MAXIMIZING SUPERVISORS’ EFFICIENCY: THE USE OF ENHANCED WRITTEN INSTRUCTIONS TO TEACH UNDERGRADUATES TO IMPLEMENT A STIMULUS PREFERENCE ASSESSMENT

A thesis in preparation for partial fulfillment of the requirements

For the degree of Master of Arts in Psychology

Clinical Psychology

By

Marnie Nicole Shapiro

May 2013
The Thesis of Marnie Nicole Shapiro is approved:

____________________________________  _________________________
Tara Fahmie, Ph.D., BCBA-D

____________________________________  _________________________
Gary Katz, Ph.D.

____________________________________  _________________________
Ellie Kazemi, Ph.D., BCBA-D, Chair

California State University, Northridge
Acknowledgement Page

I would like to thank the KLAB research team for the tremendous effort they put forth in assisting with the development and completion of this research investigation. I know each of you spent a great deal of time inside and outside lab meetings generating ideas and refining the methodology. You all have been instrumental in my growth as a scholar, and I deeply appreciate all of your support and constructive feedback.

I would like to express my sincerest gratitude to my thesis committee, Dr. Katz and Dr. Fahmie, for your continued direction, critical feedback, and assistance throughout this investigation. Dr. Katz, you have always served as a pillar of support for me throughout my time at CSUN. Dr. Fahmie, you have given me some excellent advice that I will continue to follow as I make my way through graduate school.

Most importantly, I would like to thank my advisor and thesis chair, Dr. Kazemi. By far, you have been the most influential person in my undergraduate and graduate career. I truly thank you for broadening my educational perspective, investing immeasurable personal and professional resources into my scholarly trajectory, and shaping me into the individual I am today. Without your emotional and professional support, dedication, and guidance I would have not been prepared to move on to a Ph.D. program. You believed in me more than I believed in myself and showed me that with persistence and hard work, anything could be accomplished. You ensured I was prepared for every possible challenge that I could face and have served as an excellent model of the type of scholar, professor, and scientist I hope to emulate. It is an honor to say that I have been one of your mentees. I hope one day I will be able to make a difference in a student’s life that mirrors that which you have made in mine.
# Table of Contents

Signature page ii  
Acknowledgement iii  
Abstract v  
Introduction 1  
Method 13  
Results 20  
Discussion 22  
References 27  

Appendices:

A- Table 1 31  
B- Table 2 32  
C- Figure 1 33
ABSTRACT

MAXIMIZING SUPERVISORS’ EFFICIENCY: THE USE OF ENHANCED WRITTEN INSTRUCTIONS TO TEACH UNDERGRADUATES TO IMPLEMENT A STIMULUS PREFERENCE ASSESSMENT

By

Marnie Nicole Shapiro

Master of Arts in Psychology
Clinical Psychology

The efficacy of behavior change programs depends upon staffs’ ability to identify stimuli that may function as reinforcers for individual consumers. Graff & Karsten (2012) were the first researchers to test the effectiveness of a self-instructional package (i.e., enhanced written instructions) that did not require a supervisor to model appropriate skills, rehearse with staff members, or provide feedback for staff to meet the mastery criterion. The purpose of my study was to replicate the methods used in the study conducted by Graff and Karsten. A multiple baseline design across participants was used to train three undergraduate students to implement, score, and interpret the outcomes from a paired-stimulus preference assessment with a simulated consumer. Results indicated that all participants demonstrated low to moderate levels of accuracy with the inclusion of written instructions at baseline. Once the enhanced written instructions developed by Graff and Karsten was introduced, only one out of three participants met the mastery criterion for implementation. These findings suggest that the inclusion of enhanced written instructions alone may not be sufficient for all individuals to acquire the basic skills for conducting a stimulus-preference assessment.
Introduction

Applied behavior analysis (ABA) is the science of analyzing and improving socially significant behavior (Cooper, Heron, & Heward, 2007). Researchers have demonstrated the effectiveness of ABA treatments for a wide range of individuals presenting with numerous behavioral challenges. Two hallmarks of ABA include precise and consistent recordings of observable and measurable behaviors (i.e., the behavioral dimension of ABA) and application of behavior change procedures that produce socially significant outcomes (i.e., the effective dimension; see Baer, Wolf, and Risley, 1968). By making data-based decisions that guide interventions, researchers and practitioners can further modify each treatment plan to shape individual behavior to achieve best treatment outcomes. A unique feature of ABA treatment implementation is the use of a tiered service delivery model. A doctoral or master’s level Behavior Analyst designs and monitors programs and trains relatively inexperienced bachelor’s level Behavioral Technicians to carry out behavior analytic treatments.

ABA is in a period of immense growth and the number of staff serving children in need of services exceeds the number of Behavior Analysts who can train and monitor staff performance. Therefore, it is imperative to develop time efficient training strategies to maximize Behavior Analysts’ efficiency. Researchers have demonstrated that aspects of behavioral technology can be taught to novice staff with relative ease in a relatively short duration of time (Catania, Almeida, Liu-Constant, & DiGennaro Reed, 2009; Graff & Karsten, 2012; Iwata et al., 2000; Roscoe & Fisher, 2008). Thus, numerous researchers have begun developing effective and efficient antecedent- and consequence-based training strategies to teach inexperienced staff to implement a range of behavior change
and assessment procedures, including but not limited to, managing classroom behavior through the use of a token economy system and teaching new skills through the use of least-to-most prompting procedures and task analyses (Bishop & Kenzer, 2012; Petscher & Bailey, 2006; Selinske, Greer, & Lodhi, 1991). One assessment procedure researchers often train practitioners to implement is a preference assessment because it is commonly employed in the field to identify a range of potent reinforcers to use when training.

**The Importance and Use of Preference Assessments**

Identifying potent reinforcers are a necessary component when employing positive reinforcement contingencies to increase skill acquisition or modify problem behaviors (Karsten & Carr, 2009; Wallace, Iwata, & Hanley, 2006). When practitioners develop positive reinforcement-based programs it is important to be cognizant of factors that may influence the effectiveness of such programs, including idiosyncratic consumer preferences that vary over time as a function of contextual factors (e.g., deprivation and satiation conditions prior to sessions). Therefore, it is imperative that practitioners and educators identify consumer preferences regularly during training or teaching sessions (Hanley, Iwata, & Roscoe, 2006). Numerous authors have demonstrated the predictive validity of some preference assessment procedures such that stimuli identified as preferred at the end of the assessment functioned as positive reinforcers in follow-up reinforcer assessments resulting in successful treatment outcomes (DeLeon & Iwata, 1996; Fisher et al., 1992; Ringdahl, Vollmer, Marcus, & Sloane, 1997). There are currently various methods practitioners employ to assess consumer preference, including engagement-based methods (i.e., measuring duration of interaction with stimuli) and approach-based methods (i.e., measuring any detectable movement towards stimuli). For
the purposes of this study, I will focus on two of the most commonly reported direct approach-based methods for identifying potential reinforcers. Specifically, I will address both the paired-stimulus (PS) and multiple-stimulus without replacement (MSWO) because researchers have demonstrated the predictive validity of these assessments across a wide range of individuals (Hanley et al., 2006).

**Approached-Based Methods of Assessment**

Practitioners use approach-based methods of assessment to systematically evaluate and record the number of times the consumer approaches each stimulus across multiple trials. In the PS method, the practitioner presents two stimuli simultaneously and records which stimulus the consumer chooses. Throughout the course of the assessment period, the practitioner must pair each stimulus with all other stimuli in the predetermined pool of stimuli and counterbalance the left versus right presentation of items across trials (Fisher et al., 1992; Pence, Peter, & Tetreault, 2012). For example, Fisher et al. (1992) presented 16 stimuli in pairs and, in a randomized order, paired each stimulus with every other stimulus resulting in 120 stimulus-pair presentations. Following the assessment, the practitioner must calculate selection percentages for each stimulus and interpret outcomes (e.g., discriminate between high and low-preference stimuli from the assessment results). Specifically, a consumer’s preference is determined by calculating the percentage of trials the consumer selected each stimulus relative to the number of trials the practitioner presented each stimulus (Hagopian, Long, & Rush, 2004). Results from a PS assessment yield accurate distinctions between high and low valued preference stimuli. More specifically, high-preference stimuli (i.e., items selected on at least 80% of trials) had high predictive validity and were more likely to function as
effective reinforcers than low to moderate-preference stimuli (i.e., items selected on less than 60% of trials; Fisher et al., 1992). However, the main limitation of the PS procedure is that the method is time consuming (approximately 20-50 min, depending on number of stimuli used) compared to other types of direct methods of assessment (Hagopian et al., 2004).

In the MSWO method, first introduced by DeLeon and Iwata (1996), the practitioner presents an array of stimuli and records which stimulus the consumer chooses. Once the consumer chooses a particular stimulus from the array, the practitioner removes that stimulus on subsequent trials and rotates the remaining stimuli in the array resulting in a reduced number of stimuli from which the consumer is to choose. After approximately five sessions, a consumer’s preference is determined by dividing the number of times the consumer selected each stimulus by the number of trials the practitioner presented that stimulus across all sessions (DeLeon & Iwata, 1996; Hagopian et al., 2004). DeLeon and Iwata (1996) showed that the MSWO has good concurrent validity with the PS assessment but takes less time to administer. Thus, for practitioners, the MSWO assessment is both an efficient and accurate assessment procedure. The main limitation of the MSWO procedure is that the consumer must possess the skill set to be able to scan and select from a large array of stimuli (Pence et al., 2012). Both the PS and MSWO direct assessment procedures enable practitioners to generate preference hierarchies with relative ease.

**Training Procedures Used to Teach Preference Assessments**

Training of staff to implement preference assessments is of paramount importance because the efficacy of behavior change programs depends on staffs’ ability to identify
stimuli that may function as reinforcers for individual consumers (Paramore & Higbee, 2005). DeLeon and Iwata (1996) and Fisher et al. (1992) were the first researchers to provide a task analysis outlining how to conduct a MSWO or PS assessment; however, the authors’ main objective was to determine which assessment type accurately predicted the reinforcement effect of individual stimuli rather than to determine best methods to teach staff to implement such assessments. Subsequently, researchers began to identify best practices for teaching staff to implement preference assessments with procedural fidelity (Lavie & Sturmey, 2002; Lerman, Tetreault, Hovanetz, Strobel, & Garro, 2008; Graff & Karsten, 2012; Roscoe & Fisher, 2008).

In their seminal research article, Lavie and Sturmey (2002) used a multiple baseline design across participants to investigate the efficacy of their brief instructional package (i.e., video model, rehearsal, and feedback) to train three novice staff members to conduct paired-stimulus preference assessments. Prior to training, staff self-reported that they experienced difficulty with identifying stimuli that could function as potential reinforcers for students diagnosed with autism. At baseline, trainers asked staff to determine a pool of stimuli which may function as reinforcers for the children based on dietary restrictions, ease of presentation and availability, and reports generated by other individuals regarding reinforcing stimuli. The trainers placed the stimuli identified by staff (e.g., candy, cookies, and a toy car) on the table and presented staff with a piece of paper to collect data.

Following baseline, trainers gave staff a task analysis (Fisher et al., 1992) outlining the required steps to perform the preference assessment and described the corresponding target skills. Staff underwent two 40 to 50 min individualized training
sessions during which the trainers provided staff with a checklist illustrating the target skills required for mastery and verbally described each target skill to the participants. Next, staff members viewed a videotape in which an individual modeled the target skills and assessment procedures. Following the videotape, staff members implemented preference assessments with consumers as trainers observed and provided corrective or approving feedback on all target skills contingent on staff performance. If a staff member did not meet the mastery criterion, which was accurate implementation of at least 85% of the target skills across two consecutive sessions, the staff member viewed the videotape again prior to performing subsequent trials. The authors reported that all staff members met the mastery criterion following only two training sessions lasting a total of 80 min. In sum, the authors found staff demonstrated a high level of accuracy when trainers incorporated brief instruction, video modeling, rehearsal, and feedback into two 40 to 50 min training sessions.

In a related article, Roscoe, Fisher, Glover, and Volkert (2006) evaluated a procedure to train staff with minimal experience to accurately implement both the PS and MSWO preference assessments. Participants included four individuals who earned their bachelor’s degree and had no formal training conducting preference assessments, as well as staff who served as “simulated” consumers. The authors provided the simulated consumers with a script outlining how to respond during each trial. During baseline, the authors gave staff a blank piece of paper and the stimuli needed to conduct the assessment. The authors provided no further written or verbal instructions. Following baseline, the authors gave staff the methods section from two seminal articles (see Fisher et al., 1992 and DeLeon & Iwata, 1996) but made slight modifications to the procedures
such that they indicated how staff should respond if the child were to select two items in rapid succession or concomitantly. Following baseline and instructions-only training, none of the staff met the mastery criterion.

For the next phase, the authors randomly assigned staff to one of two conditions: corrective feedback or contingent money conditions. In the feedback condition, staff viewed videotapes of their own performances from the prior sessions and the experimenter provided them with corrective feedback. In the contingent money condition, the authors presented staff with contingent money, independent of corrective feedback, in a predetermined sum commensurate to the percentage of staff’s correct responses. Roscoe et al. found that the inclusion of written instructions alone or contingent money without corrective feedback produced inconsistent responding and only small increases in accurate performance. Conversely, corrective feedback resulted in large, rapid, and consistent increases in accurate performance. Thus, performance feedback from the trainer was needed for staff to meet the mastery criterion.

Although Roscoe et al. (2006) demonstrated that feedback was necessary for staff to meet the mastery criterion, staff had multiple opportunities to conduct the assessments before the authors introduced the feedback component. Thus, the authors failed to show that feedback alone was sufficient to demonstrate mastery because it was unclear whether staff performance would have increased as rapidly if the authors had provided staff with only one or two training sessions before they introduced the feedback component. In a follow-up study, Roscoe and Fisher (2008) replicated and extended the research of Roscoe et al. (2006) by isolating the most effective training components and including only one or two baseline sessions. Specifically, the authors examined the efficacy of a
brief training procedure that integrated feedback and role-play to train eight inexperienced behavioral technicians to implement both PS and MSWO assessments. The authors found that a single 15-20 min training session consisting of feedback in conjunction with role-play was sufficient to train staff to correctly implement two types of preference assessments.

It is important to note that despite the strengths in the aforementioned studies, they have several limitations. First, the authors did not report on staffs’ ability to score preference assessments (Roscoe & Fisher, 2008) nor did they ask staff to generate preference hierarchies and evaluate assessment outcomes (Lavie & Sturmey, 2002; Roscoe et al., 2006). Thus, even though staff could accurately implement a preference assessment following training, it is unclear whether staff could interpret and translate results from this assessment into effective interventions. Finally, the feedback component in each study required that a trainer be present for at least a portion of the training sessions; however, many staff and educators do not have continuous access to a supervisor on site (Graff & Karsten, 2012). In line with these limitations, researchers are beginning to question how they can employ training methods that maximize efficiency by reducing the need for the presence of a supervisor.

**Training Procedures that Maximize an Onsite Supervisor’s Efficiency**

Graff and Karsten (2012) argued that even though researchers have advanced the efficiency of preference assessment training procedures (e.g., decreasing duration of training), researchers have not yet determined a way to minimize the need of a supervisor who must be present to model appropriate implementation or provide feedback.
contingent on performance in order for staff to reach mastery. One strategy to maximize a supervisor’s efficiency in field involves giving staff a comprehensive self-instructional package that eliminates the need for performance feedback to reach the mastery criteria. In 2012, Graff and Karsten used a multiple baseline design across behaviors to evaluate the effects of a self-instructional package to train inexperienced staff to implement, score, and interpret the results from both the PS and MSWO assessments. Participants included 11 teachers without prior experience implementing preference assessments employed at a school for children with autism and related developmental disabilities. Prior to initiating the study, teachers completed a ten item pretest (e.g., multiple choice or matching questions) to assess their knowledge of preference assessments. To be included in the study, the participant needed to score 50% or lower on this pretest.

In addition to the 11 teachers who served as participants, Graff and Karsten (2012) used 3 BCBAs who worked within the school site to serve as simulated consumers. The authors provided the simulated consumers with a script outlining how to respond during each trial. For approximately half of the trials, the simulated consumers responded atypically consisting of selecting two items at once, engaging in any behavior other than selecting any stimuli presented during that trial, or attempting to select an item that was not used in the corresponding trial. Prior to each session, the authors informed the teachers which preference assessment they would be conducting and supplied them with the essential training materials needed to conduct the assessments (e.g., data sheets, writing utensils, and a predetermined pool of stimuli). During instruction, teachers predominantly used eight edible items (e.g., pretzel, cracker, candy corn, cookie, corn chip, marshmallow, raisin, chocolate candy) to conduct the assessment.
At baseline, the authors gave all teachers written instructions from the methods section of Fisher et al. (1992), for the PS assessment, and DeLeon and Iwata (1996), for the MSWO assessment, and teachers had 30 min to review the material. The authors stated that they counterbalanced assessment types to control for sequence effects. After 30 min elapsed, the authors told the teachers to implement preference assessment procedures based on the written instructions. At the end of baseline, the authors asked staff to calculate selection percentages for all of their collected data. The authors also asked staff to select the item they would use to teach the consumer a new skill, based on their data.

The authors then exposed the teachers to two different training procedures: 1) enhanced written instructions consisting of step-by-step instructions that contained limited technical jargon; diagrams; pictures; and comprehensive data sheets, or 2) written instructions and data sheets consisting of the same written instructions from baseline with the added data sheet contained in the enhanced written instructions package. The latter training procedure afforded the authors the opportunity to analyze the effects of the data sheets alone on accurate implementation. At the end of training, the authors gave staff hypothetical data and asked them to calculate selection percentages and identify the stimulus that they would use to teach a new skill. In sum, five out of eleven teachers advanced from written instructions (baseline), to written instructions and data sheet, to enhanced written instructions, to generalization probes with real consumers from the school setting. The remaining six teachers advanced from written instructions (baseline), to enhanced written instructions, to generalization probes.
The authors collected data on five different target responses specific to implementation: (a) Stimulus presentation- replacement or presentation of items; (b) Stimulus placement - where participants placed the stimuli in relation to the consumer; (c) Postselection response - teacher response after an item was chosen; (d) Response blocking - blocking approaches to more than one stimuli; and (e) Trial termination- termination of trial if consumer engaged in any behavior other than selecting any stimuli presented during that trial. For each trial, observers recorded whether the teacher accurately performed all of the aforementioned target responses and scored the trial as incorrect if the participant incorrectly executed a target response. The authors also had secondary observers note whether teachers accurately recorded consumer responses, calculated correct percentages, and named the correct stimulus that they would use to teach a new skill. The authors found that with the written instructions and written instructions plus data sheets, none of the teachers reached the mastery criterion, which was 90% or greater accuracy across two consecutive sessions consisting of approximately eight to ten trials each. Conversely, with the use of enhanced written instructions, mean total accuracy for all teachers increased to well above mastery criterion and the authors used generalization probes conducted one week to one month following mastery to demonstrate that performance gains maintained overtime and generalized from simulated consumers to actual consumers. At the completion of the study, staff were also able to create a preference hierarchy and accurately identify which item should be used to teach a consumer new skills.

**Summary and Research Objective**

The number of staff serving individuals with special needs far exceeds the number
of certified behavior analysts who can supervise, train, and monitor staff performance (Simpson, 2004). For example, according to the National Center for Education Statistics (2006), there are 6.7 million students with special needs, but only 9,736 BCBA s in the U.S (“BACB Certificants Now Exceed 10,000 Worldwide”, 2011). Therefore, it is imperative to look for low cost, portable training programs with few technological demands that can be used to replace or supplement the frequent need for a supervisor on site. Graff and Karsten (2012) were the first researchers to address some of the previously mentioned limitations by testing the efficacy of a self-instructional package with enhanced written instructions that did not require feedback by a supervisor in order for staff to meet the mastery criterion. Furthermore, the authors taught staff to accurately implement preference assessments and to score and interpret the outcomes from these assessments. Thus, the purpose of this study is to replicate the methods used in the study conducted by Graff and Karsten (2012) using undergraduate students, who often serve as frontline staff for in-home behavioral services.
Method

Participants

Participants included three undergraduate students from California State University, Northridge. The undergraduate students between the ages of 19 - 35 (M = 1, F = 2) were enrolled in an introductory course in applied behavior analysis and self-reported that they never worked at a behavioral agency. To encourage participation in this study, the experimenter sent an IRB-approved flier to professors who taught undergraduate and graduate core requisite coursework in psychology with an emphasis in ABA. To be included in this study, a participant must not have received any formal training observing or conducting stimulus preference assessments. If at baseline a participant achieved mastery, which was a score of 90% or greater across two consecutive baseline sessions, their data were not included in the analyses. All participants gave informed consent prior to beginning the study and were debriefed at the end. The experimenter treated all participants in accordance with APA ethical guidelines for conducting research.

Settings and Materials

The experimenter videotaped all training sessions for each participant; all sessions were carried out in small observation rooms located on campus. The room contained basic materials needed to conduct a preference assessment: paper, pencil, calculator, table, chairs, and the same predetermined pool of eight edible stimuli used by Graff and Karsten in their enhanced written instructions. The experimenter arranged the items in the room such that the simulated consumer always sat on the same side of the table, across from the participant.
Simulated consumer and assessment scripts. Roscoe et al. (2006) and Graff and Karsten (2012) demonstrated that skills learned during practice with simulated consumers do in fact generalize to actual consumers. One student assistant was trained to behave as the simulated consumer during each session. The simulated consumer did not provide participants with any feedback related to the assessment procedures throughout the trials. In accordance with the procedures of Graff and Karsten (2012), the experimenter used four different scripts to minimize carryover and practice effects and the scripts designated how the simulated consumer responded (i.e., atypical and typical responses) across ten trials. Typical responses included the consumer choosing one item within 5 s of the presentation of two stimuli. Atypical responses included the consumer (a) simultaneously choosing more than one stimuli (one trial), (b) choosing two stimuli in quick succession (one trial), (c) engaging in any behavior other than selecting any stimuli presented during that trial (two trials), (d) or selecting a stimulus that was not presented on a trial (one trial). The experimenter designated atypical responses to occur in 50% of the trials and counterbalanced the order of the scripts across trials. For the remaining trials, the simulated consumer chose only one stimulus. To ensure procedural fidelity and collect treatment integrity data, a research assistant (RA) read the script to the simulated consumer for each trial via a Bluetooth device.

Written instructions at baseline. The experimenter gave all participants the methods section from Fisher et al. (1992). The written instructions from Fisher et al. (1992) contained the following directions: (a) place two items on the table approximately one foot in front of the consumer and one foot apart; (b) verbally prompt the consumer to select one item; (c) remove the item not selected and record the item selected on the data
sheet; (d) if an item is not selected within 5 s of the verbal prompt, record “no response” for that trial and position items for next trial; and (e) if the consumer attempts to simultaneously select more than one item, block the attempt and repeat the trial. Graff and Karsten (2012) noted that they made some additions and modifications to the published procedure section in Fisher et al. (1992). For example, in the written instructions they included how participants should respond when the consumer gave an atypical response (e.g., selecting more than one item in quick succession).

**Enhanced written instructions at intervention.** The experimenter used the enhanced written instructions from the study conducted by Graff and Karsten (2012). The experimenter made a slight modification to the written instructions in order to clearly discriminate when a trial begins. Specifically, the experimenter added a bullet to instruct the participant to verbally prompt the consumer to begin the trial.

**Hypothetical data at baseline and intervention.** To calculate selection percentages and interpret outcomes, the experimenter used the data sheet included in the enhanced written instructions developed by Graff and Kartsen (2012). The experimenter created two different hypothetical sets of data, and counterbalanced the order of the data sheets across participants to make sequence effects more visible. The data sheets included a place for participants to calculate selection percentages from 28 trials using the same eight stimuli presented during the assessment. The experimenter also asked participants, based on their percentage calculations, to write in the item they would use to teach the hypothetical consumer a new skill.

**Design and Procedure**

All RAs conducting research with the participants completed NIH IRB Training
Modules prior to the initiation of the study for the protection of human subjects. Prior to beginning the study, all participants gave informed consent, and the experimenter informed them that she would videotape all training sessions at baseline and intervention phases.

The experimenter used a multiple baseline design across participants to evaluate the effects of the self-instructional manual (Graff and Karsten, 2012). At baseline, the experimenter gave all participants the modified version of the methods section from Fisher et al. (1992). The experimenter provided all participants with a piece of paper and a pen to record the data as well as the same predetermined pool of eight edible items that Graff and Karsten used in their enhanced written instructions (e.g., popcorn, goldfish crackers, teddy graham’s, etc.). The experimenter allowed participants to review the written instructions for a maximum of thirty min before initiating the first session and asked participants to inform the experimenter when they were ready to begin. The experimenter allowed participants to have free access to the written instructions throughout all sessions. The simulated consumer provided no feedback to the participant except to indicate the initiation and completion of a session. At the end of baseline, The experimenter gave the participants hypothetical data and asked them to calculate selection percentages. The experimenter also asked participants, based on this hypothetical data, to select the item they would use to teach the consumer a new skill.

In the intervention phase, the experimenter provided participants with enhanced written instructions (see Graff and Karsten, 2012) and repeated the same procedures from baseline. The experimenter concluded by asking participants if they had any experience working at a behavioral agency as well as a few basic demographic questions. All
participants had the option to receive a break approximately one hour into the training session. Each training session contained 10 trials, and the experimenter conducted the training session for each participant over the course of one day.

**Dependent Measures.** The mastery criterion was for participants to implement the preference assessment with 90% or greater accuracy across two consecutive sessions. The experimenter used the DV measures from Graff and Karsten (2012) and calculated percentage of trials implemented correctly by following their guidelines. Specifically, for Stimulus presentation, observers scored a correct response only if two stimuli were placed on the table in front of the simulated consumer during each trial. For Stimulus position, observers scored a correct response only if the stimuli were placed 1 ft in front of the simulated consumer and 1 ft from one another. For Postselection response, observers scored a correct response only if the participant removed the item that was not selected before recording the consumer’s response. For Response blocking, observers scored a correct response only if the participant moved his/her hand forward towards the consumer’s hand to block attempts to select more than one stimulus during the trial. Finally, for Trial termination, observers scored a correct response only if the participant removed all stimuli from the table when the consumer made no response within approximately 5 s of the stimuli presentation.

The experimenter calculated accuracy of each specific target response by dividing the number of correct target responses by the overall number of opportunities to engage in the correct target response during each session and multiplying by 100%. Total accuracy for implementation was determined by dividing the number of trials during which the participant correctly emitted all target responses by the total number of trials.
implemented and multiplying by 100%. If a participant implemented any one of the five target responses incorrectly, the entire trial was scored as incorrect. The experimenter also collected data pertaining to accuracy of scoring and interpreting outcomes. At the completion of each condition, the experimenter asked each participant to calculate selection percentages from a set of hypothetical data to determine the accuracy of scoring assessment outcomes. A RA then compared the participant’s summary score for each stimulus to their corresponding summary scores. The experimenter ended by asking each participant to write down which stimulus he or she would select to teach the consumer a new skill.

**Interobserver Agreement (IOA).** See Table 1 for IOA data for individual sessions. The experimenter used the same method Graff and Karsten (2012) used to calculate IOA data. In-vivo, the experimenter scored each participant’s target response pertaining to implementation. After training, a secondary RA independently scored 33% of the data via videotape. An agreement was defined as both observers recording an error or correct implementation for a specific target response during a given trial. A disagreement was defined as one observer recording a specific target response as incorrect and the other observer recording that target response as correct for that trial. Specifically, the experimenter obtained IOA data for individual participant responses by dividing the number of agreements for each target response by the number of agreements plus disagreements per session and multiplying by 100%. The mean agreement score across specific participant target responses was 90.4% (range, 78% to 100%).

The experimenter also calculated interobserver agreement for participants’ total accuracy. An agreement was defined as both observers recording the trial as correctly
implemented (i.e., all target responses scored as correct). A disagreement was defined as only one observer scoring correct implementation for a trial. The mean agreement for total accuracy across participants was 83.3% (range, 30% to 100%). The experimenter collected data on the simulated consumer’s implementation of the assessment script. One RA in-vivo recorded, on a trial-by-trial basis, accuracy of the simulated consumer’s responses. A secondary RA independently scored 33% of the data via videotape. A correct response was scored if the simulated consumer performed the behaviors prescribed in the script. An incorrect response was scored if the simulated consumer did not follow the behaviors prescribed in the script. The simulated consumer implemented the assessment scripts with 99.33% accuracy. The mean agreement IOA score for the simulated consumer’s behavior was 100%.
Results

See Figure 1 for the percentage of correct participant responses when implementing preference assessments. At baseline, none of the participants met the mastery criterion using the written instructions alone, and all participants demonstrated low to moderate levels of accuracy conducting the paired-stimulus assessment: Daniel (M = 22%; range, 10% to 30%); Alexandra (M = 20%; range, 10% to 40%); and Christina (M = 60%; range, 50% to 70%). With the addition of the enhanced written instruction, accurate performance increased to moderate levels for Daniel (M = 68%; range 59% to 80%), but did not increase for Alexandra (M = 36%; range, 30% to 40%). Conversely, only one participant met the mastery criterion, which was 90% accuracy across two consecutive sessions: Christina (M = 80%; range, 60% to 90%).

See Table 2 for information regarding specific target responses for each participant. When the experimenter provided the participants with the written instructions alone, all participants demonstrated high scores on stimulus presentation (i.e., presenting two items in front of the consumer; M = 100%). However, accuracy of stimulus placement, post-selection response, response blocking, and trial termination varied across participants. Specifically, mean accuracy of stimulus placement for Daniel, Alexandra, and Christina at baseline ranged from 34% to 98%. For post-selection response, mean accuracy ranged from 40% to 95%. Finally, for response blocking and trial termination, mean total accuracy ranged from 0% to 46.6% and 66.7% to 75%, respectively.

The addition of the enhanced written instructions led to an increase in all participants’ scores from baseline on stimulus placement (M = 96.7%; range, 90% to 100%). For post selection response, Daniel and Christina’s mean total accuracy increased
to 100% whereas Alexandra’s mean total accuracy decreased to 8%. The experimenter observed similar trends for response blocking and trial termination. Specifically, for response blocking, Alexandra’s and Christina’s mean total accuracy scores increased to 76.7% and 86.6%, respectively, whereas Daniel’s scores decreased to 19.8%. For trial termination, Daniels’ mean total accuracy score increased to 100% whereas Alexandra’s and Christina’s scores decreased to 40% and 33.3%, respectively.

To evaluate scoring and interpretation of assessment results, the experimenter asked participants to calculate selection percentages and to name the stimulus he or she would use to teach the simulated consumer a new skill. At baseline, none of the participants were able to calculate selection percentages correctly; however, all participants were able to identify which stimulus the consumer selected most frequently. After the experimenter introduced the enhanced written instructions, only one participant (Alexandra) calculated the selection percentages correctly; however, all participants were again able to identify which stimulus the consumer selected most frequently.
Discussion

Graff & Karsten (2012) were the first researchers to demonstrate that enhanced written instructions, independent of modeling, rehearsal, and feedback, could be used to train teachers to implement a stimulus preference assessment thereby maximizing a supervisor’s efficiency. Therefore, my main objective was to replicate their study with a different population, undergraduate students, who often serve as frontline staff for in-home behavioral services. In line with other studies, students without any formal training in implementation of preference assessments were not able to accurately implement and score the outcomes from a paired-stimulus preference assessment with the written instructions alone (Graff and Karsten, 2012; Lavie & Sturmey, 2002; Roscoe et al., 2006; and Roscoe & Fisher, 2008). Conversely, after reading the enhanced written instructions, only one participant was able to meet the mastery criterion. For the remaining two participants, mean total accuracy for implementation slightly increased (from 21%, at baseline, to 52%, at intervention). Furthermore, after the introduction of the enhanced written instructions only one participant was able to accurately calculate selection percentages; however, all participants were able to identify which stimulus the consumer selected most frequently at both baseline and intervention phases.

Some general differences between my replication study and the study conducted by Graff and Karsten (2012) are worth noting. First, our participants self-reported that they had never worked at a behavioral agency implementing any type of direct instruction. Thus, not only did our participants (similar to Graff and Karsten’s) have no formal training with conducting preference assessments, but they also may have had no experience working with or teaching children in any capacity. However, in the study conducted by Graff and Karsten, all 11 participants held either a bachelor’s or master’s
degree, had worked for several years in a school for children with autism or developmental disabilities, or were new hires to the facility. Given that only one undergraduate participant met the mastery criterion in this replication, researchers whose main objective is to replicate the study conducted by Graff and Karsten should consider identifying variables that may affect performance accuracy (e.g., history of working with children or delivering direct instruction during which blocking or experience with removing potentially distracting stimuli from the child’s environment during instruction has been reinforced).

Even though the published article by Graff and Karsten (2012) was fairly technological for the purpose of this replication, I did encounter a few difficulties such that it was not always possible to carry out the replication by referring only to their manuscript. For example, the article contained limited information regarding participants’ accuracy in scoring and interpreting the paired-stimulus assessment results. In addition, the authors did not thoroughly describe the method by which they asked participants, at intervention, to calculate selection percentages from either hypothetical or real data and to interpret outcomes based on their calculations. Although detailed procedures for calculation of selection percentages and interpretation of outcomes were provided in this manuscript, the method by which Graff and Karsten presented their hypothetical data at the end of intervention may have differed. Given that only one participant in this replication study was able to accurately calculate selection percentages after the intervention, one recommendation for researchers would be to include detailed procedures for scoring and interpretation as well as verbatim instruments used throughout the study.
Researchers should also consider developing a screening tool to identify differences between responders and non-responders so that supervisors can develop efficient methods to bring non-responders to mastery in a short duration of time. If researchers are able to identify reasons as to why some participants were able to reach mastery with the addition of the enhanced written instructions while others were not, the enhanced written instructions could be used to maximize a supervisor’s efficiency in field because modeling, rehearsal, and feedback would not be needed to achieve mastery. Additionally, in the enhanced written instructions developed by Graff and Karsten, instructions specific to atypical responses such as blocking and engaging in behavior other than selecting a stimulus were italicized, bolded, and underlined. Thus, participants may have attended more to these subcomponents because of the added within stimulus prompts and less to other important components (e.g., post-selection responses and stimulus placement). Anecdotally, at the completion of the study, one participant who exhibited difficulty placing two stimuli 1ft apart self-disclosed that she did not know the length of one foot. In line with future research suggestions discussed by Graff and Karsten in their 2012 article, a component analysis should be conducted to isolate the most effective components of the enhanced written instructions (e.g., non-technical language, diagrams, etc.). Future investigators may also wish to determine the extent to which the addition of prompts in the enhanced written instructions can help participants discriminate a foot (e.g., approximately the vertical length of a piece of letter-size paper, 11 x 8 1/2 inches). Finally, throughout the assessment period participants repeatedly referred back to multiple pages in the enhanced written instructions manual resulting in a
delay between trials. In an applied setting, this delay may be problematic if it sets the occasion for the child to engage in problem behavior.

This study contributes further evidence that enhanced written instructions containing pictures, diagrams, and limited technical language could potentially be used to teach novice staff to implement certain behavior analytic procedures. However, we had a number of limitations that may reduce the generality of the findings including undergraduate students who did not work at a behavioral agency and had no immediate reason to learn the necessary skills for implementation, scoring, and interpretation of the assessment. Furthermore, results regarding scoring and interpreting outcomes should be interpreted with caution because the exact methods used to capture the dependent variables (e.g., postselection response) in the study conducted by Graff and Karsten were difficult to discern. Another limitation was that although the experimenter used the exact operational definitions provided by Graff and Karsten to score the DVs specific to implementation, the inter-rater agreement for participants’ mean total accuracy scores related to implementation ranged from poor to excellent. These results may be due to poor operational definitions or lack of multiple exemplars for the raters with regards to the DVs related to blocking and stimulus placement. Despite these noted limitations, I hope that results from this study will help supervisors and researchers begin to identify best practices for teaching staff to implement, score, and interpret the outcomes from preference assessments with procedural fidelity.

In sum, although major advancements have been made in the development of effective and efficient methods to teach accurate implementation of a stimulus-preference assessment, additional research is needed to determine which procedures could be
disseminated via the use of instructions, video, and other mediums of didactic instructions that may increase the efficiency of a supervisor. Results from this study indicate that individuals with no prior history with working with children or implementing direct instruction may require more extensive performance-based training possibly combined with feedback to achieve mastery. Therefore, researchers should focus on identifying efficient individualized training strategies (e.g., brief session of feedback) that can be easily incorporated into the training package if an individual does not meet mastery with enhanced written instructions alone.
References


Table 1

*IOA Data Across Sessions for Individual Target Responses and Means (M) for Specific Target Responses (%)*

<table>
<thead>
<tr>
<th>Participant Pool</th>
<th>Stimulus Presentation^{10}</th>
<th>Stimulus Placement^{10}</th>
<th>Postselection Response^{5}</th>
<th>Response Blocking^{3}</th>
<th>Trial Termination^{2}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daniel</td>
<td>100</td>
<td>96.7</td>
<td>100</td>
<td>77.8</td>
<td>83.3</td>
</tr>
<tr>
<td>Alexandra</td>
<td>100</td>
<td>83.3</td>
<td>100</td>
<td>88.9</td>
<td>83.3</td>
</tr>
<tr>
<td>Christina</td>
<td>100</td>
<td>60</td>
<td>100</td>
<td>88.9</td>
<td>83.3</td>
</tr>
<tr>
<td>M Totals</td>
<td>100</td>
<td>80</td>
<td>100</td>
<td>85.2</td>
<td>83.3</td>
</tr>
</tbody>
</table>
Appendix B

Table 2

*Mean Accuracy (%) for Specific Target Responses Across Participants*

<table>
<thead>
<tr>
<th>Participant Pool</th>
<th>Instruction Type</th>
<th>Stimulus Presentation&lt;sup&gt;10&lt;/sup&gt;</th>
<th>Stimulus Placement&lt;sup&gt;10&lt;/sup&gt;</th>
<th>Postselection Response&lt;sup&gt;5&lt;/sup&gt;</th>
<th>Response Blocking&lt;sup&gt;3&lt;/sup&gt;</th>
<th>Trial Termination&lt;sup&gt;2&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daniel</td>
<td>Written Instructions</td>
<td>100</td>
<td>34</td>
<td>80</td>
<td>46.6</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Enhanced Instructions</td>
<td>100</td>
<td>90</td>
<td>100</td>
<td>19.8</td>
<td>100</td>
</tr>
<tr>
<td>Alexandra</td>
<td>Written Instructions</td>
<td>100</td>
<td>53.3</td>
<td>40</td>
<td>0</td>
<td>66.7</td>
</tr>
<tr>
<td></td>
<td>Enhanced Instructions</td>
<td>100</td>
<td>100</td>
<td>8</td>
<td>86.6</td>
<td>40</td>
</tr>
<tr>
<td>Christina</td>
<td>Written Instructions</td>
<td>100</td>
<td>97.5</td>
<td>95</td>
<td>8.25</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Enhanced Instructions</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>77.7</td>
<td>33.3</td>
</tr>
</tbody>
</table>

Note. Superscript numbers next to the five DVs represents the number of opportunities participants could engage in the target response each session.
Figure 1. Percentage of trials implemented correctly across baseline (i.e., written instructions) and intervention (i.e., enhanced written instructions) conditions for all participants.