervention, either by development of new, often social, contingencies in the absence of intervention (e.g., Boyce & Geller, 2001; Fleming, 2001; Kessler, 2001) and/or extension of contingencies present in the existing environment (e.g., Reid et al., 2015). Such changes in maintaining variables, absent the original intervention, seemingly confirm what multiple commentaries have noted: that OBM mainly deals largely with rule-governed and verbally mediated behaviors (e.g., Austin et al., 1999; Malott, 2001). In other words, initial interventions can set the stage for creation of verbal cultures and subsequent transformation of stimuli functions that produce ongoing sustained responding.

In this sense, the appeal to strict operant conditioning procedures is no longer necessary or warranted. Moreover, methodology which limits generality and restricts focus
to certain performance problems, consumers, and/or settings by compromise of those principles and concepts should be amended. Thus, a feasible starting point in investigating predictors of maintenance is to eschew the standard premise of many generality tactics. The principles underlying the initial behavior change remain the same principles maintaining the performance. That is, if a performance contacts effective contingencies, then the performance will continue. If the performance is to recur at desired rates, it must continue to contact effective contingencies, whether it is the initial intervention contrivance or emergent social pressures resultant of the intervention. No repertoire is indefinitely resistant to extinction. Ensuring that performance occurs improbably under unideal or deleterious conditions is, at best, a temporary solution. Eventually, environments must be arranged to be highly conducive to the occurrence of socially important behavior. In most cases, the approach then is straightforward: high probability behaviors are best maintained by the occurrence of high probability consequences (Malott, 2001). Or, as Skinner states, “Behavior continues to have consequences, and these continue to be important. If consequences are not forthcoming, extinction occurs (1953, p. 98)”. Therefore, a major concern in research of maintenance should be the interplay between the intervention and the targeted milieu over the relative strength of the intervention itself.

A general process can then be extracted from previous studies to accomplish this goal. The first key factor appears to be a form of pre-intervention assessment to identify the nature of the problem and gauge environmental receptivity to a particular solution. There is precedent in successful, long-term interventions assessing these variables either through informally surveilling the entire organization (e.g., Christian, 1984) or conducting a formal
systems analysis (e.g., Conard et al., 2016; Grindle et al., 2000; Sulzer-Azaroff et al., 1990). Assessment likely allows researchers and consultants to identify current effective and deficient contingencies within the organization as well as any ancillary barriers to intervention implementation. A second key part of the process is ensuring such assessments are conducted collaboratively with members of the representative organization, both for accuracy of observations as well as to solicit buy-in and participation. The latter appears to be essential in sustaining organizational intervention (e.g., Boyce & Geller, 2001; Hagge et al., 2017; Myers et al., 2010; Reid et al., 2015). A performance that eventually produces natural, beneficial outcomes also promotes sustained change (Alavosius & Sulzer-Azaroff, 1990; Reid et al., 2015). So long as a performance contacts high probability positive outcomes, this may also promote the creation of new cultural rules and social contingencies that can contribute to performance maintenance, even if some intervention implementation has waned, or implementers have left or changed (Fleming, 2001; Foxx, 2013; Kessler, 2001; McSween & Matthews, 2001). When high probability contingencies are present, staff are involved in design, and social contingencies develop, this can potentially lead to a secondary performance increase in the institutionalization of intervention components (Alavosius & Sulzer-Azaroff, 1990; Boyce & Geller, 2001; Myers et al., 2010; Reid et al., 2015; Sigurdsson & Austin, 2006).

Institutionalization as a Distinct Phenomenon

Institutionalization, like maintenance, has also suffered some definitional inconsistency. It has been described as both separate from maintenance (e.g., Sigurdsson & Austin, 2006) or as an indistinguishable part of maintenance (i.e., Kessler, 2001). Kessler
defends the latter assertion in that the performance outcomes of an institutionalized intervention are the same as a performance that contacts an already occurring contingency. However, this comparison indicates the evident distinction between the two terms. That is, one outcome (i.e., maintenance) can potentially occur independent of the other (e.g., if a training deficit is remedied and a performer contacts already present environmental contingencies). In this regard, institutionalization definitionally requires consideration of at least two distinct performances, one of intervention implementation and one of the target performance. Presumably, maintenance could be examined in both cases (i.e., maintenance of the intervention effect and maintenance of implementation). In this sense, maintenance is an examination of a specifically targeted behavior or performance, most typically defined within a traditional behavioral contingency. Institutionalization, is, at a minimum, the performance of one individual interacting with another’s performance. In other words, institutionalization, by default, specifies an interlocking behavioral contingency (Malott, 2003, p. 36).

Determinants of Institutionalization

The presence of two or more distinct behavers introduces additional conceptual nuances. As noted previously, OBM research (i.e., VanStelle et al., 2012) occurs mainly at the performer level, and, thus, the lens through which change is examined is through the lens of operant learning. However, the stated intent of institutionalization is for the behavior of the researchers and/or consultants to be perpetuated, with fidelity, by constituent members of the organization (i.e., the intervention is maintained internally; Sigurdsson & Austin, 2006). In this sense, there are two potential and related ways to
examine and influence institutionalization. The first is to proceed much in the vein of traditional OBM research and affect change at the performer level. Institutionalization can be targeted as a separate performance to be intervened upon. In other words, teaching supervisors, managers, and other in-house implementers adequate repertoires and providing environmental supports to increase likelihood of maintenance of intervention implementation.

The second way to analyze the problem is as a broader deficiency in the system of management. Gilbert (2007, p. 76) noted that most performance problems, either as the result of lacking individual repertoires or inadequate environmental control, is ultimately a problem of systems. That is, regardless of the variables inhibiting the desired performance(s), that the terminal failure lies in properly defining and managing the critical features of those performance(s). Presumably, this assertion is equally applicable to frontline employees and their managers as it is to the managers of those managers. Effective intervention occurs when all levels of performance adjust to serve to produce the terminal product (Malott, 2004; p. 175). Managers unable to manage is as equally critical an issue as deficiencies in the repertoires of those they manage. However, in contrast to basic principles of operant learning (i.e., variation, selection, and retention at the individual level) and measured outcomes of most typical OBM research, these changes in dynamics must occur across multiple levels of the organization and organizational constituents. Therefore, there is a second process that likely results in the perpetuation of certain interventions in organizations, a larger target outcome of selection and transmission of practices across individuals (i.e., cultural selection; Atkins et. al, 2019, p. 17).
Neither level of analysis is necessarily without merit, as both operant learning (i.e., selection by consequences) and cultural selection intersect frequently (Atkins et al., 2019, p. 18). However, this suggests that the optimum outcome of institutionalization is to produce important performance at the group level and generate a culturo-behavioral lineage (Glenn et al., 2016). Thus, one potential way to analyze the process of institutionalization is through the lens of cultural phenomena. Glenn (2004) posits that cultural selection occurs through units analogous to those in operant selection, the metacontingency. A metacontingency comprises a group of interlocking operant contingencies which produce an aggregate product (i.e., a culturant), which are then selected by culturant consequences (i.e., environmental variables that “reinforce” the group behavior; Glenn et al., 2016). The concept of the metacontingency and interlocking behavior is not foreign to OBM both in JOBM discussion (e.g., Clayton et al., 1997; Ludwig, 2017; Mawhinney, 1992; Robertson & Pelaez, 2016) as well as in its subdisciplines (i.e., Behavioral Systems Analysis; Malott, 2003).

However, while the metacontingency as a basic unit of culture has been proposed, refined, and analyzed over multiple decades, definitive basic empirical demonstration has been elusive (Mattaini, 2004) and complementary terminology and technology remains in its relative infancy (Glenn et al., 2016). Even subdisciplines that, to some degree, utilize the metacontingency paradigm (i.e., Behavioral Systems Analysis; BSA) lack clear procedures and demonstrated empirical results, despite perceived practical utility (Diener et al., 2009; Johnson et al., 2014). Consequently, though it may be feasible to assert that the broad, overall goal of institutionalization is cultural in nature, empirical examination...
should remain restricted closer to analysis at the performer level. Malott (2004, p. 173-176) provides a framework for producing change in this regard, suggesting design of PM systems that account for interlocking behavioral contingencies at multiple levels. Contingency analysis, Malott suggests, should not be restricted to simply the level of targeted performance change, rather at all levels that interlock and influence the aggregate product. This also suggests a fundamental philosophical difference in traditional approaches to maintenance and institutionalization. The critical feature of maintaining change is not adding more components to the intervention, but rather adding more layers of behavioral connectedness and implementing control systems that identify and manage these interlocking contingencies.

In this sense, accomplishing institutionalization through ancillary design modification (e.g., involving employees) is likely to be inefficient at any level of analysis. At the performer level, assuming that features can be uniformly added to the intervention (i.e., the institutionalization taxonomy; Sigurdsson & Austin, 2006) and produce linear improvements to performance is somewhat reductive and mimics some of the issues inherent to default use of maintenance taxonomies. At the operant level, because institutionalization is, itself, just another performance to be managed, the variables inhibiting or limiting that performance are likely to vary based on context. As a matter of culturo-behavior, institutionalization is part of a functional unit, related to some aspect of the primary target performance to produce a shared outcome. While Sigurdsson and Austin tacitly acknowledge this by positing broader, process-based strategies (e.g., training, formal systems of consequences and data collection) to promote institutionalization
compared to the more contingency focused maintenance strategies (e.g., use of indiscriminable contingencies; Conard et al., 2016), indiscriminate application is likely to produce unclear impacts if the solutions are not bespoke to the specific contexts of the performer and organization. For example, a researcher or consultant could intervene upon a deficient performance that requires some antecedent intervention (i.e., task clarification) while adding a formal system of dispensing consequences (i.e., a tactic for institutionalization) for use as a secondary supportive management tool. Presumably, the performance targeted already has a clear, beneficial outcome present in the environment if assessment indicates that it is solely resultant of deficient antecedents. Once remedied, this performance would likely occur regardless of the institutionalization variable. In this specific instance, the institutionalization variable would likely correlate with, but not actually contribute to maintenance. Moreover, the institutionalization variable would have little to do with the actual continuation of the intervention.

Summary of Terms

In sum, maintenance and institutionalization can work theoretically and pragmatically as separate concepts. Maintenance, in most applied research, is probably best characterized as a post-intervention experimental phase that removes some formal variable regarding the intervention. Successful maintenance can be defined as continued or exceeded criteria of the intervention effect, both topographically (i.e., regardless of the presence of the intervention variables) and/or functionally (i.e., as a continued result of the intervention components). Description of the latter remains limited in long-term follow-up accounts. Institutionalization, by contrast, generally specifies the maintenance of an
intervention procedure. At its most limited, an institutionalized intervention is a specific interlocking contingency between an internal managerial behavior (i.e., delivery of the intervention components) and the target performance. At its most expansive, institutionalization would imply the incorporation of an intervention into a larger metacontingency (i.e., a group of interlocking behaviors, aggregate product, and selecting environment). Taxonomies of strategies to promote both have been popular both broadly in behavior-analytic literature and in JOBM. However, the demonstrated generality of techniques has been limited due to the choice of underlying explanatory mechanisms and derived techniques (e.g., generalization) as well as the idiosyncratic nature of variables that often underlie successful interventions (e.g., Boyce & Geller, 2001). Therefore, a pragmatic evaluation of maintenance and institutionalization should not involve atheoretical application of broad tactics. This is not to say that previous work has not identified valuable information about procedures that have produced generality. However, successes and failures of maintenance and/or institutionalization, as with any performance or behavior change, should be considered functional in nature and occasioned by variables specific to the target environment. In complex organizational environments, this change seems more readily accomplished by identifying as many contributing variables as possible to the production of lasting change. These procedures to identify and remedy problems seemingly work better when they involve input of organizational constituents throughout to solicit other potential sources of control, or to perhaps, formally, harness them. In this regard, there is an unexplored procedure in OBM that utilizes both elements: the use of functional assessment.
Chapter 4
Functional Assessment in OBM in Relation to Maintenance and Institutionalization

Identifying maintaining variables to influence socially significant behavior is not new to behavior analysis. Since Iwata et al. (1982) first specified an experimental procedure to identify environmental variables maintaining self-injurious behavior, functional analysis (FA) has become a standard best practice in ABA. The significance of the FA procedure cannot be understated, both due to demonstrated generality across myriad behavioral concerns and populations (Beavers et al., 2013) and its critical role in changing the philosophical direction of the applied field (i.e., functional replacement compared to indiscriminate use of strong reinforcers and punishers; Mace, 1994). The latter has given rise to a variety of functional assessment methodologies, which typically comprise informant-based (e.g., questionnaire), descriptive (e.g., scatterplots, ABC analysis), and experimental analyses (e.g., FA procedures). However, while such procedures became increasingly default in therapeutic ABA settings, they remained atypical in OBM research by the turn of the 20th century. Austin et al. (1999) hypothesized potential reasons why such procedures did not develop concurrently with advances in ABA practice. First, OBM achieved significant successes in research and practice absent formal assessment. Adding formal assessment procedures could add minimal benefit to already potent interventions while increasing effort, cost, and potential conflicts with organizational needs. Moreover, conventional ABA assessment procedures would require significant adaption to effectively treat the typical population and problems of concern in
OBM intervention. Austin et al. noted that language and rule-governance is prevalent in typical OBM subjects, and most non-behavioral tools were indirect in nature. Relatedly, because OBM is almost solely focused on increasing behavior in complex work arrangements, simple arrangement of analogue contingencies would be inadequate to consider all the barriers to effective performance.

Austin et al. (1999) reported that formal functional assessment was virtually non-existent in OBM research at the time of publication. In the intervening decades, it appears that the article served as an accelerant for OBM functional assessment procedures in empirical work. Wilder et al. (2017) reviewed JOBM from 2000-2015 and found that 28% of all empirical articles utilized a preintervention assessment. Within those articles, indirect methods seem to be preferred over direct or experimental methods. In the case of the latter, experimental analysis continued to be rare, likely due to constraints common to organizational environments (e.g., inability to withdrawal intervention or separate components). However, combinations of indirect and descriptive methods were not uncommon, representing approximately one-third of studies that included pre-intervention assessment. Two of the most popularized OBM functional assessment methodologies, Behavioral Systems Analysis (BSA; Diener et al., 2009) and the Performance Diagnostic Checklist (PDC; Austin, 2000) in fact recommend the use of both. Johnson et al. (2014) reviewed their use in OBM intervention and found that the assessments comprised roughly 15% of all empirical JOBM articles. Or, in relation to Wilder et al. (2017), this was more than half of all articles that utilized preintervention assessment. Johnson et al. (2014) noted that, despite perceived popularity, empirical data surrounding the use of these assessments
was minimal, especially in the BSA literature. Most BSA studies were conceptual in nature and, when data were present, rarely demonstrated efficacy of the approach over other functional assessment procedures or conventional approaches. Literature on the PDC, a far more compact question-based assessment approach, was much more likely to produce some empirical data, but, at the time of publication, lacked evidence of validation.

Therefore, significant questions remained regarding how well these assessments predicated functional approaches in OBM, both in terms of intervention effectiveness and comparative efficacy. Thus, their role in producing effective maintenance of performances remained unknown as well. However, while measurement of maintenance has not increased, the years following the Johnson et al. (2014) and Wilder et al. (2017) reviews saw literature on the PDC expand significantly through the development of multiple variants and studies that have investigated validity of these instruments. This, combined with a relatively circumscribed procedure compared to BSA (i.e., discrete guided questions compared to multilevel, multivariate analyses with several distinct tools), suggest a protocol more amenable to investigation of maintenance and institutionalization.

History of the Performance Diagnostic Checklist (PDC) and Variations

The Performance Diagnostic Checklist (PDC) was derived from research studying the method and content of inquiries when managers or experts attempt to find solutions to performance problems. Austin (1999) used these data to distill the best problem solvers’ questions into four distinct domains, adapted from a variety of Performance Analysis
Literature (e.g., Gilbert, 2007, Komaki et al., 1986; Rummler & Brache, 1990):
Antecedents, Equipment and Processes, Knowledge and Skills, and Consequences. A
consultant tasked with solving a prospective problem would presumably use these twenty
questions to pinpoint the specific area with the most salient opportunities for intervention,
thus providing a direction to choose an appropriate functional intervention. Though
Johnson et al. (2014) noted that empirical use of the original PDC had recently waned, it
remained one of the most frequently cited OBM functional assessments through the first
decade of the 21st century.

The PDC in Empirical Literature

Pampino et al. (2003) were among the first to use the PDC as a preintervention
assessment in published research. The PDC was administered prior to evaluation of weekly
and daily feedback effects within a package intervention to improve completion of
inventory and maintenance tasks in an art and framing store. Results of the PDC indicated
potential problems in two domains: Antecedents and Consequences. Intervention was
therefore designed to account for both domains, comprising task clarification, goal setting,
and contingent access based on goal attainment as well as performance feedback. Across
two experiments, there were mean increases in completion of duties, though direct
improvements to staff behavior did not extend to the owners of the store. The general trend
of intervention in retail settings characterizes much of the research utilizing the PDC.

Shier et al. (2003) conducted a study similarly in a retail setting and in focusing
intervention on completion of cleaning tasks at a grocery store. The PDC was used to
assess and determine intervention components, indicating issues related, again, to two
domains, Antecedents and Consequences. A package of task clarification, checklists and feedback were implemented, resulting in mean performance increases in all departments over baseline. Shier et al. note that data collection was reliant on self-monitoring, which limited their confidence in the source of intervention effects.

Pampino et al. (2004) examined the use of the PDC as an assessment tool to guide intervention choice for increasing maintenance tasks in an independent coffee shop. The PDC indicated that two domains, Antecedents and Consequences, were likely responsible for deficient performance and researchers implemented a package intervention of task clarification, a training checklist, public posting of tasks completed as well as a reward lottery. The intervention was introduced across two task groups and completion of closing tasks increased for both. Authors in each of the aforementioned studies (Pampino et al. 2003, Pampino et al., 2004; Shier et al., 2003) note that use of a packaged intervention obscured which variable(s) represented the primary change agent. Moreover, they were unable to ascertain how much the use of the PDC helped beyond informal assessment.

Austin et al. (2005) examined the use of the PDC to assess problematic performance amongst staff at a privately-owned restaurant. PDC results indicated that employees were unaware of job duties and lacked consequences for task completion. To address performance deficiencies in completion of closing tasks, a package intervention was introduced across servers and dishwashers. The package comprised graphic and verbal feedback as well as task clarification. Unlike previous PDC research, delivery of graphic feedback was staggered within the server group allowing for a limited component analysis of the impact of task clarification and verbal feedback. There was seemingly minimal
improvement with the addition of graphic feedback, possibly influenced by a ceiling effect as both groups improved mean performance over baseline and were relatively close to 100% of possible tasks completed. However, researchers noted that some individuals were likely more responsible for the improvement than others.

Eikenhout and Austin (2005) sought to improve customer service behaviors at a Midwestern department store and utilized the PDC to functionally assess deficient performances. Noting the general dearth of feedback and relative response effort for engaging customers, Consequences were noted as the primary domain to improve. Customer service behaviors (e.g. greetings, offering assistance, smiling, small talk) were targeted in a multiple baseline across behaviors with embedded reversal design. Intervention comprised a package of graphic feedback, goal setting and reinforcement. The intervention resulted in increases in all behaviors relative to baseline. Consistent with earlier trends in PDC studies, researchers noted that the package intervention casts doubt upon the primary component(s) that produced the effect and the extent to which the intervention impacted individual performers.

Rodriguez et al. (2005) used the PDC to assess the variables responsible for low rates of promotional stamp offers at two locations of a restaurant franchise. Antecedents, Equipment and Processes and Consequences were identified as problematic in both restaurant locations. A package intervention comprising task clarification, self-monitoring, equipment modification, goal setting and graphic feedback was introduced in a multiple baseline design across settings. In both stores, there were increases in promotional stamp offers over baseline rates. The authors note that due to the large amount of intervention
components, the relative importance of each component was not tested. Moreover, in contrast to other studies where the overall contribution of each component was unknown, the significant number of intervention components also cast doubt on whether the preintervention assessment yielded differential results over informal assessment of the problem.

Doll et al. (2007) devised an intervention to target store maintenance and cleaning (e.g. cleaning countertops, mirrors, hanging jackets) in a ski shop. The PDC was used as one of two functional assessments. Though researchers did not identify the deficient domains identified through the PDC, they concluded that Antecedents and Consequences were missing to occasion and maintain the desired performance, respectively. Once a package intervention (consisting of task clarification, checklist, graphic and task-specific feedback, was implemented) cleaning behaviors began to occur regularly. The authors note that measurement by permanent product and design choice render conclusions about intervention tenuous as well as encourage future research to make use of more potent consequences (e.g., monetary incentives).

Amigo et al. (2008) sought to decrease table busing times at a popular pizza restaurant and used a modified form of the PDC as part of their functional assessment. While the authors do not report the specific results of the PDC findings, their initial intervention was to add antecedents through task clarification and goal setting. In a later phase, they add group and individual feedback as consequences to bolster the performance. Mean group table busing times decreased in each phase, most notably when consequences were added. Performance returned to baseline following removal of the intervention. The
authors note that individual busing times were difficult to represent due to varied work schedules, and that experimental control was difficult to assess with the inability to separate intervention components.

Gravina et al. (2008) used the PDC to inform intervention to increase morning preparation tasks at a physical therapy clinic. A package intervention was devised from PDC results, with a component from each indicated domain: task clarification, equipment manipulations and feedback corresponding to Antecedents and Information, Equipment and Processes and Consequences, respectively. Percentage of completed tasks increased from baseline across behaviors and maintained above baseline in a probe following intervention. The authors acknowledge that a component analysis would be beneficial and that future studies should compare indicated and non-indicated interventions.

Rice et al. (2009) extended literature on improving customer service behaviors using functional assessment by using the PDC to inform intervention on correct greeting and closing behaviors. The PDC indicated that Antecedents and Consequences were the problematic domains, leading to an intervention of task clarification and social praise. Staggered across each behavior, intervention led to increases in mean greetings and closings over baseline. Despite the effective intervention, the authors call for additional work validating the use of the PDC.

Lebbon et al. (2011) used the PDC to develop an intervention to improve safe performance of lifting procedures at a nursing care facility. The PDC indicated issues in all domains, resulting in a package intervention that included training, training on checklist
usage, weekly graphed feedback and daily verbal feedback. They evaluated the interventions using a reversal design and found mean performance increases across each of the three possible lifting methods at the facility. The authors mention that limitations included the absence of treatment integrity data, the failure to intervene on all indicated domains of the PDC and the failure to account for employee choice of lift procedure used. The authors broadly call for more safety research, which would be expanded on in subsequent iterations of the PDC.

In summary, utilization of the PDC in empirical studies is primarily restricted to informing intervention choice related to specific applied performance problems, in contrast to other subsequent PDC variants wherein the specific application of the tool is a variable of interest (e.g., Carr et al., 2013) or to investigate its psychometric properties (e.g., Wilder et al., 2019). Thus, most studies that utilized the PDC only investigated the singular impact of a PDC-informed intervention rather than how well the PDC identified an efficacious intervention. The potential accuracy of the assessment was hard to ascertain as well, as most interventions were designed to produce immediate robust change (i.e., through package interventions) and individual components were rarely examined (e.g., Austin et al., 2005). The PDC literature also does not make clear the extent to which researchers relied on the PDC. Typically, only the problematic domains were reported and, in some instances (e.g., Doll et al., 2017), the PDC was used in combination with other functional assessment procedures. Therefore, research involving the PDC suggests little about its impact on intervention selection. Data to ascertain the impact of PDC assessment on durability and intervention continuity is also limited.
Maintenance and Institutionalization in the PDC Literature

Follow-up data was measured in three of studies that utilized the PDC as a preintervention assessment. While the overall dearth of follow-up data is consistent with the broad trend in OBM publications (e.g., VanStelle et al., 2012), measurement of post-intervention effects is minimal within the studies as well. Follow-up data were collected between two and eleven months after conclusion of the studies, with Rice et al. (2009) having the most conclusive evidence of maintenance after the conclusion of the experiment (i.e., over 3 data points above baseline levels for both target behaviors). In contrast, Amigo et al. (2008) collected maintenance data following removal of the intervention and after an eight-week interruption in data collection. The average table busing time had increased to levels closer to baseline, suggesting minimal maintenance of the intervention effects. Gravina et al. (2008) also collected a single maintenance probe per activity checklist measured, suggesting performance somewhere between intervention and baseline levels.

Evidence of institutionalization is also limited in these studies. For example, Rice et al. (2009) demonstrated maintenance of employee performance for those that had been present for the initial intervention. Employees hired after the intervention, however, did not demonstrate similar levels of performance. In other words, the intervention effects maintained, but, presumably, the intervention itself was not maintained, evidenced by the disparate performance of new hires. Rice et al. provided some explanation of this effect, noting that it was both difficult and non-preferred for the manager to deliver positive feedback. Amigo et al. (2008) similarly noted that their follow-up data was solely based on
new hires. The mean busing times of subsequent hires were at levels above the target goal, again suggesting that intervention was not maintained by management. However, it should be noted that this is not to criticize these studies, as the long-term outcome was not the express purpose of these studies. Relatedly, it is relatively unclear if any procedures were designed in such a fashion to incorporate previously identified generality tactics (e.g., Conard et al., 2016), though it is arguable that “Involvement in Design” occurs by default through administration of the PDC.

The Performance Diagnostic Checklist- Safety (PDC-Safety) in Empirical Literature

Despite use in clinical, retail and other settings, the Performance Diagnostic Checklist had only been used to assess safe performances in one study (Lebbon et al., 2011). Incidentally, Behavior-Based Safety (BBS) lacked a formal, performer level tool to assess environmental variables surrounding risky performances. Martinez-Onstott et al. (2016) adapted the original PDC to meet this need, maintaining the same domains but writing questions for consultants and managers responsible for employee safety. The authors found that PPE usage was minimal likely due to ineffective consequences to maintain safe performance and devised a graphic and oral feedback intervention to remediate the deficit. Across three performers, PPE usage increased over baseline means.

In the initial PDC-Safety study, Martinez-Onstott et al. (2016) noted that additional research to compare the effectiveness of interventions is necessary to further validate the tool, akin to calls in PDC research (e.g., Rice et al., 2009). Cruz et al. (2019) conducted a
study to further evaluate the tool by comparing a PDC-Safety indicated intervention to a non-indicated intervention for three employees at a clinic serving individuals with intellectual disabilities. Following assessment, the Antecedents and Information domain was noted as most problematic for impeding frequency of hand sanitation. A non-indicated intervention, increased access to resources, was implemented and failed to improve performance. An indicated intervention, task clarification and prompting, was subsequently added resulting in performance increases for all three participants. The authors note that additional investigation is required due to constraints of the setting (i.e., clinical vs. industrial) and the arbitrary choice of non-indicated intervention. Moreover, the authors suggest further investigation into the validity of the tool.

Pugliese et al. (2021) extended the PDC-Safety literature further by applying using the assessment to troubleshoot poor adherence to policies requiring the use of personal protective equipment (PPE) in a private school. Direct care staff were supposed to wear various protective gear (e.g., thick clothing, pads, helmets) when working with students who exhibited certain topographies of aggressive behavior. Unique to this study, results from the PDC-Safety across three classrooms suggested three different treatment packages. One classroom received an intervention comprising feedback only, while in another additional PPE and an incentive program were selected. The third classroom received a combination of feedback and the incentive program. Each classroom showed marked increases in adherence to safe behaviors (i.e., wearing the prescribed PPE), while an analysis over time showed a reduction in the overall rate of injuries. Like other studies in the PDC family of literature, the researchers noted that the combination of different
interventions and the lack of a non-indicated control condition made it difficult to ascertain the relevant benefit of the PDC-Safety.

**Maintenance and Institutionalization in the PDC-Safety Literature**

Maintenance and Institutionalization in the PDC-Safety Literature. Of the three published studies utilizing the PDC-Safety, none contained follow-up data to investigate the post-study intervention effects. Like the PDC literature, lack of measurement of these variables is not problematic and likely ascribed to being outside of the scope of the experimental purpose (i.e., general validation of the tool). However, this remains a distinct concern for a BBS tool, given the history of literature on the long-term effects of more formal BBS programs (e.g., Myers et al., 2010) and the general need for safety programs to be maintained and internalized. For the latter concern, there is also minimal suggestion that these procedures were institutionalized. Martinez-Onstott et al. (2016) delivered feedback to participants themselves and, though Cruz et al. (2019) had supervisors deliver the PDC-Safety indicated intervention (i.e., feedback emails), no data were presented on the continuation of those emails. Pugliese et al. (2020) noted that the organization’s safety mission was consistently rated poorly on the PDC-Safety, suggesting that there were larger organizational changes required to sustain the goal, despite the initial success of the intervention.
The Performance Diagnostic Checklist- Human Services (PDC-HS) in Empirical Literature

The original PDC was developed to apply generally to performance problems in businesses and industries. However, like safety assessment, industries that provide intervention, treatment, care, and other supports to others have unique concerns. Human service settings are a common setting for OBM intervention (e.g., Gravina et al., 2016), and, thus, a variant of the PDC to address concerns specific to them was created. Carr et al. (2013) developed the Performance Diagnostic Checklist- Human Services (PDC-HS) by identifying problems unique and/or common to human services settings and adjusting the original PDC domains (i.e., Training, Task Clarification & Prompting, Resources, Materials, & Processes, and Performance Consequences, Effort, & Competition) as well as questions around them (e.g., question order and specific wording). The PDC-HS holds the distinction of having the most varied empirical support of all the PDC family of assessments (e.g., predictive validity, reliability; Wilder et al., 2020) and demonstrated generality across a variety of settings, performance problems, and implementers (Brand et al., 2021).

Predictive Validity

One common theme of the PDC-HS literature base is the focus on predictive validity. Researchers have commonly examined this by comparing implementation of a PDC-HS indicated intervention to an intervention not indicated by the results of the assessment. Carr et al. (2013), in developing and testing the tool, conducted one such study to improve staff cleaning of treatment rooms in an autism treatment center. Administration
of the PDC-HS indicated issues in the Training and Performance Consequences, Effort, & Competition domains. Thus, experimenters planned an intervention that included both pre-session training and graphed feedback. This intervention was implemented in a multiple baseline design across eight treatment rooms. Two rooms also received a non-indicated intervention consisting of task clarification and modifications to the availability of cleaning materials. The indicated intervention was effective in all treatment rooms and, for the two rooms that received a non-indicated intervention, completion did not improve until introduction of the indicated intervention.

Ditzian et al. (2015) extended this line of research by replicating the Carr et al. (2013) study with another performance concern. The PDC-HS was similarly administered to supervisors at an autism treatment center, but used to troubleshoot the performance of four staff members who were failing to secure treatment room doors, resulting in client elopement. Results of the PDC-HS indicated that the failure to secure doors was resultant of insufficient Performance Consequences, Effort, & Competition. Thus, intervention was planned which incorporated multiple forms of feedback (i.e., verbal and graphed). The consequence-based intervention was compared against a non-indicated intervention of Task Clarification & Prompting, in which a written prompt was affixed outside the treatment room. Results confirmed that of Carr et al. (2013), in that the indicated intervention was more effective than one not indicated by the PDC-HS.

Bowe and Sellers (2018) also compared the results of a PDC-HS indicated intervention against a non-indicated one. PDC-HS results from teacher-conducted assessment suggested that a lack of effective Training was contributing to poor
implementation of error correction during discrete trial teaching sessions conducted by paraprofessionals. Behavioral Skills Training (BST) was implemented first by the experimenters, then by the teachers to instruct the paraprofessionals. The four paraprofessionals met mastery criteria directly following the BST sessions, whereas an intervention not indicated by the PDC-HS, written task clarification and vocal prompting, did not. In addition to further confirmation of the predictive validity of the PDC-HS, Bowe and Sellers also demonstrated that assessment results could lead to effective intervention even with novice implementers (i.e., without formal behavior analysis experience).

Smith and Wilder (2018) continued to demonstrate the utility of the tool with novel implementers. The PDC-HS was implemented by supervisors to assess variables inhibiting accurately pricing items in a community thrift store. Assessment results indicated that Training was the most problematic domain, and an intervention that utilized components of BST was implemented. The training intervention contributed to increased accuracy of pricing, suggesting that the tool was effective in settings where both the staff and the supervisors administering the assessment were diagnosed with intellectual disabilities.

The validity of the PDC-HS has also been demonstrated even when implementing interventions from underutilized domains. Previous research typically suggested issues with Training and Performance Consequences, Effort, & Competition, and interventions were derived from these categories. Wilder et al. (2018), across two studies, assessed variables contributing to natural environment teaching of various verbal operants and use of prescribed equipment (i.e., timers). Results of study 1 and study 2 demonstrated the utility of interventions derived from the less frequently identified domains (i.e., Task
Clarification & Prompting and Resources, Materials, & Processes, respectively). In both instances, the indicated intervention was the most successful to improve the performance relative to non-indicated domains.

Most recently, Russell et al. (2020) extended research on predictive validity of the PDC-HS, using it to diagnose variables surrounding low rates of staff engagement with clients to promote social skills development. Across three participants, results of the PDC-HS indicated that the problematic performance was likely due to the absence of variables assessed in the Performance Consequences, Effort, & Competition domain. Russell et al. used an ABACAD design to evaluate the use of three interventions: one explicitly non-indicated intervention (i.e., prompts), a graphic feedback intervention, and a PDC-HS indicated intervention (i.e., public recognition for meeting criteria). Similar to other validation studies, only the indicated intervention produced consistent change to criterion for all three participants.

Applied Utility of the PDC-HS

Merritt et al. (2019) examined the use of the PDC-HS results to inform an intervention related to tardiness of direct care-staff. Application of the PDC-HS noted three generally problematic domains for the staff, including Task Clarification & Prompting, Resources, Materials, & Processes, and Performance Consequences, Effort, & Competition. Interventions from each domain were implemented, including the supervisor clarifying performance expectations, meeting with each staff to remediate barriers to timeliness, as well as providing graphic and corrective feedback. Moreover, interventions were subsequently modified as changes failed to sustain for two of the three participants.
and, in the case of the third participant, failed to produce timeliness. These modifications comprised removal of corrective feedback and addition of positive reinforcement contingencies (i.e., a token and lottery system) as well as individualized intervention for one participant (i.e., specific goal setting for token earning). Overall, interventions derived from the PDC-HS and subsequent modifications were able to produce some change in the tardiness of the three employees.

Hays and Romani (2021) used the PDC-HS to assess the variables maintaining inconsistent staff compliance with hand sanitation practices in an inpatient psychiatric unit. The PDC-HS results indicated that issues with hand sanitation practices were likely resultant of issues in the Performance Consequences, Effort, & Competition domain. Nine participants were provided with pre-shift feedback on the aggregate group hygiene performance, their personal performance for the previous shift, and the current unit goal for hand hygiene (i.e., 93%). Eight of the nine participants were able to reach the 100% compliance following the intervention. One participant required task clarification (i.e., information on missed/potential opportunities to engage in hand hygiene) to reach 100% compliance. Effects of the intervention maintained for at least one-week following intervention.

Melendez et al. (2020) conducted a replication of Wilder et al. (2018), targeting the mand training performance of three staff that provide in-home ABA service delivery. Administration of the PDC-HS indicated that all participants likely failed to engage in the target still due to a lack of adequate training. BST was conducted at two levels of frequency (i.e., every 1-4 sessions, after every session), with participants contriving the
highest frequency of mand opportunities after conclusion of the post-session BST phase. Following stability of performance during fading of BST, maintenance was assessed between 2-3 weeks later for each participant. All three participants maintained levels above baseline in follow-up measurement and both participants and caregivers rated the outcome of the training procedures favorably.

Collier-Meek et al. (2021) further demonstrated the utility of the PDC-HS to treat performance problems of paraeducators. Across five paraeducators, data from the PDC-HS indicated that a lack of Training and Task Clarification & Prompting were the underlying functional issues surrounding failure to implement student behavior plans. The participants were given a combination of Behavioral Skills Training (BST) and prompts regarding their students’ specific behavior plans. Four of the five paraeducators improved their implementation fidelity with addition of the PDC-HS indicated intervention. Collier-Meek et al. also collected data on the acceptability of the PDC-HS tool and its suggested interventions, as well as data on student outcomes. Despite the robust performance improvements of most participants, acceptability ratings and student outcome data were middling. Thus, it is suggested that applying the PDC-HS in traditional school settings with typical implementers may require additional considerations or revisions to the tool.

Analogue Studies on Reliability and Validity

While the PDC-HS literature demonstrated applied utility and predictive validity, fewer studies have evaluated psychometric properties of the assessment tool. Wilder et al. (2018) were the first to assess reliability and validity of the PDC-HS in an analogue setting. Twenty-one participants watched vignettes describing three different hypothetical
performance problems and scored the PDC-HS accordingly. Each of the three scripted videos featured a consultant and manager completing the PDC-HS together and a demonstration of the performance problem. The three problems ranged in the number of problematic domains indicated as an index of problem complexity. Despite these variations, participants scored each scenario highly accurately and consistently (i.e., across 2-4 weeks). These data suggested that the PDC-HS was both reliable and valid, though disparity between experience levels (i.e., degree type) suggest that the PDC-HS is best administered by a consultant or manager experienced in ABA. However, the high accuracy of the scores was potentially resultant of extremely straightforward nature of the scripted videos.

Cymbal et al. (2020) attempted to reduce the contrived nature of the videos by the Wilder et al. (2018) procedure, using scripts derived from actual consultant administrations of the PDC-HS. Twenty-one new participants watched and scored these videos twice, approximately 4 weeks apart, to assess reliability and validity. Notably, scores were lower than those measured in Wilder et al., though remained still accurate enough to diagnose the performance problem correctly. Measured interrater reliability was also lower than the previous study, though still within acceptable ranges to assert consistency. Differences in accuracy by experience level again confirmed that PDC-HS administration is likely best conducted by experienced practitioners.

Procedural Refinements to the PDC-HS

Despite utility across different supervisory populations (e.g., Merritt et al., 2019; Smith & Wilder, 2018), settings, performance problems and general empirical interest
(e.g., Wilder et al., 2020), there remain multiple ways in which administration and implementation of the tool may be difficult for inexperienced implementers. Implementer skill may reduce confidence in results (e.g., Cymbal et al., 2020; Wilder et al., 2018), while specific intervention choices derived from assessment have not yet been validated (Wilder et al., 2020). One potential solution is providing supplemental guidelines for use and administration of the PDC-HS (i.e., Brand et al., 2021). Alternatively, procedural modifications could reduce the amount of subjectivity in intervention selection. Vance et al. (2022) suggested two such refinements to the PDC-HS, that comprised a formal decision-making model (DMM) and establishing an empirically derived cutoff score for prioritizing problematic domains. Using a between-groups design, Vance et al. examined the impact of DMM and cutoff score availability on participants ability to select an intervention that correlates with intervention choices made in published empirical studies (i.e., intervention selection by experienced implementers). The group that had access to the DMM and cutoff score were much more likely to select an intervention that aligned with previous research and subjective ratings suggested that participants found the scoring to be more user-friendly. However, some participants ignored the DMM and instead continued to intervene upon the domain with the most “No” answers and only one-third of the potential DMM outcomes were validated.

Maintenance and Institutionalization in the PDC-HS Literature

Like the other PDC variants, research on the PDC-HS has typically eschewed measurement of post-intervention effects. Some studies (i.e., Cymbal et al., 2020; Vance et al., 2022; Wilder et al., 2018) utilized simulated performance problems, did not entail
applied intervention, and, thus, follow-up measurement was not possible and/or warranted. Most of the remaining empirical studies were intended to evaluate the predictive validity of the assessment relative to non-functional (i.e., not indicated by the PDC-HS) intervention, rather than measure the strength of post-intervention impacts and did not measure maintenance of the targeted performance. Two studies formally measured maintenance of intervention effects. Melendez et al. (2020) conducted follow up and found that participants maintained their performances 2-3 weeks following training. Hays and Romani (2021) also measured handwashing compliance after a one-week delay and with a single day re-introduction of the performance feedback intervention. Notably, though participants did not meet overall institutional goal for hand hygiene (i.e., 93% of opportunities), performance in the maintenance phase, albeit brief, was at 100% for all participants. The inclusion of additional components beyond what was indicated by the PDC-HS (i.e., goal setting) made it difficult to assert the degree to which the PDC-HS contributed to the observed maintenance effect as well as the specific environmental variables that produced the change. Melendez et al. (2020) also noted that, due to lack of comparison to non-indicated intervention or use of other assessments, how well the PDC-HS effectively identified the correct intervention. Both studies also conducted brief follow-ups which limit conclusions on the stability of maintenance.

Two studies in the PDC-HS literature (Collier-Meek et al., 2021; Merritt et al., 2019) did not formally assess maintenance, but were notable in that intervention produced inconsistent results across all participants. Moreover, these non-results occurred despite experienced researchers administering the assessment, selecting the intervention, and
monitoring implementation. Reasoning behind the recent proposed procedural modifications to PDC-HS administration could explain these outliers (i.e., Brand et al., 2021; Vance et al., 2022), in that there is subjectivity inherent to the PDC-HS intervention selection process (i.e., the domain was chosen correctly, but different and/or additional intervention selections could have been more effective). Therefore, even though maintenance was not measured, these studies can potentially provide valuable insight into the complexity of distilling a performance problem down to its strongest maintaining variables and how the PDC-HS might relate to the institutionalization of intervention practices. Merritt et al. (2019), intervening on tardiness, note that most participants had a long history of being late to work (e.g., greater than one year) and that the site was in an area that frequently experienced traffic and construction related delays in addition to regular interruptions of public transportation. Because tardiness seemed to be partly accounted for by external variables outside of researcher/performer control, was unlikely to contact detrimental job-related outcomes (e.g., termination of employment) and faced competing direct contingencies to formal policy (e.g., pay reduction only occurred after 7 minutes late arrival), it is possible that interpretation of functional assessment missed key opportunities in the altering processes, environmental barriers, and contingencies (i.e., Resources, Materials, & Processes, and Performance Consequences, Effort, & Competition categories). Collier-Meek et al. (2021) too found that student outcomes were relatively unchanged by functional intervention on the paraeducators implementation of behavior plans. Notably, the paraeducator whose performance was not impacted by intervention was assigned the only two students whose measured outcomes (i.e., academic engagement and rate of disruptive behavior) improved throughout intervention.
In other words, the improvement was resultant of factors outside behavior plan implementation and may have produced consequences that competed with the targeted performance. To bolster this assertion, the development and quality of the behavior plans was a noted uncontrolled variable, and the written plan, if poor quality, may not have been the best avenue to desirable student outcomes. In addition, Collier-Meek et al. (2021) suggest that layered supports within classrooms may be necessary to account for the various variables that influence paraeducator behavior. Paraeducators that participated in the study reported that, despite broad administrative favor for implementation of plans, they likely needed more formal external supports to engage in the targeted performance, ranking systems support for implementation relatively low. Thus, there is the suggestion that the intervention could have been bolstered through effort to institutionalize the intervention. Such results highlight the importance of institutionalization. If intervention resultant of assessment produces important results (i.e., client outcomes) and has high acceptability, it has been hypothesized that organizations may be more likely to adopt procedures (Boyce & Roman, 2002). Melendez et al. (2020) confirmed this, albeit anecdotally, reporting that the organization adopted the intervention (i.e., BST procedures) for all new staff. However, variables to support institutionalization might still need to be diagnosed independent of other concerns. Existing processes, policy, and contingencies may not support the target of the performance change (i.e., Merritt et al., 2019) and functional treatment of a specific problem, even if well perceived, may not necessarily ensure that practices will sustain or impact important results (i.e., student outcomes; Collier-Meek et al., 2021).
Rationale and Purpose of the Present Study

In summary, while the study of lasting change has historically been cited as a critical feature of OBM and ABA applications, techniques to produce generality have not progressed in line with other behavioral technologies. In the past, extrapolation and imprecise invocation of terms and concepts from basic research literature tended to overstate the importance of specific factors contributing to maintenance by primarily ascribing maintenance of change to transfer of training techniques. Stimulus generalization, a phenomenon that has been thoroughly studied in basic research, denotes organisms’ responses under similar stimulus conditions. Superimposing this basic assay to applied technology (e.g., Stokes & Baer, 1977) yielded a set of methods that remain narrowly applicable, mainly to a particular set of performance deficits (i.e., transfer of control past training conditions). While such taxonomies in OBM have expanded beyond the scope of generalization (e.g., Conard et al., 2016; Sigurdsson & Austin, 2006), general techniques are underused, perhaps due to the tendency to overlook the core conceit of behavior analysis, that ongoing performance is resultant of functional relationships. The result has been a simultaneous overcomplication in terms of recommending myriad probabilistic techniques and an oversimplification in failing to encapsulate the mechanism that produces the phenomena of interest (i.e., the relationship between variables that predict or preclude intervention success). Thus, these methods, having limited utility and explanatory power in research, in concert with prevailing concerns in research methodology, have likely led to infrequent experimental evaluation of maintained intervention effects.
The current, common method to examine maintenance is to aggregate and examine studies post-hoc (e.g., Conard et al., 2016; Sigurdsson & Austin, 2006). Aggregating study results to assess the overall utility of a given intervention (e.g., reviews and meta-analyses) is a valid way to demonstrate utility of certain methodologies, but this is more likely to assess an overall generality of procedures rather than demonstrate the functional relationships present in the specific intervened upon environments. Additionally, such tactics may be subject to publication biases or other methodological issues (Baer, 2001; Sigurdsson & Austin, 2001). Regardless of method, no studies to date have asserted that temporal generality criteria are required or actively contribute to the spread of intervention effects across time. Most techniques cited specifically to promote temporal generality remain unevaluated empirically, particularly in OBM literature. In applied literature, because the maintenance of an intervention is, generally, secondary to its efficacy in most research studies, variables maintaining the intervention effects are rarely observed, if manipulated at all. This extends to examining the degree to which intervention practices are institutionalized by the target organizations as well.

However, the methodology to examine the persistence of intervention effects already exists within behavior analysis. Diagnosis of maintaining variables is already a hallmark of ABA (e.g., Iwata et al., 1982; Mace, 1994) and OBM (e.g., Austin et al., 2000) practices. The latter may have an advantage in examining maintaining change. In comparing OBM to ABA, Austin et al. (2000) noted that one key difference in functional assessment methodology is that assessment focuses on the variables needed to produce a target performance, rather than identifying variables maintaining a problematic
performance. As such, the functional assessment process of OBM typically entails gauging the target environment’s suitability and receptivity toward the proposed performance change. In some ways, this type of functional assessment looks at similar variables to the previous reviews on maintenance. For example, some items from the PDC-HS (e.g., asking a supervisor about natural consequences of task completion and what types of training a performer received; Carr et al., 2013) might correlate well to some of the temporal generality tactics (e.g., Introduce to naturally maintaining contingencies and Use of instructional design features; Conard et al., 2016). However, most critically, these are diagnosed in relation to the specific target environment (i.e., functionally). Rather than attempting to identify how many tactics were necessary to sustain performance based on previous literature, a functional assessment asks which of those variables are absent from the target environment to produce the necessary change. OBM assessment tools (e.g., Johnson et al., 2014; Wilder et al., 2018) thus, can potentially better conceptually account for factors important to maintenance and institutionalization than taxonomies.

Since the initial call of Austin et al. (2000) to broaden the use of functional assessment in OBM, the use and variety of tools used has arguably diversified. However, a prevailing concern in the literature surrounding OBM functional assessment is the predictive validity of such tools. Of those most cited tools, the PDC-HS remains the most empirically backed, both in validity, reliability, and applied utility. Yet, to date, no thorough examination has been made of how well any of these tools directs intervention to effective performance change. While there is a necessary empirical focus on initial intervention efficacy, focus on post-intervention evaluation is equally important. As Malott
(2001) noted, mere implementation and awareness of behavioral principles will not lead to long-term course correction of a deficient environment, otherwise the performance problems and demand for behavior analysts would be relatively non-existent. Rather, a deterministic perspective suggests that an organization already operates as optimally as possible, subject to the current and historical interplay of environmental variables and its constituents. Moreover, the inherent complexity of organizational environments suggests that whether the performance is well pinpointed or if a behavior-analytic intervention produces a demonstration of effect, this does not speak to whether these practices will be selected and transmitted through the organization. In this regard, institutionalization should likely not be treated as an ancillary effect to intervention but as another key area to investigate in the functional assessment literature.

As OBM technology develops, so should tools used to diagnose and treat the functional causes. Even in the more well-established ABA functional assessment literature, the link between assessment and outcome is tenuous, given the definitional issues that exacerbate study of maintenance patterns (e.g., Pennington et al., 2019) or the potential ephemeral nature of problem behavior functions (e.g., Lerman et al., 1994). Examining OBM tools’ ability to predict maintenance or adoption patterns may provide some critical insight into why some interventions may be rejected or discontinued, despite obvious effectiveness in isolation (i.e., under stringent research conditions). As in therapeutic ABA applications, OBM practitioners often intervene when risks are high (e.g., in concerns of employee safety). In this regard, it remains essential to demonstrate that functional assessment tools not only provide impacts that are robust and effacious, but also have the
potential to endure. Thus, a needed first step is evaluating how well current functional assessment tools predict maintenance of the intervention and its impacts when formal involvement is less obvious. That is, when natural contingencies have a chance to develop and compete. Therefore, the purpose of the present study is two-fold. The first purpose is to investigate and elaborate on validation of a commonly used OBM assessment, the PDC-HS, regarding its ability to identify effacious, enduring interventions absent apparent researcher involvement. The second purpose is to expand and, potentially, provide an experimental paradigm to study maintenance and intervention adoption by examining the influence of function-based treatment on institutionalization of an OBM intervention.
Chapter 5
Method

Setting and Participants

A sample of eight behavior technicians and four supervisors (i.e., Board-Certified Behavior Analysts® [BCBAs®] and Board-Certified Assistant Behavior Analysts [BCaBAs®]) were recruited through the intraoffice communication platform within an ABA service delivery organization located in the southeastern United States. The behavior technicians who reported demographic information (n = 6) all identified as female, ranging in age from 22 to 36 (M = 26.4 years old). Three were non-Hispanic and White, two were Black, and one was Asian or Pacific Islander. Most participants reported between 1-2 years of experience in ABA services, with half holding a Bachelor’s degree, while there was one participant with a non-ABA Master’s Degree and one with an Associate’s degree. All but one participating behavior technician held their RBT® certification. Participating supervisors (i.e., BCBAs® and BCaBAs®) predominantly identified as female (3 of 4) and held their respective certifications between 1-3 years. Three of four supervisors were non-Hispanic and white, while one was Asian and/or Pacific Islander. A third non-participating group of four behavior technicians’ session data was sampled as an untreated control group. In this group, three of four identified as female and ranged in age from 22 to 39 years old (M = 30 years old). Control group behavior technicians had between one and five years of experience in ABA services. The organization comprises two locations in a major metropolitan area, providing primarily early intervention services to children with autism and other developmental disabilities in clinic, school, and community settings. The present
study was conducted across both clinic locations. At the time of assessment, there were 32 active individuals providing direct therapy and 9 clinical supervisors serving approximately 50 clients. Among those clinical supervisors, three also had administrative duties. Two were Clinical Directors, while the third oversaw Operations Management.

Supervisors with administrative duties and their cases were excluded from either intervention group, as their input informed the functional assessment data. However, session data from their cases comprised the basis for the control group. Behavior technicians at these clinics are commonly, though not always, Registered Behavior Technicians® (RBTs®) tasked primarily with implementation of function-based behavioral support procedures (e.g., behavior reduction and crisis management), skill acquisition and self-care programs, as well as operational duties including documentation, data collection, and cleaning tasks. All behavior technicians in the study met all necessary organizational training (e.g., onboarding, didactic, and client-specific training) and credentialing requirements (e.g., a 40-hour RBT® Training, insurance approval) to work independently with clients; this process usually ranged from one to three months from the initial hire date. This is both for purposes of experimental design (i.e., to avoid any potential training and knowledge deficits) and as a pragmatic criterion (i.e., attrition was common in the first calendar month of employment, prior to completion of training/credentialing requirements). Within the organization, BCBA® and BCaBA® duties are generally related to case management; they are responsible for treatment planning, upkeep of documentation and programming, and supervision. To establish experimental groups, the four supervisors were randomly sorted into two groups by a web-
based randomizer. Behavior technicians were then recruited from their caseloads, specifically those that 1) served a single client consistently throughout the week (i.e., 3 or more days of service), 2) were tasked primarily with teaching language acquisition and foundational skills, rather than treating challenging behavior, and 3) who had noticeable deficits or consistent low levels (relative to opportunities) in providing learning opportunities per hour, based on visual inspection of level and trend for current sessions. These criteria were confirmed by the scheduling team and their respective supervisors. Inclusionary criteria were established to maximize the potential for improvement and frequency of session data generated, as well as to minimize the impact of challenging behavior on providing learning opportunities. As the desired level of performance varied with supervisor preference and due to idiosyncratic variables (i.e., individual client history, session structure), there were no firm exclusionary performance criteria for the primary dependent variable.

**Dependent Variables**

Informal interviews were conducted initially with clinical and executive leadership that supervise service delivery or manage programs at these locations to identify the presence of performance problems pervasive at the behavior technician level. While there were multiple potential pinpoints identified, variables were considered for intervention only if they were a problem of operational significance, supervisors expressed willingness to intervene and monitor within the parameters of the present study, and were, presumably, not solely resultant of knowledge or skill deficits. Last and, perhaps, most critically, intervention on the target performance problem could not directly impede or reduce current
standards of care (e.g., crisis management procedures, hygiene, health, and well-being, etc.). With these criteria, interviews indicated that a pervasive problem existed in the rate of learning trials presented and recorded during session. With fewer trials per data point or simply no data present at all, this impeded the supervisors’ ability to make timely and accurate data-based decisions about updates and programmatic modifications.

Primary Dependent Variable

All supervisors and behavior technicians within the organization use a software platform for electronic data entry, case management, and billing. At onboarding, staff were trained on the software platform and were provided the expectation of providing learning opportunities and data recording throughout their session. Learning opportunities were defined as single trials (i.e., instruction/cue, behavior, consequence) of programs for acquisition skills (e.g., communication, language, daily living skills). While there was no consistent organization-wide performance goal for learning opportunities (or a desire to create such a goal), most supervisors interviewed suggested at least five trials per target would be preferred; clinical directors expressed a desire for at least fifty to sixty learning opportunities per hour. The median number of trials ran by participants in baseline were 17.84 per hour; occurrences or intervention on problem behavior were not included in these numbers at any point during the study. Therefore, the primary dependent variable selected was the frequency of learning opportunities per hour (LOPH). This was calculated as the number of acquisition targets run in a session, divided by the total duration of session, in hours. These data were primarily measured by permanent product, as the primary researcher could sample session data from the system after sessions concluded,
which were already converted into a per hour count. Any discrepancies in time recording were resolved by comparing the recorded session duration to the appointment times billed to insurance.

Independent Variables and Secondary Dependent Variable

Initial reports, confirmed formally through administration of the PDC-HS to clinical directors, suggested the performance deficit was predominately caused by a dearth of available, effective resources and processes for completing trials. Therefore, the indicated intervention would be providing more quality resources for program stimuli as well as a process for ensuring that they are accurate and organized for ease of use. Multiple items on the PDC-HS require confirmation through observation, including items 2, 3, and 4 in the Resources, Materials, and Processes domain. The primary researcher noted that client materials were stored in cloth storage bins that were 27.94 cm x 26.67 cm x 26.67 cm. Without any additional organizing materials or storage, materials were often stacked on top of each other, and, in many cases, buried under multiple layers of items. This caused some materials to be damaged or frayed and many programs outlined in the data tracking software lacked the necessary visual stimuli to implement them (see Appendix A for baseline sample images). After some discussion with onsite clinical directors and the company owner, a storage solution was approved that could house visual stimuli for programs, smaller preferred items, and other necessary task stimuli in a clear, three-tiered stacking storage box. The storage box was large enough to house most program stimuli for clients, but small enough to fit within the larger client bin (24.13 cm x 16.51 cm x 18.31 cm; see Appendix B). The dimensions of each compartment could be altered to
accommodate stimuli of various sizes, up to the size of a 7.62 cm x 12.70 cm index card. By condensing and organizing these stimuli, it also permitted organization of other, larger stimuli within the primary bin to make them more readily accessible. The stackable bins were also chosen so that behavior technicians could lay them out in the instructional area to have easier access during session (see Appendix C for a picture of the storage solution and an organized client bin). Furthermore, all compartments were labelled, and a template was created for “program cards” that could be inserted in the compartment. This was added as an element to the material creation process to provide dated inventory of necessary stimuli for the program, as well as establish program parameters as behavior technicians noted that instructions were difficult to find while also running session and recording data (see Appendix D for a sample program card).

In addition to serving as the primary intervention, implementation of the PDC-HS indicated intervention (i.e., accessible, accurate, and organized resources as well as an easier inventory process) by supervisors served as a secondary dependent variable to assess institutionalization. Specifically, supervisors were measured on procedural integrity related to maintaining these resources. Procedural integrity was chosen as the primary measure of the secondary dependent variable because lapses in fidelity have been observed to potentially have detrimental impacts on intervention outcomes (see Brand et al., 2019 and Fryling et al., 2012 for discussions). Within OBM research, procedural integrity remains under measured and investigated (e.g., Cymbal et al., 2021) and low rates of reporting has warranted exploration in multiple behavior analytic journals (e.g., Han et al., 2022). To create a standard measure for the secondary dependent variable, another set of PDC-HS
interviews were conducted to determine which variables might promote supervisors maintaining the quality of resources (i.e., institutionalization). These interviews suggested that supervisors did not receive feedback on the upkeep of client materials, nor did they have any sort of job aid or checklist for maintaining client materials. A job aid/checklist was developed with approval from the clinical directors and this document served to score the fidelity of the resource intervention and a scored checklist was also delivered as a component of feedback to supervisors (see Appendix E). The checklist comprised thirteen items, and each item was scored as Yes, Somewhat, No, or, N/A for materials related to antecedent strategies, teaching stimuli, and consequence-based interventions. Yes answers indicated that all stimuli met criterion for that question and were scored as 1 point. Somewhat indicated that most (i.e., more than fifty percent) stimuli met criterion for that question and were scored as 0.5 points. An answer of No meant that none of the stimuli met criterion for that question and were scored as 0 points, while an answer of N/A meant that procedures as written did not specify any stimuli that would meet that criterion. A percentage was calculated by adding all the scores for each question, divided by 13 minus the number of N/A answers, multiplied by 100. This score was used to determine fidelity of the intervention each week across both experimental groups but was only delivered to the second group.

The second group of supervisors received feedback as a completed checklist with an accompanying written statement approximately bi-weekly (see Appendix F for the feedback template). Prior research suggests that effective feedback should likely include evaluative (i.e., valence of the performance) and objective (e.g., featuring a numerical
score) components (Johnson, 2013), with specificity (e.g., highlights various dimensions of the performance; Park et al., 2019) and accuracy (Brand et al., 2020; Lee et al., 2020; Palmer et al., 2015). Moreover, feedback is likely more effective when delivered frequently (e.g., Park et al., 2019) and most efficacious when provided daily. However, such frequency of feedback may not be feasible for some applied settings, such as human service organizations, and longer intervals between feedback have proven effective (e.g., bi-weekly or monthly; Sleiman et al., 2020). Initially, for this study, weekly feedback was proposed, as this has been demonstrated previously to be potentially as effective as daily delivery (e.g., Pampino et al., 2003). However, to promote effectiveness (i.e., supervisor-delivered feedback is amongst the strongest demonstrated effects; Sleiman et al., 2020) and to cultivate a more natural contingency, clinical directors were asked to deliver feedback. After minor delays to completion of the initial cycle scoring the checklist, discussion with the participating clinical director resulted in the procedure being amended to accommodate their other operational duties and to mirror the typical interval for regular program and material updates. Thus, the written feedback and job aid was delivered bi-weekly. For the final participant in the second group, the other clinical director initially agreed to deliver feedback, but then subsequently refused participation when feedback was due. Therefore, feedback was delivered by the primary researcher to that supervisor.

Thus, procedural integrity parameters for the feedback were defined as 1) the presence of the completed checklist as graphic and written feedback, 2) an evaluative statement, 3) objective data, 4) specific information about performance, and 5) delivered bi-weekly via email, per performer.
Interobserver Agreement and Procedural Integrity

Interobserver agreement (IOA) was calculated by the primary researcher and a secondary observer. For the primary dependent variable, LOPH, the primary researcher collected data for all participants via review of permanent products (e.g., review of session data and billed appointment lengths) and reviewed these twice as an accuracy check. The secondary observer independently reviewed a random sample comprising 40% of de-identified participant data across baseline and intervention phases. IOA was calculated as total agreement (i.e., the number of agreements divided by the total number of agreements plus disagreements, multiplied by 100; Johnston et al., 2020). Observed IOA for the primary dependent variable was 97.80% across all participants. IOA was 100% for all participants in Group A and the Control group. In Group B, IOA was 100% for Participant 1B, and 90% for Participants 2B and 2C. For Participant 1B, this was a calculation error of the session time by the primary researcher, resulting in a lower LOPH value during baseline. This was corrected in the final data set. For Participant 2C, this was an error introduced during review as the file provided was for the incorrect session.

IOA was also calculated for measurement of the secondary dependent variable, procedural integrity of the PDC-HS indicated intervention for behavior technicians. The primary researcher and participating clinical director independently scored fidelity for 30% of total observations across both groups. IOA was calculated by dividing the smaller integrity score by the larger integrity score and multiplying that by 100. IOA averaged 94% for the first group and 92% for the second, for an average of 93% across both groups.
Procedural integrity for the primary independent variable was scored as outlined above, based on the job aid/checklist. For the first group, which only received the PDC-HS indicated intervention for behavior technicians, procedural integrity ranged from 38% to 88% ($M = 64\%$). Within the second group, which received a PDC-HS indicated intervention for both behavior technicians and supervisors, procedural integrity ranged from 79% to 100% ($M = 81\%$). The mean procedural integrity across both groups was 79%. For the secondary independent variable, delivery of the checklist and feedback, procedural integrity was monitored via a manipulation check wherein the primary researcher provided a sample script and checklist with each specified element (e.g., including objective and evaluative information) to the clinical director intraoffice communication prior to each delivery. The participating clinical director then blind copied the feedback email to the primary researcher for review. While the director was encouraged to deviate from the script in non-critical ways to reflect their personal communication style, they chose to adhere closely to the prompt in all observed instances.

**Experimental Design and Procedure**

Within subject examination of functional assessment impacts was examined across two groups, each represented by a concurrent multiple baseline designs across participants, one for each assessment combination. The first group received a PDC-HS indicated intervention at the behavior technician level only, to observe the impact of the intervention over time with minimal researcher involvement. This group simulated a traditional examination of maintenance in research literature (i.e., introduction of the intervention and removal of researcher involvement). The second group received the same PDC-HS
indicated intervention for behavior technicians, as well as a secondary functional
intervention derived from the PDC-HS to promote supervisory behavior maintaining the
initial intervention (see below for a detailed description). Data collection was conducted
throughout the week, with the primary researcher pulling session data the following day.

In a multiple baseline design across participants, intervention is introduced to the
first tier (i.e., participant) when responding is stable (i.e., zero or minimal trend) or when
directionality of the data is opposite the presumed treatment effect in all baseline tiers
(Johnston et al., 2020; Kazdin, 2011). Intervention is introduced in staggered fashion to
subsequent tiers utilizing this same logic. Experimental control in multiple baseline designs
is asserted through comparison within tiers (i.e., a single AB design) as well as across tiers
(i.e., between the different behaviors or participants). When the intervention produces
change in the dependent variable relative to baseline data as well as in relation to the
baseline of the untreated tiers, experimental control can be asserted (Lanovaz & Turgeon,
2020; Slocum et al., 2022). Each additional tier in the design that replicates the treatment
effect strengthens confidence in experimental control, though three tiers and control
demonstrated in at least two of those three tiers are likely sufficient in most cases (Lanovaz
& Turgeon, 2020). In the present study, this means that experimental control was
demonstrated if the PDC-HS derived intervention(s) produced visible change in level or
trend, while all subsequent tiers’ baselines remain stable.

In addition to the single-subject design, another purpose of the present study was
to conduct a preliminary analysis of differences in each condition over time. To examine
these between subject interactions, these data were analyzed using a mixed repeated
measures analysis of variance (ANOVA). An additional participant per group was introduced to the intervention phase simultaneously with the third participant in each multiple baseline design. Data from a group of behavior technicians not exposed to the intervention were also collected during the same period as the multiple baseline designs. These technicians were randomly selected from the clinical directors’ caseloads, by a web-based randomizer. Their data served as a control group for the statistical analyses. Data for these analyses were measured as weekly averages through baseline and five weeks of intervention (i.e., 1 week of baseline, 1st week of intervention, 2nd week of intervention, 3rd week of intervention, etc.).

Power Analysis

To estimate an adequate sample size for the statistical portion of the study, an a priori power analysis was conducted with G*Power, version 3.1.9.2 (Faul et al., 2007). Previous empirical work validating the PDC-HS (see Wilder et al., 2020 for a review) has almost exclusively utilized single-subject research designs. Thus, published comparisons between PDC-HS indicated and non-indicated interventions have been evaluated visually and, though comparisons have shown robust differences, no statistical representation of effect size have been noted. However, metanalyses of prospective intervention components (e.g., performance feedback from supervisors) utilizing non-parametric tests have yielded relatively large effect sizes in behavior analytic literature (i.e., Sleiman et al., 2020). Therefore, conservatively assuming a medium effect (i.e., Cohen, 1988) in concert with the planned parameters for the present study (i.e., number of groups and measurements), results suggested that a total sample of N = 18 at 80% power would be adequate. These
data suggested that each group should comprise approximately 6 participants. However, due to rates of turnover, operational changes in scheduling, and some supervisors electing to not participate, the final total of participating behavior technicians was 12, with 4 behavior technicians per group.

Pre-intervention Assessment

Prior to the experiment, the primary researcher scheduled two independent meetings with the clinical directors that oversee staff at each clinic location. During these meetings, the primary researcher briefed them on the purpose and phases of the project and administered the PDC-HS. The PDC-HS is a 20-question informant- and observation-based functional assessment tool (see Appendix G) used to identify variables contributing to a pinpointed performance deficit across four domains: Training, Task Clarification & Prompting, Resources, Materials, & Processes, and Performance Consequences, Effort, & Competition (Carr et al., 2013). The PDC-HS was administered twice to each, first to identify potential causes of low learning opportunities provided per hour and, second, to identify potential barriers to supervisors (i.e., BCBAs® and BCaBAs®) providing those variables regularly. The assessment tool is conventionally completed by a consultant and manager, with seven items additionally verified through direct observation by the consultant (Cymbal et al., 2020). In research, formal scoring has customarily been quantified by calculating a “negativity” score, or the percentage of questions answered “No” in each domain (Vance et al., 2022). This negativity score indicates the domain most rife with opportunity for intervention, though it should be noted that any “No” answer can indicate a potentially useful intervention (Carr et al., 2013).
As the intervention is troubleshooting a pervasive performance problem identified organization wide, the two clinical directors scores were compared for agreement on the prospective intervention components. If results of the assessment had not agreed on a problematic domain (i.e., one indicates only issues with Training while the other indicates only issues in Performance Consequences, Effort, & Competition) or if there was no agreement within the problematic domain (i.e., Performance Consequences, Effort, & Competition is indicated, but the answers differ in all questions), then the PDC-HS would have been administered a third time to the company’s owner to further clarify the performance problem; interventions would have been chosen based on the highest agreements between the three results. However, agreement between clinical directors’ PDC-HS answers for both behavior technicians and supervisors using the total agreement method was 90%. Results indicated that the most problematic domain (i.e., exceeding the proposed “cutoff” threshold of 50% “no” answers; Vance et al., 2022) was Resources, Materials, & Processes for behavior technicians (see Figure 1). For supervisors, it was suggested that the most problematic domain was Performance Consequences, Effort, & Competition (40% “no” answers) with Task Clarification & Prompting (20% “no” answers) as a less significant concern (see Figure 2).

Baseline

Participant data for baseline phases were collected through review of electronic session records. During review of baseline data, the researcher did not contact any behavior technicians, supervisors, or clinical directors regarding performances or data collection. These data were aggregated as part of previous operational duties of the primary
researcher. When participants had completed informed consent and met inclusionary
criteria, they moved to the relevant intervention phase.

**Group A: PDC-HS Indicated Intervention Only**

The first group of behavior technicians received only the PDC-HS indicated
intervention, improved access to, redesign, and reorganization of materials, as well as
changes to the task process. This intervention comprised sorting and organization of
materials, provision of the stackable storage solution, additional labeling for stimuli, and
prompts (i.e., program cards). During time outside of client sessions, the primary
researcher reviewed all client programs in the data collection platform and took inventory
of existing stimuli in the client bins. All stimuli were organized and, if they were not
already, were clearly labelled. Program cards were added for all existing programs, with
instructions and stimuli copied from protocols specified in the case management platform.
These cards were added to allow supervisors to conduct inventory more readily and to
serve as a quick reference for behavior technicians when running a session. When the
materials were completed and organized, they were introduced and demonstrated to the
supervisor and behavior technician, separately. Participants were told that the new
materials were designed to help sessions run more smoothly. The primary researcher
showed them how the stacked storage boxes worked and where program stimuli were
located or added. No additional instructions were given nor were learning opportunities or
data collection mentioned at any point. Because the indicated intervention was primarily
centered around providing stimuli and organization, there was no formal intervention or
maintenance phase delineated, as researcher involvement could be ceased as soon as new
materials and organizers were put into place. If supervisors from this group asked the primary researcher for additional materials to be made following subsequent program changes, they were redirected to the normal, pre-existing process for material creation (i.e., staff whose clients cancelled were paid at a reduced rate for time making requested stimuli and other programmatic materials).

Group B: PDC-HS Indicated Intervention and Indicated Intervention for Supervisors

The second group of behavior technicians also received the PDC-HS indicated intervention (i.e., material organization and availability), like the first group. Similarly, once the intervention was put into place, maintaining the intervention was left to the supervisors. However, in contrast to the first group, supervisors also received a PDC-HS indicated intervention to promote maintenance of the materials (i.e., institutionalization) simultaneously with the behavior technician intervention. For supervisors, the indicated intervention based on PDC-HS results was written feedback (i.e., from the Performance Consequences, Effort, & Competition domain) and a job aid/checklist (i.e., from the Task Clarification & Prompting domain) delivered bi-weekly. The feedback statement was sent from the relevant clinical director’s email account every other week, typically at the start of the work week as checklist observations were conducted at the end of the week. Their email contained both evaluative and objective data on client material organization as outlined above.
Procedural Acceptability Survey

In the present study, social validity was assessed through procedural acceptability. Participant report on the appropriateness and usefulness of the experimental procedures and intervention is pertinent in OBM research, both to avoid undesirable impacts and to, potentially, influence the long-term adoption of intervention strategies (Gravina et al., 2021; Nastasi et al., 2021). Gravina et al. (2021) specifically recommend that procedural acceptability measures are designed to assess a representative sample of the participant population and/or organization and that measurement occurs at multiple points throughout the experiment. These data can then be used to inform changes to intervention (e.g., to maximize intervention impacts or improve cultural sensitivity) or examine the relationship between intervention effectiveness and subjective reports of acceptability. In this study, procedural acceptability was assessed through a five-question survey scored on a five-point Likert scale, with an additional open-ended response for additional feedback. For supervisors that received the secondary intervention, they received a form with four additional questions regarding the checklist and feedback received. Questions asked about the perceived acceptability of procedures for behavior technicians, supervisors, and clinical directors (see Appendix H for the survey content). The survey was provided to all participants at the conclusion of the formal experiment.
Chapter 6
Results

Baseline Group A

Overall, participants in Group A provided learning opportunities per hour (LOPH) ranging from 0 to 33.58, with a median of 13.74 ($M = 14.38$). Participant 1A provided LOPH ranging from 3.42 to 11.56 ($M = 8.28$). Participant 2A’s LOPH ranged from 0 to 13.74 ($M = 6.53$). LOPH for Participant 3A ranged from 8.52 to 33.58 ($M = 20.01$). Figure 3 depicts the baseline and intervention phases across consecutive sessions for Group A.

Intervention Phase Group A

Following introduction of the resource and material-based PDC-HS indicated intervention, all participants in Group A generally increased provision of LOPH. During the entirety of the intervention phase, participants’ LOPH ranged from 4.08 to 74.59, with a median of 24.23 ($M = 28.85$). Daily session data tended to be variable, but there were consistent observable increases across weekly averages except for the fifth week of Participant 3A (see Figure 4 for average weekly data for Group A participants).

Participant 1A

Participant 1A was one of the lowest LOPH scores through baseline ($M = 8.28$) and, following intervention, demonstrated the smallest amount of change across both groups following the introduction of the independent variable. In the intervention phase, Participant 1A’s LOPH ranged from 4.08 to 31.07 with a median of 14.42 ($M = 15.11$). A
slight level change was observed for the first week of intervention but with significant variability that, at multiple points, overlapped baseline data. Following the initial intervention week, an overall increasing trend was observed, and though, Participant 1A’s improvement was, at times, more than triple their baseline mean, this still represented the smallest overall change in either experimental group (see Figure 5 for isolated data for Participant 1A).

Participant 2A

Participant 2A had the lowest mean LOPH scores through baseline ($M = 6.53$), though this was skewed by multiple sessions in which they did not record any trials. Although there was greater variability and decreases in LOPH in the latter half of intervention, Participant 2A remained on an increasing trend overall throughout the study and reached LOPH approximately four times their highest baseline observation (see Figure 6 for isolated data for participant 2A). Overall, in intervention, Participant 2A’s LOPH ranged from 16.83 to 53.49 with a median of 26.56 ($M = 30.57$).

Participant 3A

Of the three participants in Group A, Participant 3A had the highest LOPH mean in baseline ($M = 20.01$), though the trend was decreasing overall. In intervention, their data had the highest peak in Group A, ranging from 30.49 to 74.59, with a median of 54.48 ($M = 52.28$). Although some variability was observed in the initial stages of intervention, this was common in multiple participants and Participant 3A was mostly on an increasing trend for the first four weeks of the study (see Figure 7 for isolated data for Participant 3A).
Baseline Group B

Participants in Group B generally provide LOPH at a higher rate in baseline than Group A, ranging from 0.77 to 47.83, with a median of 34.62 (M = 29.56). Participant 1B provided LOPH ranging from 11.39 to 14.97 (M = 13.82). Participant 2B presented LOPH ranging from 0.77 to 18.76 (M = 8.92). Participant 3B’s LOPH ranged from 26.71 to 47.81 (M = 39.33). Figure 8 depicts the baseline and intervention phases across consecutive sessions for Group B.

Intervention Group B

Group B received the resource and material-based PDC-HS indicated intervention while their supervisors also received an intervention comprising performance feedback and a job/aid checklist. In Group B, all participants generally increased provision of LOPH. Throughout the intervention phase, participants’ LOPH ranged from 11.38 to 93.83, with a median of 40.98 (M = 42.27). Like Group A, most participants in Group B had variable performances across days, but overall level of LOPH increased and remained above baseline levels, which is apparent in the weekly data (see Figure 9 for average weekly data for Group B participants).

Participant 1B

Following baseline measurement of LOPH (M = 13.82), Participant 1B demonstrated immediate change in LOPH for the first session that the intervention was delivered. Within both groups, only Participants 1B and 3A saw immediate changes in
level of LOPH following intervention. However, Participant 1B reverted closer to baseline levels until approximately one calendar week of sessions had occurred. This pattern was more akin to Participants 1A and 2B, as there was some delay to the effect of the intervention. In intervention, Participant 1B’s LOPH ranged from 19.08 to 62.57 with a median of 37.81 ($M = 37.82$). After this initial delay, LOPH stayed between approximately 2-3 times baseline averages (see Figure 10 for isolated data for Participant 1B).

Participant 2B

The effect of the intervention on Participant 2B’s provided LOPH compared to baseline levels ($M = 8.92$) was not apparent upon first introduction to the intervention. Like other participants (e.g., 1A and 2A), this lasted for approximately one week of sessions until an increase in level occurred. In intervention, Participant 2B ranged from 11.38 to 61.17 with a median of 44.83 ($M = 40.08$). After the initial delayed effect and following some variable daily session LOPH totals, Participant 2B’s data were more stable, with around 45-50 LOPH each session. This was approximately 5-6 times the baseline average of LOPH (see Figure 11 for isolated data for Participant 2B).

Participant 3B

Participant 3B had the highest LOPH in baseline ($M = 39.33$), but these totals remained relatively stable. Following intervention, Participant 3B demonstrated a significant increase in LOPH in the second and third sessions. However, these LOPH totals were not matched in any subsequent sessions. Moreover, daily LOPH data remained variable. For Participant 3B, provided LOPH ranged from 33.30 to 93.83, with a median of
60.32 \( (M = 60.25) \). This represented an overall approximate 50% increase from baseline averages (see Figure 12 for isolated data for Participant 3B).

Secondary Dependent Variable: Supervisor Integrity

For supervisor procedural integrity, the secondary dependent variable in the study, integrity varied initially between the two groups. Group A’s scored checklists were 17% lower on average compared to Group B. While both groups received an intervention wherein materials and organization were improved by the primary researcher (i.e., a score of 100% of checklist items), this disparity remained throughout the study. Through five weeks of the study, Group B, whose supervisors received bi-weekly feedback and checklists on material status, stayed, on average, 19% higher than Group A, whose supervisors did not receive a checklist or feedback (see Figure 13 for weekly average integrity data for both groups).

Between Subjects Comparison

A mixed-design ANOVA was conducted to examine the effects of the PDC-HS indicated interventions in changing LOPH over baseline levels across five weeks, as well as if the effect was different between Group A, which received a single intervention for behavior technicians, and Group B, which received the additional supervisor-level intervention, compared to an untreated control sample. Mauchly’s test indicated that the assumption of sphericity had not been violated for the main effects of time, \( \chi^2(9) = 8.63, p = .484 \). Therefore, degrees of freedom were not corrected. Results showed that the time, in
weeks, did not have significant main effect on change in LOPH over baseline, $F(4, 36) = 2.01, p = .113$ partial $\eta^2 = .18$ (see Figure 14).

The homogeneity of variance assumption was fulfilled for the five weeks sampled. Group membership did have a significant main effect on LOPH above baseline, $F(2, 9) = 7.99, p = .010$, partial $\eta^2 = .64$. Post-hoc tests with the Bonferroni correction demonstrated that participants’ change over baseline in LOPH was significantly higher for Group B ($M = 23.75$), who received the intervention for both technicians and supervisors, over the control group ($M = -1.78$). The difference between Group A ($M = 14.55$), who received the intervention for behavior technicians only, was not significantly higher than the control group nor was it significantly lower than Group B. There was no significant interaction effect between group membership and week.

Procedural Acceptability Data

All single-case participants in Group A completed procedural acceptability or social validity surveys, while 2 of 3 participants in Group B completed surveys. The abstaining participant left the company before completing the survey. Across all submissions, participants found the resource and process intervention derived from the PDC-HS to be helpful in providing learning opportunities for the clients, with most selecting Agree regarding the statements on LOPH and making their jobs easier. In addition, all participants rated the intervention highly regarding the benefit to the organization and the clients, with most agreeing that they would recommend a similar intervention for other key job duties (see Figure 15). While there was minimal additional
feedback provided, one participant noted that the organizers made it easier to keep smaller program stimuli organized, while another mentioned that their organizer had a manufacturing defect which was corrected during the intervention phase (i.e., Participant 2A).

Only one supervisor, from Group B, completed the procedural acceptability survey. Like the behavior technicians, they agreed that the PDC-HS indicated resource and material intervention was easier for behavior technicians, their job, and benefitted the client and organization. However, for the secondary intervention, they were neutral about the job/aid checklist and disagreed with the statement that it was helpful in maintaining case materials and program stimuli. In addition, they were also neutral regarding the statement that feedback helped them better maintain program stimuli and case materials and disagreed with the statement that they liked receiving feedback from their clinical director. The other supervisor from Group B anecdotally noted to their clinical director that they liked and appreciated the feedback but did not complete the survey.
Chapter 7
Discussion

The presumed intent of any behavioral intervention is that the observed change persists beyond formally arranged conditions (e.g., Foxx, 2013). In many studies, data on maintenance of the performance change and adoption of the change procedures remains secondary to producing an effect (Redmon, 1991). As this is rarely the primary focus of studies, follow-up measurement of maintenance is typically relegated to probe data or long-term post-hoc analysis (e.g., Boyce & Geller, 2001). However, behavior analysts view performance in context of its environmental controls. That is, performance sustains if there is some environmental variable to maintain it (e.g., Malott, 2001). This leaves a critical gap in studies in functional assessment, as function-based treatment reliably produces more robust results in empirical literature, but it remains largely unknown if function-specific intervention produces more durable performance over time. In addition, though suspected (e.g., Boyce & Roman, 2002), empirical validation is necessary in determining if diagnosing and correcting the appropriate variables also makes institutionalization of the change procedures more likely. This is extremely important in organizational settings as OBM practitioners are often tasked with creating environmental controls (i.e., contingencies, processes, and systems) rather than relying on naturally occurring ones.

Therefore, the purpose of the current study was two-fold. First, to observe the impact and change over time in a function-based performance intervention, specifically when apparent researcher control was absent. Second, to see if the same functional assessment tool could be used to craft a supportive intervention for in-house implementers.
to maintain the intervention (i.e., to achieve institutionalization). Clinical directors interviewed by the primary researcher both indicated that behavior technicians onsite generally lacked the necessary task materials to complete their job effectively. Moreover, there was process ensuring that these materials remained in good condition and organized.

An intervention was planned that included 1) increased access to materials, 2) organization solutions that permitted easier access and arrangement of these materials during and after session, and 3) supplemental stimuli (e.g., program cards and labels) to ensure that inventory was accurate to the protocols specified by supervisors within the online data collection platform. Groups A and B both received access to this intervention derived from the Resources, Materials, & Processes domain of the PDC-HS. The intervention, once implemented with fidelity, was immediately left to staff (i.e., behavior technicians and supervisors) to maintain over a five-week period to examine maintenance of the intervention. However, Group B also received a secondary intervention to ensure that supervisors had adequate supports to maintain materials. A subsequent round of PDC-HS interviews indicated that supervisors would benefit from feedback as well as a job aid or checklist to highlight key features of organization. This secondary intervention was designed to be delivered by clinical directors at the organization to encourage the institutionalization and maintenance of the primary intervention.

LOPH provided by behavior technicians were measured, taken from the permanent product when session data was uploaded to the organization’s case management platform. Of the six behavior technicians whose data was examined within a single-subject design, all improved over baseline levels after receiving the intervention, without any additional
modifications from the primary researcher for five consecutive weeks. There were two
general patterns that emerged after the intervention. For half of participants in the within-
subjects designs (i.e., 3 of 6) this was a variable LOPH count for approximately one
consecutive week of sessions (i.e., between 3 and 5), before more consistent responding
and increases in level. For the other half of participants, there was an initial increase in
level from the last baseline data point to intervention, but then multiple consecutive
sessions of minimal change until a larger change in level. These patterns cannot be ruled
out as an artifact of group membership, as the former pattern was somewhat more common
in Group A (i.e., 2 of 3) while the latter occurred in 2 of 3 of Group B participants.
However, it was Group B overall, whose supervisors also received bi-weekly feedback and
communication from their clinical director that tended to perform better across the duration
of data collection.

Participant 1A was the initial behavior technician recruited for the study as their
overall performance was the lowest. While examination of their isolated data (see Figure 5)
shows significant increases over their initial performance, they were the smallest change
overall compared to other participants both within and across groups (cf. Figure 3). These
changes are difficult to evaluate visually when their y-axis is scaled according to the other
group members’ performances, though it should be noted that 88% of data in intervention
does not overlap with baseline performance which meets general criteria to deem the
intervention “effective” (Carlin & Costello, 2022). There are multiple idiosyncratic
variables that potentially impaired this change. First, in addition to the global
organizational problem of poor resource provision and organization, this performer may
have had other motivational variables impeding performance. Based on supervisor report, they were regarded as one of the lower performing behavior technicians overall. This was anecdotally verified by reported low engagement with other clients and being observed taking personal calls multiple times within client sessions. Furthermore, these variables may have led to a less effective supervisory relationship. For instance, the supervisor may have been more reluctant to provide corrective feedback and support given these perceived performance deficits. Specific to this case, the behavior technician also complained that the supervisor was not providing enough programming for the client (i.e., targets were mastered more quickly than they were added) and did not update stimuli when additional targets were added. For example, their supervisor had advised them to look up stimuli for tact targets on their cell phone. Given the timing of these comments, this may explain the increasing trend between data points 9 and 13 and then the subsequent decreasing trend between 13 and 19. That is, the intervention in concert with the client’s progress may have served to produce additional LOPH until the client reached perceived or actual mastery criterion, with LOPH declining until new instructional goals could be added and/or corresponding materials were produced.

Participant 2A initially did not produce many trials with their client (i.e., sessions 1-3) but displayed a relatively stable pattern through the last four data points of baseline. However, the last data point in baseline was slightly higher the previous two sessions (i.e., approximately 2 more LOPH). Given the minimal initial level change following intervention, this represents a methodological limitation in these data. There was a generally increasing trend in the first eight intervention sessions. This changed to a pattern
wherein the increases in LOPH and variability were greater, including a decreasing trend through sessions 19 to 22. Like Participant 1A, this may have been related to delays to the supervisor adding additional targets once mastery criterion was achieved, given that supervisory sessions often occurred once weekly, while updates seemingly occurred more commonly bi-weekly. Regardless, final LOPH were approximately 4-5 times baseline rates. Notably, Participant 2A rated the intervention favorably overall, but, early in the intervention (i.e., during the initial increasing trend) had complained to their supervisor that some flashcards were difficult to access from the storage solution. Their supervisor noted that a manufacturing defect with their organizer resulted in a gap between dividers and the container base, resulting in cards sliding between sections. This was corrected by the supervisor and researcher in the second intervention week by placing pieces of tape where the divider met the base so stimuli could not pass beneath the barriers.

Despite being the only behavior technician without the RBT® credential, Participant 3A had the highest baseline performance of any participant in Group A but was noticeably declining in LOPH. After intervention, there was a clear increase in trend, but their data remained highly variable across days. There was a decreasing trend between sessions 20 and 22, but there was an immediate rebound in the subsequent three sessions, before ending on a significant drop in level in session 25. Like other participants in Group A, it is likely that intervention functioned to increase the LOPH overall, but that case management factors (e.g., programming and target updates) led to variability within daily performance. As the two supervisors comprising Group A did not receive any intervention, this might explain why the effects were diminished (e.g., Participant 1A) or less consistent.
Participant 1B initially increased their LOPH in the first session after intervention, but then had consecutive sessions with close to baseline levels. Afterwards, performance sharply increased and remained somewhat consistent (i.e., sessions 9 through 18). In the latter portion of the intervention phase, performance began to vary, with a decreasing trend in sessions 19 through 22, and then an increasing, albeit variable, trend for the remaining sessions. The observed consistency in the middle phase could, potentially, be ascribed to the supervisor receiving regular information about the state of materials and resources, considering similar patterns were also observed in Participant 2B. In addition, later variability could also be attributed to idiosyncratic factors. Participant 1B reported personal, unspecified, health issues which disrupted sessions and may have resulted in inconsistent performance, though similar patterns of variability were not uncommon across groups.

Participant 2B initially had a low rate of LOPH, with a gap in baseline data due to a technical issue where session data were lost. However, the final three data points of baseline indicated a decreasing trend. Following intervention introduction there was some latency to a clear impact on LOPH. This was not unusual for participants, but specific to Participant 2B, this was partially due to a unique factor with their supervisor’s programming. Their LOPH would not be recorded unless they reached a minimum count per session, meaning that it was likely that there was an increasing trend not reflected in the data as Participant 2B approached the minimum LOPH for the electronic platform to count and record targets. Once they reached this level, their performance varied. There was a decreasing trend for approximately one week of sessions, followed by an increasing trend
before stabilizing. Overall, this tends to be consistent with the other Group B participants, in that the increase was greater with intervention but the level remained more consistent, compared to the continually increasing but more variable data of Group A.

Participant 3B was somewhat unique amongst participants in that they were generally higher in performance than other participants and were, anecdotally, regarded as a much stronger technician than other participants. They were amongst the most experienced behavior technicians and were the only within subject participant pursuing BACB® certification. However, with this client, their trials were low compared to other behavior technicians on the case and given the number of discrete trial training (DTT) targets present in this client’s programs. After the first session, their LOPH increased greatly, more than doubling their baseline average. Their pattern following was generally lower, with variable performance at one point dropping below baseline levels. This can partly be ascribed to unforeseen scheduling changes, as administrative staff shifted the client’s daily session from having one behavior technician to a split between two technicians. Lower LOPH were generally observed on their shorter sessions, as Participant 3B was scheduled with the client near the end of the day. Moreover, Participant 3B received the intervention at the site where the clinical director refused to deliver performance feedback to the BCBA® supervisor responsible for the case. This resulted in delays to the supervisor receiving the feedback and, unlike the previous two participants, was delivered by the primary researcher. These factors likely contributed to the inconsistent performance post-intervention.
The dependent variable for the supervisors, the maintenance of the intervention resources and process, was measured as procedural integrity regarding the organization, quality, and availability of materials. Baseline levels of integrity based on the checklist created for the study show that, despite random selection of supervisors, there was an initial disparity in the overall availability and organization of the materials between groups. However, this variance can be mostly attributed to Participant 2B, who had the highest pre-intervention fidelity score at 63%, which was at least 25% higher than participants in both groups. Regardless of initial differences, all client materials were organized to attain a 100% score on the checklist/job aid and scores declined most steeply in Group A, which did not have any supervisory intervention. While there are many potential caveats to these findings, the relatively greater change and steadiness of Group B is suggestive that multi-tiered intervention might be important in producing longer lasting change and maintenance of interventions.

One significant caveat is that, although the present study was designed to limit formal researcher involvement, it nonetheless remained a research study in an applied environment. However unobtrusive measurement and monitoring was implemented, participants were aware that there was an intervention (i.e., they consented to participation) and there were various shifts in ancillary organizational processes throughout. For example, there were multiple scheduling and staffing changes throughout the study that necessitated change in participants and session lengths (e.g., Participant 3B). Many more ideal baselines (i.e., lower and greater declines in LOPH) were not incorporated into the study as these staff left the organization during data collection and/or were re-assigned to
different cases. In addition, the supervisory population diminished greatly from the onset of the study. For example, three supervisors left during baseline data collection, while two additional supervisors had their in-center cases redistributed to other analysts after the first participants were introduced to the intervention.

Furthermore, one supervisor chose to not participate in the study and only one clinical director participated in the study. The latter was a significant barrier to promoting institutionalization, as their choice to not provide feedback coincided with them also reassigning the case near the end of the intervention phase for Participant 3B. Thus, to preserve the integrity of the intervention, rather than its intent, the primary researcher measured and delivered feedback to Participant 3B’s supervisor. Another curious finding regarding supervision relates to the rate of LOPH during sessions in which supervisors were present. Supervisory presence had an idiosyncratic effect amongst participants, in that performance varied in both directions when supervisors were in session. These data were not included in the analysis, given that examined relation was intended to be the impact of the intervention on behavior technician performance in independent sessions. However, there were multiple supervisors in both groups in which supervision sessions resulted in below baseline levels of LOPH, suggesting that oversight, and perhaps embedded training, feedback, and other monitoring procedures, might have restrictive impacts on treatment delivery. Further research ought to investigate this relation further, given that the effect of observers might, in some cases, be utilized as a positive influence on performance (e.g., King et al., 2018).
While the issues with participants and institutionalization limit the group level comparison as well, the results do have some significant implications. The expectation was that initially, function-based intervention would be equally impactful across both groups, with the multi-level intervention of Group B eventually proving more durable over time. This was borne out somewhat in that the difference between Group A and Group B was not statistically significant. That is, the function-based treatment produced not overly dissimilar levels of change, which can also be confirmed in visual inspection of the single subject data. However, these differences were great enough for Group B differences to be statistically significant compared to the untreated Control group, whose average change in LOPH over the baseline week was generally declining slightly or static across weeks.

Interestingly, the interaction effect was non-significant as well, which did not validate the expectation that Group B would prove stronger over time. It should be noted, that, though non-significant, the average change in LOPH over baseline was higher than Group A consistently, save for the third week. So, despite Group A having more room for improvement, by virtue of lower average baseline measurements in LOPH, the multiple component intervention produced higher results in nearly every week. Notably, the disparity was most pronounced in the 2nd, 4th, and 5th week. These differences happen to coincide with the lowest integrity scores for Group A as well as the feedback timings (i.e., every two weeks) for Group B. Whereas the increases in LOPH were relatively linear for Group A, there were greater upward shifts in Group B participants, suggesting that supervisory involvement might produce more robust initial change. This could be a critical feature in producing long term adoption of procedures, as Boyce and Roman (2002) note.
that demonstrated value to organizational decision makers could make them more likely to
maintain the intervention once formal research or consultancy has ceased.

The observed relationship between performance and the secondary dependent
variable, procedural integrity of the intervention, is not unexpected, given research that has
examined the clinical effects of poor treatment integrity. Varying dosages of treatment
accuracy or errors have been examined in a variety of procedures targeting learner
compliance (e.g., Leon et al., 2014), differential reinforcement of alternative behavior (e.g.,
St. Peter-Pipkin et al., 2010) and pediatric feeding concerns (Ulloa et al., 2020). While
most of studies in procedural integrity have been in relation to consumer behavior and
acquisition rates, there have been some common patterns observed. Brand et al. (2019)
synthesized the findings of these and other studies that examined parametric manipulations
of procedural integrity, finding that consumers overwhelmingly performed best under
100% integrity conditions (i.e., when a procedure was implemented exactly as designed),
with increased variability and delays to acquisition when integrity was lowered. Moreover,
Brand et al. observed that multiple studies suggest that exposure to 100% integrity initially
in intervention may inoculate learners to later degradation in procedural integrity.

Brand et al. (2023) conducted a further study examining parametric manipulations
of procedural integrity and their impact on student performance on a computerized match-
to-sample program. Their results suggested that counter to earlier literature (e.g., Brand et
al., 2020), exposure to lowered integrity (i.e., errors) following mastery of a skill
inconsistently disrupted performance. Many participants still performed accurately on the
task at 80% integrity levels, and their data overall suggested that a critical threshold for
maintaining a skill, in their experimental paradigm, might lie between 60 and 80 percent. In the present study, the observed outcome of varying procedural integrity was not consumer behavior, rather it was staff behavior. Yet, there were similar patterns observed (i.e., lower integrity resulting in lower performance), which could suggest that integrity might impact staff behavior in ways akin to consumer learning. However, this statement should be approached with caution, as this there are multiple design elements that do not necessarily translate directly to this literature. For example, Brand et al. (2020) noted that most of these parametric modifications were made regarding consequence delivery and prompts, and errors were, in multiple studies, split into those of omission (i.e., missed procedural steps) and commission (i.e., extraneous steps in the procedure).

The present intervention was neither consequence nor prompt based, and the measure of integrity was based solely around the organization, availability, and quality of resources provided. In addition, the “errors” in integrity could generally be classified as those of omission (i.e., the materials were not easily accessible or available for learning opportunities). Errors of commission would be harder to classify in this sense. For example, Participant 3B’s supervisor received feedback that, while all stimuli for written programs were present, there were multiple sets of extraneous stimuli in the client’s storage bin. This could be conceptualized as having unnecessary resources (i.e., an error of commission) added to the intervention, but sometimes these stimuli were used programmatically, to intersperse previously mastered targets or to add distractors in an array. However, there were so many that it made current materials difficult to find and rendered them, relatively, unavailable (i.e., producing an error of omission). Furthermore,
participants were arguably introduced to the intervention at 100% integrity levels, yet performance was mostly variable across participants and some small increases were still observed even when the checklist observation was scored lower than 50% (e.g., Participant 1A).

A counter argument could be that, in baseline, they were already introduced to materials and resources far below “perfect” implementation which would explain the unstable subsequent performance, despite the initial intervention integrity. As the primary intention of the study was to model the change in intervention integrity and performance over time, participants were also not formally presented an intervention with perfect integrity for any extended period. That is, researchers did not alter the integrity of the intervention once it was introduced or maintain high integrity until stable responding was achieved. This could have produced patterns more in line with other parametric procedural integrity literature, as achievement of mastery under perfect conditions might stabilize the performance over time. In addition, as Brand et al. (2019) note, measurement of integrity is generally idiosyncratic due to the number of different procedures, populations, and skills targeted; this study is no different in that regard. Contextual variables likely account for some of the variability in performance, such as client target behavior and environmental changes. Therefore, further research might investigate more formally how procedural integrity impacts staff behavior at work while accounting for these variables.

Given that both intervention conditions resulted in improvement across participant groups, the present study also generally adds to the literature investigating the predictive validity of the PDC-HS (e.g., Wilder et al., 2020). In addition, there are some more specific
additions to the PDC-HS literature. For example, interventions when the Resources, Materials, & Processes domain is indicated have been less frequently investigated (e.g., Merritt et al., 2019; Wilder et al., 2018) and, in analogue studies, tends to be the domain in terms of participant scoring accuracy with contrived vignettes (e.g., Cymbal et al., 2020; Wilder et al., 2018). Therefore, the relative impact of these interventions demonstrates further utility of this domain and demonstrates the reliability of the tool, given the relative agreement between clinical directors at two levels of performance (i.e., clinical supervisors and behavior technicians). This could also potentially be due to the primary researcher’s experience administering and utilizing the tool, which seems to moderate its accuracy in analogue studies (e.g., Wilder et al., 2018). Another potential contribution is in the participant population and performance problem examined. Typically, the PDC-HS has been used by supervisors and managers to analyze performance problems of consumer-facing roles (e.g., direct-care staff) rather than examining variables impacting supervisory performance. While this was not the primary dependent variable in the study, the stronger intervention effects in Group B suggest that the tool may have some utility in treating performances beyond front-line roles and may inform efforts toward maintenance.

Relatedly, the primary goal of the study was to examine how an intervention might maintain over time once introduced. One critique of traditional approaches to examining response maintenance are attempts to distill performance changes to basic operant conditioning principles and behavioral phenomena, such as schedules of reinforcement, extinction, and stimulus generalization (e.g., Malott, 2001; McSween & Matthews, 2001). Such explanations may fit better when changing clinically significant or aberrant behavior,
when the goal is to craft repertoires that fit in and align with pre-existing verbal communities and social norms in populations with limited verbal repertoires. In this sense, McSween and Matthews point out that basic principles can be a critical facet when the primary target is transfer of training effects. However, in changing behavior in organizations or, more broadly within society, the focus is more typically on correcting deficient systems and processes (e.g., Malott, 2003). That is, creating an environment conducive to more productive, prosocial, and/or effective performances, often when these systems and component contingencies do not already exist. The general complexity of organizations and work performance means that it would be beneficial to examine not only the variables that produce change (e.g., examples of predictive validity with the PDC-HS), but also if these variables maintain change or remain the same as when initially introduced in the intervention. Despite appeals to improbable or indiscriminable contingencies and resistance to extinction (e.g., Boyce & Geller, 2001), even absent apparent intervention contingencies, some environmental variable likely controls workplace performance (Malott, 2001). Many studies have indicated (e.g., Alavosius & Sulzer-Azaroff, 1990; Boyce & Geller, 2001; Myers et al, 2010) that verbally mediated contingencies may often maintain long term behavior change even when formal systems have not been adopted past intervention. Although this study was not, compared to others, long enough to evaluate the development of some these extant contingencies, experimentally examining assessment procedures (e.g., the PDC-HS) as a predictor of sustained performance change is a significant first step.
Often, many post-hoc evaluations of sustained organizational performance change have noted that pre-intervention assessment seems to produce stronger, sustainable results. In other words, when consultants and/or researchers use diagnostic tools to pinpoint the performance issue and find solutions that are environmentally appropriate, the intervention tends to be more successful. Many times, these assessments take the form of BSA (e.g., Grindle et al., 2000; Sulzer-Azaroff et al., 1990). In addition to identifying functional issues, data can also provide information on any potential barriers and permit an opportunity to convene and collaborate with organizational leaders and stakeholders. This process is important to solicit buy-in and may also identify the natural contingencies (i.e., consumer outcomes, production) that can help bolster change efforts (e.g., Alavosius & Sulzer-Azaroff, 1990, Boyce & Geller, 2001; Hagge et al., 2017; Myers et al., 2010; Reid et al., 2015). Moreover, a performance that eventually produces natural, beneficial outcomes also promotes sustained change. Because the PDC-HS is designed in such fashion to examine individual performance and not the entirety of the organizational milieu, it likely did not capture all the necessary components to produce the most robust, enduring effects. However, in contrast, more holistic methods are so large and dense in components they are hard to examine empirically. Thus, given the empirical support for the PDC-HS, the present study was designed around the tool, to see if even a performer level assessment could be sufficient to involve stakeholders and identify functional variables in such a way that the performance sustained. In addition, it was seen if using the tool at multiple levels (i.e., to assess sources of supervisory performance deficits) could add to the impact of the intervention, thus approximating a step toward a more BSA-like intervention.
Compared to the taxonomies established previously, such as the “temporal generality” tactics compiled by Conard et al. (2016), utilizing functional assessment provides numerous advantages. First, assessment of function is a critical feature of behavior analysis generally, and the introduction of function-based intervention as a gold standard has reduced the need for probabilistic tactics or imposition of potent reinforcers and punishers (e.g., Mace, 1994). In addition, there is the suggestion that such generality tactics may be additive (i.e., that more would produce better outcomes), something that has yet to be examined fully with components of the Conard et al. taxonomy. Those tactics that have been empirically examined (e.g., Sigurdsson & Austin, 2006) have limited conclusive support. More broadly, the composition of these tactics, in some ways, is nonsensical. For example, BSA, a tactic for generality identified by Conard et al., uses procedures that subsume multiple other tactics, such as those derived from the Sigurdsson and Austin (2006) institutionalization taxonomy (e.g., involving staff in design) as well as incorporation of instructional design factors. Instructional design factors, in turn, often include procedures that subsume many of the tactics derived from the Stokes & Baer (1977) taxonomy to promote transfer of training effects, all of which are present as independent items in the temporal generality taxonomy. Likewise, although on a smaller scale, the use of tools like the PDC-HS could, by default, utilize some of these tactics, in assessing already present natural contingencies or involving workers in the design of the intervention. More broadly, functional assessment derives its potency from predictive validity, in that it can identify and guide the correct intervention(s) for the targeted environment. Such evaluations are much more amenable to experimental evaluation,
evidenced by the comparative lack of experimental evaluation regarding the taxonomical approach.

For the current study, it seems that the PDC-HS was useful in identifying an intervention that both changed immediate performance and sustained changes above baseline levels without much additional or obvious support from the primary researcher. While five weeks of examining the intervention seems short compared to the duration of measured follow-ups in JOBM (e.g., VanStelle et al., 2012), this was the longest amount of time that the relationship between the supervisor, intervention, and performer could be sustained for all participants in the present environment. In other words, this was the extent to which the behavers remained the same and the functional relationships could be preserved intact, prior to turnover and other changes in client assignments. So, in this sense, the evaluation remains somewhat unique. Moreover, the observation of performance change following a researcher led intervention is not always measured continuously across time, and the differences in participant performance seem to suggest that further layers of intervention might be necessary to produce the strongest change, given the generally higher performance of Group B.

Maintenance, as a term in applied literature, remains somewhat murky in definition. While the term in this study proposed a definition that maintenance is simply the removal of some experimental contrivance and sustained performance, there are multiple potential ways to analyze durability. In the present experiment, maintenance was evaluated as the length of the functional intervention and subsequent changes in participant performance. It may be beneficial to further differentiate the term based on what is being
measured, rather than create and adopt broader terms (e.g., temporal generality). For example, in longer follow up studies, the original functional intervention may no longer be present at all (e.g., Boyce & Geller, 2001). Instances like these might suggest the measure of maintenance is more akin to social validity measurement (e.g., measuring if the performance change was important to consumers and stakeholders), because the performance persisted despite the intervention’s absence. Such investigations might yield more resource lean alternatives to formal intervention when there are potential potent natural or social contingencies present. Alternatively, studies might detail maintenance of the intervention as a cultural development (i.e., the practices were transmitted to subsequent generations of performers). For example, Hagge et al. (2017) noted even a fraction of employees participating in a BBS program over time resulted in lower rates of injuries overall across multiple years. Therefore, it may be beneficial for researchers to plan examinations of maintenance accordingly, with methods and designs that fit the type of durability being assessed. Those researchers looking to examine maintenance through creation of cultural practices might utilize the metacontingency paradigm, while those looking to evaluate the usefulness of the performance targeted might pursue more qualitative research. Researchers looking to extend experiments such as the present study, modelling the effects of a functional intervention over time, might consider utilizing more participants in larger scale group designs, to diminish the impact of attrition or organizational policy shifts on measurement. Another potential avenue is extending investigation of institutionalization.
The maintenance or adoption of the intervention, called “institutionalization” by Sigurdsson and Austin (2006) was another focus of the present study. Institutionalization has oft been suggested as the true goal of organizational intervention (e.g., McSween & Matthews, 2001), in that the system should subsume the intervention to ensure that important performances continue to occur reliably and not due to chance. It should be noted that Group B specifically incorporated all tactics of the institutionalization taxonomy proposed by Sigurdsson and Austin. The PDC-HS was used to involve workers and managers in developing the intervention and the clinical director was trained and monitored on implementing the intervention for supervisors. Moreover, all directors were made aware of how to formally monitor these data, using the electronic data collection platform and they had a system to formally dispense consequences to their subordinates.

Yet, in many respects, it could be said that institutionalization failed, as the primary researcher could never fully fade support for the supervisory intervention. In the case of one clinical director, they never implemented the intervention for supervisors, even with assistance and training offered. Thus, the present data suggests that the Sigurdsson and Austin taxonomy might oversimplify the nature of these tactics. In other words, asserting the presence (or absence) of a tactic may be too simplistic and might be better represented as a gradient. For instance, while stakeholders were generally involved in the design, involvement was not equitable. Supervisors and behavior technicians had far less choice in design than other levels of leadership due to the experimental design. One supervisor reported negative perception of the feedback, which may have been remedied by more involvement in planning but also would have confounded measurement. In addition, formal monitoring systems, while present, were not utilized reliably by leadership and supervisors.
based on anecdotal reports and individuals often had idiosyncratic approaches to monitoring supervisees. Thus, a stronger, cohesive system likely would have been better over the mere presence of one.

Moreover, the adoption of the tools and organizers used for the independent variable did not yet translate to organization-wide use, though the remaining specific participants involved still utilize the tools provided. However, this represents roughly twelve percent of all consumers served by the agency. So, given that the intervention has not yet become pervasive organization-wide and as Gravina and Austin (2018) noted, the institutionalization taxonomy likely does not also encompass all the various facets that might produce intervention adoption, such as support from leadership. Even then, this intervention was initially supported by the agency owner, but other variables such as cost, might be a prohibitive barrier. Each organizer, materials for stimuli and program cards for one client cost approximately 20 USD. While this was insignificant for one client, even given that this was a smaller agency, it would cost roughly 1000 USD for the necessary materials for all clients. Moreover, these solutions were only ideal for clients that had primarily language-based instruction; more bespoke solutions (e.g., if materials and reinforcers were larger, or if problem behavior were more severe and required protective equipment) could easily increase the resource cost greatly.

Some procedural acceptability reports also may hint at why these procedures were not quickly adopted. As mentioned above, one supervisor in Group B reported the feedback from their clinical director as a negative feature and were ambivalent towards the checklist/job aid, despite unknowingly having some of the highest performing behavior
technicians and procedural integrity scores across groups. A follow-up conversation with that supervisor indicated that they thought the combination of checklist and feedback was excessive and would have preferred longer intervals between feedback. In addition, the participating clinical director noted that they liked the feedback and found the intervention easy to implement and asked for the tools and extra organizing materials purchased for the study. However, they have also continued to utilize their preferred method of monitoring, which is delivering feedback verbally during supervision to behavior technicians on their specific case teams, rather than to supervisors they monitor within the clinic. That is, they had a pre-existing, albeit ineffective, process that they continued to use in lieu of the intervention. In this sense, this may be a facet not easily examined with tools such as the PDC-HS, and more readily investigated with BSA, as process creation and alignment is generally beyond the scope of the assessment.

So, in sum, the participating behavior technicians, despite some barriers to full adoption of the full intervention (i.e., Group B) or in the absence of the supervisory intervention (i.e., Group A) all improved their performances with a PDC-HS indicated intervention. The supervisors who received feedback on the state of the intervention provided mixed reports on the utility of the feedback and checklist/job aid, though performance was generally better in their group. In this case, like literature on parametric manipulation of integrity (e.g., Brand et al., 2023), higher integrity seems to promote stronger performance. One potential mechanism explaining the increase regardless of group is straightforward: clients that received more LOPH may have learned more skills or been more engaged in session, resulting in positive reinforcement contingencies that
produced continued performance improvement over time, relative to those that received no intervention at all. In addition, given that most behavior technicians surveyed noted that they thought the intervention was favorable to their sessions and their clients, it is possible that the presence of materials and more successful sessions resulted in rule generation that supported the performance change. These potential verbal relations remain an area to investigate further in OBM research.

Conversely, variability or smaller than desired changes in participant performance could likely be attributed to the nature of intervening on one specific interlocking contingency within the organization. As Rummler and Brache (1990) note, optimization of one component in a system can destabilize other parts of the system. Having behavior technicians suddenly increase their rate of LOPH might put more pressure on clinical supervisors to change and/or update programs and protocols, something they may be ill-equipped for if they already have a large caseload or an unoptimized schedule. Most clinical supervisors at the company were hired as independent contractors, so increased demands for case management duties, which were unpaid outside of client facing sessions, would likely not be completed quickly, or done so reluctantly. Alternatively, more LOPH might have resulted in more challenging client behavior with increased demands, resulting in some days where rates were lower. Regardless of the mechanism, the present study provides some preliminary evidence that OBM functional assessment tools, such as the PDC-HS can produce reasonable change even without stringent controls or implementers.
Limitations and Future Research

There are, however, numerous limitations. First, identification of an issue within the Resources, Materials, & Processes domain of the PDC-HS was fortuitous regarding the experiment’s main conceit (i.e., measuring maintenance), as the entirety of the intervention could be introduced almost immediately to the behavior technicians and its upkeep did not require regular researcher input. Alternatively, if the issue identified necessitated a regular consequence (e.g., performance feedback), the primary researcher may have had to deliver feedback until supervisors could be trained and proficient at delivery. However, given the sequencing effects noted by Brand et al. (2020) when examining integrity failures, it may be beneficial in subsequent studies to expose the participants to multiple sessions of a researcher-maintained intervention, to observe the impact on the longevity of change.

Second, literature on the PDC-HS has grown recently, leading to empirically based refinements to intervention selection (e.g., Vance et al., 2022) as well as an updated version of the tool itself (i.e., the PDC-HS 1.1; Brand et al., 2022). It is possible that use of either of these refinements could have improved the quality of the intervention chosen or perhaps changed the results of the domain selected for intervention. However, using the original tool was a conscious choice as most of the empirical support in terms of applied utility, validity, and reliability was conducted using the original Carr et al. (2016) version of the tool. Thus, because the purpose of the study was to examine durability of interventions, given the lack of subsequent empirical evaluation of these modifications, it was deemed best to use the original tool. This would avoid simultaneously evaluating variations to the tool and the intervention selection component while simultaneously
answering the primary experimental question. However, future research needs to examine the utility of these tools and refinements in this area as well.

Third, the PDC-HS indicated interventions are often compared to a non-indicated intervention (e.g., Bowe & Sellers, 2018; Carr et al., 2013; Ditzian et al., 2015). In this study, there was no formal non-indicated intervention implemented, as there were not enough participating supervisors, behavior technicians, and/or adequate baselines to create additional groups for study. However, it should be noted that the control group informally received non-indicated interventions throughout the study, as the clinical directors often delivered feedback to technicians regarding LOPH, set goals that were not met for LOPH, and mentioned low rates of LOPH at a center wide meeting while participants were in baseline measurement. Feedback and goal setting would likely fall under the Performance Consequences, Effort, & Competition domain, while the reminder at the meeting would seemingly fit in the Task Clarification & Prompting domain. Neither domain was indicated by the PDC-HS and no change was noticeable in the Control group.

A fourth limitation is the relatively short duration of intervention evaluation. While this is not problematic for the study’s purpose and definition of maintenance, it is undeniable that long term evaluation represents a common measure of durability in applied research. Ideally, the study would have been able to provide both short- and longer-term evaluation as originally planned. However, as maintaining the experimental design and collecting further data in the future is unlikely, given the dwindling participant pool, this limitation would likely have to be addressed in subsequent studies. Relatedly, measurement duration also made it difficult to ascertain precisely how much the intervention was
adopted by the organization. While it remains in place for any remaining behavior
technicians/client teams that participated, it remains unknown what, if any, components of
the intervention might be maintained long-term by leadership and clinical supervisors. This
could be addressed in subsequent studies by arranging for longer-term measurement by
utilizing different or additional experimental designs.

Another limitation is the relatively small sample, namely for the supervisors.
Because there were fewer available supervisors with ideal behavior technician
performances, this means that the conclusions generated from the group level remain
limited. Despite attempted random assignment, there is some chance that the differences
between groups were related to supervisory skill and quality of programming. Such
concerns have limited the findings of previous PDC-HS studies (e.g., the quality of
different behavior plans; Collier-Meek et al, 2021), although experience levels were
relatively uniform (i.e., 2-4 years) between supervisory groups and certification level, in
fact, favored Group A, as Group B contained the only participating BCaBA® supervisor.
Likewise, the assortment of technicians was somewhat random, but tied to supervisory
group assignment. There is potential that there were individual differences that contributed
to the varying magnitude of change (e.g., Participant 1A) and conclusions may be tenuous
in cases where intervention did not produce large, immediate effects relative to baseline
level and trend (e.g., Participant 1B). Conversely, one of the greatest changes were
observed with Participant 2B and Participant 3A, both of whom were amongst the least
experienced in either group or were relatively new to the organization. It is possible that
less exposure to poorly maintained materials resulted in more stable performance even
when the intervention integrity degraded, given the observed sequence effects in most procedural integrity literature (e.g., Brand et al., 2019). Future studies should use a more robust sample to confirm the present findings.

An additional limitation is the type of performance chosen. While LOPH was troublingly low in baseline for many participants, indicating that there were no data collected for entire sessions (e.g., Participant 2A) or that one or fewer learning opportunities were provided per session hour (e.g., Participant 1A), it is difficult to evaluate the relative worth of the intervention. The provision of learning opportunities, trials, or teaching does not have a fixed industry standard, as ABA treatments should be tailored to the specific client, idiosyncratic environmental and contextual variables, and are subject to the preferences of the supervisor. In addition, there was no further evaluation of client outcomes related to the recorded LOPH, so there is scant direct evidence that the behavior change was valuable to organizational consumers. Even if these client data were to be examined (i.e., outcomes or learning targets achieved), it would be difficult to ascribe any changes in rate of skills mastered to the intervention while controlling for the impact of other variables (e.g., teaching procedures used, types of skills taught, challenging behavior evoked through additional demands, etc.). As such, there is no way to assert that the performance change was a net positive for the organization. Subsequent studies should likely choose a pinpoint with clear value to the organization or its consumers, as this might be a critical factor in whether organizational leaders adopt the intervention in the long term. Alternatively, future research could also examine the same variable, LOPH, and further measure how this relates to client outcomes over time.
Finally, the study remains a very preliminary foray into using functional assessment to investigate durability of performance change and adoption of intervention procedures. Thus, there are likely many other ways to enhance the complexity of the experiment (e.g., incorporating further layers of functional assessment and intervention) as well as examining other variables that might produce sustained results. While previous generality tactics (e.g., those outlined in Conard et al., 2016) do not provide compelling explanation of maintenance and institutionalization, nor do they entirely encapsulate the sum of potential variables responsible for lasting change, there are presently no other tools that accomplish this either. The PDC-HS is a highly useful and empirically validated functional assessment, but, like any other procedure, can always be further refined, validated, and iterated (e.g., Brand et al., 2022). Thus, there are myriad ways to expand upon this framework. Future research should investigate specific variables suspected to be important to adoption of procedures (e.g., leader support; Gravina & Austin, 2018) or test tools designed specifically around the sustainability of interventions (e.g., McGee & Crowley-Koch, 2021).

Much of this paper has centered discussion around generality, the dimension of ABA most closely related to maintenance of performance change and adoption of behavior analytic procedures. Many presumably choose to study and research behavior analysis with this in mind, hoping to apply those principles to the betterment of society at large. And, like lunar travel, we might choose to do these things “not because they are easy, but because they are hard, because that goal will serve to organize and measure the best of our energies and skills…” (Kennedy, 1962/2021). And, for much of the existence of behavior
analysis that skill and energy has been poured into those hard demonstrations of rigor and experimental control. However, though we now may know how to get to the moon, the next great challenge might be getting the general populace into space suits. In this sense, this may involve not choosing to do something hard, but rather transforming something hard into something easy. In other words, precise procedures and perfect integrity will likely always produce the most robust results. But, for dissemination and common use, research needs to also investigate what features encourage persistent performance changes, as well as what might drive consumer preference toward broad adoption of procedures in the face of ubiquitous systematic barriers (e.g., costs, education) and larger societal ills (e.g., pandemics, climate change, inflation). Therefore, while this is an extremely preliminary study, hopefully these data encourage adoption and further methodological refinements of the techniques herein.
References


https://doi.org/10.1007/s10864-018-09311-8

https://dx.doi.org/10.1080/01608061.2020.1715319


https://doi.org/10.1007/s40614-021-00317-8


http://dx.doi.org/10.1080/01608060902874526


https://doi-org/10.1177/014544557714001


http://dx.doi.org/10.1016/j.rasd.2015.06.012

https://www.americanrhetoric.com/speeches/jfkriceuniversity.htm (Original work published 1962)

http://dx.doi.org/10.1300/J075v21n01_06


164


http://dx.doi.org/10.1300/J075v21n01_08

http://dx.doi.org/10.1080/01608061.2016.1152209

http://dx.doi.org/10.5210/bsi.v13i2.20

http://dx.doi.org/10.1300/J075v18n04_06


https://doi-org/10.1901/jaba.2010.43-47

http://dx.doi.org/10.1080/01608061.2021.1920543

http://dx.doi.org/10.1901/jaba.1991.24-441

http://dx.doi.org/10.1007/s40617-015-0101-0


174

Table 1.

*Mixed Design ANOVA Results of Post-Baseline Week and Intervention Group on Learning Opportunities per Hour*

<table>
<thead>
<tr>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Subjects</td>
<td>10456.16</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>6689.08</td>
<td>2</td>
<td>3355.54</td>
<td>7.99*</td>
</tr>
<tr>
<td>Error</td>
<td>3767.08</td>
<td>9</td>
<td>418.57</td>
<td></td>
</tr>
<tr>
<td>Within Subjects</td>
<td>5476.95</td>
<td>48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week</td>
<td>902.48</td>
<td>4</td>
<td>225.62</td>
<td>2.01</td>
</tr>
<tr>
<td>Week*Group</td>
<td>541.02</td>
<td>8</td>
<td>67.63</td>
<td>0.60</td>
</tr>
<tr>
<td>Error</td>
<td>4033.45</td>
<td>36</td>
<td>112.04</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15933.11</td>
<td>59</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1: Bar graphs depicting the results of the PDC-HS, administered to clinical directors, regarding variables impeding frequency of learning opportunities provided and recorded during session by behavior technicians.
Figure 2: Bar graphs depicting the results of the PDC-HS, administered to clinical directors, regarding variables impeding supervisors from ensuring that behavior technicians had adequate materials to provide frequent learning opportunities.
Figure 3: *Graph depicting the performance of Group A participants providing learning opportunities in consecutive daily sessions when provided the resource and material intervention.*
Figure 4: Graph depicting the weekly average performance of Group A participants providing learning opportunities when provided the resource and material intervention.
Figure 5: Graph depicting the isolated performance of Participant 1 in Group A providing learning opportunities in consecutive daily sessions when provided the resource and material intervention.
Figure 6: Graph depicting the isolated performance of Participant 2 in Group A providing learning opportunities in consecutive daily sessions when provided the resource and material intervention.
Figure 7: Graph depicting the isolated performance of Participant 3 in Group A providing learning opportunities in consecutive daily sessions when provided the resource and material intervention.
Figure 8: Graph depicting the performance of Group B participants providing learning opportunities in consecutive daily sessions when provided the resource/material intervention and when their supervisors received bi-weekly feedback on maintaining the intervention.
Figure 9: Graph depicting the weekly average performance of Group B participants providing learning opportunities when provided the resource/material intervention and when their supervisors received bi-weekly feedback on maintaining the intervention.
**Figure 10:** Graph depicting the isolated performance of Participant 1 in Group B providing learning opportunities in consecutive daily sessions when provided the resource/material intervention and when their supervisor received bi-weekly feedback on maintaining the intervention.
Figure 11: Graph depicting the isolated performance of Participant 2 in Group B providing learning opportunities in consecutive daily sessions when provided the resource/material intervention and when their supervisor received bi-weekly feedback on maintaining the intervention.
Figure 12: Graph depicting the isolated performance of Participant 3 in Group B providing learning opportunities in consecutive daily sessions when provided the resource/material intervention and when their supervisor received bi-weekly feedback on maintaining the intervention.
Figure 13: Graph depicting the average performance of each group’s supervisors’ fidelity in maintaining the resource/material intervention for behavior technicians. This includes the baseline average and weeks 2-5 of the intervention phase. Week 1 is not pictured as this was when the researcher arranged the resources and materials for the behavior technicians (i.e., fidelity was 100%).
Figure 14: Graph depicting the effect of intervention group on mean changes in learning opportunities per hour.
Figure 15: Graph depicting the results of the procedural acceptability surveys for behavior technicians participating in the single-case designs (i.e., Group A and B) and their average responses to each item.
Appendix A

Images of Client Material Storage Prior to Intervention
Appendix B

Sample Image of Tiered Storage Box for Program Stimuli
Appendix C

Images of Client Material Storage Post Intervention
Appendix D

Sample Program Card

Depicts a blank program card, front (top) and back (bottom).

![Sample Program Card](image)

<table>
<thead>
<tr>
<th>PROGRAM:</th>
<th>STIMULI/EXAMPLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>INSTRUCTION/CUE</td>
<td>1.</td>
</tr>
<tr>
<td>RESPONSE</td>
<td>3.</td>
</tr>
<tr>
<td>CORRECT</td>
<td>6.</td>
</tr>
<tr>
<td>INCORRECT</td>
<td>7.</td>
</tr>
<tr>
<td></td>
<td>8.</td>
</tr>
</tbody>
</table>
Appendix E

Blank Case Stimuli and Resource Checklist

<table>
<thead>
<tr>
<th>Case Stimuli and Resource Checklist</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client Initiate: [ ]</td>
</tr>
<tr>
<td>Have programs been updated since the last observation?</td>
</tr>
<tr>
<td>Are all necessary stimuli for specified preventative strategies and problem behavior... (if none specified, mark N/A)</td>
</tr>
<tr>
<td>...accurate compared to descriptions in Central Reach and/or other written protocols (i.e., correspondence between storage and CR session)?</td>
</tr>
<tr>
<td>...present and readily accessible in the designated storage area (i.e., stored in a centralized area)?</td>
</tr>
<tr>
<td>...clearly labelled and organized (i.e., easily identifiable and grouped together)?</td>
</tr>
<tr>
<td>...in good working condition (i.e., not damaged, frayed, or illegible)?</td>
</tr>
<tr>
<td>Are all necessary stimuli for specified acquisition programs... (if none specified, mark N/A)</td>
</tr>
<tr>
<td>...accurate compared to descriptions in Central Reach and/or other written protocols (i.e., correspondence between storage and CR session)?</td>
</tr>
<tr>
<td>...present and readily accessible in the designated storage area (i.e., stored in a centralized area)?</td>
</tr>
<tr>
<td>...clearly labelled and organized (i.e., easily identifiable and grouped together)?</td>
</tr>
<tr>
<td>...in good working condition (i.e., not damaged, frayed, or illegible)?</td>
</tr>
<tr>
<td>Are all necessary stimuli used as rewards/reinforcers... (if none specified, mark N/A)</td>
</tr>
<tr>
<td>...accurate compared to descriptions in Central Reach and/or other written protocols (i.e., correspondence between storage and CR session)?</td>
</tr>
<tr>
<td>...present and readily accessible in the designated storage area (i.e., stored in a centralized area)?</td>
</tr>
<tr>
<td>...clearly labelled and organized (i.e., easily identifiable and grouped together)?</td>
</tr>
<tr>
<td>...in good working condition (i.e., not damaged, frayed, or illegible)?</td>
</tr>
<tr>
<td>...if edible, are these items dated within 1 week and appear fresh (i.e., no mold or signs of decay)?</td>
</tr>
<tr>
<td>TOTAL</td>
</tr>
</tbody>
</table>
Appendix F

Bi-Weekly Feedback Email Sample and Checklist

---

Gmail

Case Resources for REDACTED

REDACTED <redacted@gmail.com>

To: <redacted@gmail.com> <REDACTED@gmail.com>

Hello REDACTED,

As part of improving behavior technician access to materials and resources, we are also looking at the condition of materials and organization of those materials on a bi-weekly basis. We reviewed REDACTED’s materials last week to see if staff had the necessary materials prescribed for antecedent manipulations, acquisition programs, and consequence-based strategies. Our observation is attached.

The materials look great! Most of necessary materials are present and are in reasonable, working condition.

Thank you!

REDACTED
Get Outlook for iOS

---

Case Stimuli and Resource Checklist

<table>
<thead>
<tr>
<th>Client Initials</th>
<th>Supervisor</th>
<th>Date of Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Have programs been updated since the last observation?

- Yes
- No
- N/A
- Active
- Inactive
- Date received

Are all necessary stimulus for specified preventative strategies and problem behavior, (If more specified, mark N/A):
- Accurate compared to description in Central District and/or other written protocols (e.g., correspondence between storage and OR session(s))
- Present and readily accessible in the designated storage area (e.g., stored in a designated area, clearly identified and organized (e.g., easily identifiable and grouped together))
- In good working condition (i.e., not damaged, frayed, or illegible)

Are all necessary stimulus used as rewards/reinforcers. (If more specified, mark N/A):
- Accurate compared to description in Central District and/or other written protocols (e.g., correspondence between storage and OR session(s))
- Present and readily accessible in the designated storage area (i.e., stored in a designated area, clearly identified and organized (i.e., easily identifiable and grouped together))
- In good working condition (i.e., not damaged, frayed, or illegible)

100%
Appendix G

The PDC-HS (from Carr et al., 2016)
<table>
<thead>
<tr>
<th>3°</th>
<th>Yes</th>
<th>No</th>
<th>N/A</th>
<th>Are the materials necessary to complete the task well designed for their intended purpose?</th>
</tr>
</thead>
<tbody>
<tr>
<td>4°</td>
<td>Yes</td>
<td>No</td>
<td>N/A</td>
<td>Are the materials necessary to complete the task well organized for their intended purpose?</td>
</tr>
<tr>
<td>5</td>
<td>Yes</td>
<td>No</td>
<td>N/A</td>
<td>Can the task be completed without first completing other tasks? If not, indicate below the tasks that must be completed first.</td>
</tr>
<tr>
<td></td>
<td>Task 1:</td>
<td>Task 2:</td>
<td>Task 3:</td>
<td>Task 4:</td>
</tr>
<tr>
<td>6</td>
<td>Yes</td>
<td>No</td>
<td>N/A</td>
<td>If you answered NO for Question 5, are other employees responsible for completing any of the earlier tasks in the process? If so, indicate the employee(s) below.</td>
</tr>
<tr>
<td></td>
<td>Task 1:</td>
<td>Task 2:</td>
<td>Task 3:</td>
<td>Task 4:</td>
</tr>
</tbody>
</table>

**PERFORMANCE CONSEQUENCES, EFFORT, & COMPETITION**

<p>| 1 | Yes | No | Is the employee ever directly monitored by a supervisor? If so, indicate the frequency of monitoring. |
|   | ☑ hourly | ☑ daily | ☑ weekly | ☑ monthly | ☑ Other: |
| 2 | Yes | No | Does the employee ever receive feedback about the performance? If yes, indicate below. |
|   | By whom? | How often? | Delay from task? |
|   | Check all that apply: |
|   | Feedback Focus: ☑ Positive | ☑ Corrective |
|   | Feedback Type: ☑ Written | ☑ Verbal | ☑ Graphed | ☑ Other: |
| 3 | Yes | No | Does the employee ever see the effects of accurate task completion? If yes, how? |
| 4 | Yes | No | Is the task simple or does it involve relatively low response effort? |
| 5 | Yes | No | Does the task generally take precedence over other potentially competing tasks? If not, indicate these competing tasks below. |
|    | Task 1: | Task 2: | Task 3: | Task 4: |</p>
<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Are the materials necessary to complete the task well designed for their intended purpose?</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>Are the materials necessary to complete the task well organized for their intended purpose?</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td>Can the task be completed without first completing other tasks?? If not, indicate below the tasks that must be completed first.</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td>Task 1:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Task 2:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Task 3:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Task 4:</td>
</tr>
<tr>
<td>6</td>
<td>Yes</td>
<td>No</td>
<td>If you answered NO for Question 5, are other employees responsible for completing any of the earlier tasks in the process? If so, indicate the employee(s) below.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Task 1:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Task 2:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Task 3:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Task 4:</td>
</tr>
</tbody>
</table>

**PERFORMANCE CONSEQUENCES, EFFORT, & COMPETITION**

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>No</th>
<th>Is the employee ever directly monitored by a supervisor? If so, indicate the frequency of monitoring.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>◼ hourly ◼ daily ◼ weekly ◼ monthly ◼ Other:</td>
</tr>
<tr>
<td>2</td>
<td>Yes</td>
<td>No</td>
<td>Does the employee ever receive feedback about the performance? If yes, indicate below.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>By whom? ◼ How often? ◼ Delay from task?</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Check all that apply: ◼ Positive ◼ Corrective ◼ Feedback Type: ◼ Written ◼ Verbal ◼ Graphed ◼ Other:</td>
</tr>
<tr>
<td>3</td>
<td>Yes</td>
<td>No</td>
<td>Does the employee ever see the effects of accurate task completion? If yes, how?</td>
</tr>
<tr>
<td>4</td>
<td>Yes</td>
<td>No</td>
<td>Is the task simple or does it involve relatively low response effort?</td>
</tr>
<tr>
<td>5</td>
<td>Yes</td>
<td>No</td>
<td>Does the task generally take precedence over other potentially competing tasks? If not, indicate these competing tasks below.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Task 1:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Task 2:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Task 3:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Task 4:</td>
</tr>
</tbody>
</table>
## Intervention Planning

**Instructions:** Each item scored as NO on the PDC-HS should be considered as an opportunity for intervention with priority given to areas in which multiple items are endorsed. Interventions may be implemented concurrently or consecutively, with the latter option being preferred for settings in which staff resources are limited. Sample interventions and illustrative literature citations for each area are provided below.

<table>
<thead>
<tr>
<th>Area</th>
<th>Item #</th>
<th>Sample Intervention(s)</th>
<th>Literature Citations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>1, 2, 3, 4</td>
<td>Behavioral skills training (i.e., instructions, modeling, rehearsal, feedback)</td>
<td>• Barnes, Dunning, &amp; Rehfeldt (2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Improved personnel selection</td>
<td>• Nabeyama &amp; Sturkey (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Gatewood, Feld, &amp; Barrick (2008)</td>
</tr>
<tr>
<td>Task Clarification &amp; Prompting</td>
<td>1, 2</td>
<td>Task clarification &amp; checklists</td>
<td>• Cunningham &amp; Austin (2007)</td>
</tr>
<tr>
<td></td>
<td>3, 4</td>
<td>Prompts</td>
<td>• Gravina, VanWagner, &amp; Austin (2008)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Change/alter task location</td>
<td>• Bacon, Fulton, &amp; Malott (1982)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• May, Austin, &amp; Dymond (2011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Petscher &amp; Bailey (2006)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Green, Reid, Passante, &amp; Canipe (2008)</td>
</tr>
<tr>
<td></td>
<td>2, 3, 4</td>
<td>Improve access to (2), redesign (3), or reorganize (4) task materials</td>
<td>• Casella, Wilder, Neidert, Rey, Compton &amp; Chong (2010)</td>
</tr>
<tr>
<td></td>
<td>5, 6</td>
<td>Reassess task process and personnel</td>
<td>• Diener, McGee, &amp; Miguel (2009)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• McGee &amp; Diener (2010)</td>
</tr>
<tr>
<td>Performance Consequences, Effort, &amp; Competition</td>
<td>1</td>
<td>Increased supervisor presence</td>
<td>• Mozingo, Smith, Riordan, Reiss, &amp; Bailey (2006)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Performance feedback</td>
<td>• Arco (2008)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Regularly highlight task outcomes</td>
<td>• Green, Rollyson, Passante, &amp; Reid (2002)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Reduce task effort</td>
<td>• Methot, Williams, Cummings, &amp; Bradshaw (1996)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Reduce aversive task properties</td>
<td>• Casella, Wilder, Neidert, Rey, Compton, &amp; Chong (2010)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Green, Reid, Passante, &amp; Canipe (2008)</td>
</tr>
</tbody>
</table>
REFERENCES


Appendix H

Procedural Acceptability Survey

The following images depict the social validity survey that participants received along with debriefing on experimental procedures. The top image is the version that behavior technicians and the two supervisors from the first group received. The bottom image is the amended version that the two supervisors in the second group received. As they were subject to an additional intervention, four questions (see highlighted items) were added regarding the checklist and feedback received from clinical directors.
### Social Validity

Please read each question and answer honestly.

<table>
<thead>
<tr>
<th>Having an organizer and materials helped run trials and provide learning opportunities more frequently.</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Having the organizer and the materials was a positive change for my job.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Having the organizer and the materials was a positive change for the organization.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I would be likely to recommend organization of materials and resources as a strategy to promote other important job duties.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>Having more resources and materials helps the kids at [redacted].</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
<tr>
<td>I liked the checklist/job aid for case stimuli and materials.</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>

Is there any other feedback you’d like to provide based on the intervention (the materials and organizer) or the feedback you received?