9-2018

The Effects of Pyramidal Training on Staff Acquisition of Five Behavior Analytic Procedures

Lindsay M. Maffei-Almodovar

The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://academicworks.cuny.edu/gc_etds

Part of the Applied Behavior Analysis Commons

Recommended Citation

Maffei-Almodovar, Lindsay M., "The Effects of Pyramidal Training on Staff Acquisition of Five Behavior Analytic Procedures" (2018). CUNY Academic Works.
https://academicworks.cuny.edu/gc_etds/2907

This Dissertation is brought to you by CUNY Academic Works. It has been accepted for inclusion in All Dissertations, Theses, and Capstone Projects by an authorized administrator of CUNY Academic Works. For more information, please contact deposit@gc.cuny.edu.
THE EFFECTS OF PYRAMIDAL TRAINING ON STAFF ACQUISITION OF FIVE BEHAVIOR ANALYTIC PROCEDURES

by

LINDSAY MAFFEI- ALMODOVAR

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2018
The Effects of Pyramidal Training on Staff Acquisition of Five Behavior Analytic Procedures

by

Lindsay Maffei-Almodovar

This manuscript has been read and accepted for the Graduate Faculty in psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

August, 30, 2018
Date

Peter Sturmey
Chair of Examining Committee

August, 30, 2018
Date

Richard J. Bodnar
Executive Officer

Supervisory Committee:

Peter Sturmey
Emily Jones
Joshua Jessel

THE CITY UNIVERSITY OF NEW YORK
ABSTRACT

The Effects of Pyramidal Training on Staff Acquisition of Five Behavior Analytic Procedures

by

Lindsay Maffei-Almodovar

Advisor: Peter Sturmey

Direct care staff members serving people with intellectual and developmental disabilities are often required to implement several behavior analytic procedures with only limited training soon after being hired. Pyramidal training is an effective model for disseminating applied behavior analytic skills to employees that treat individuals with developmental disabilities. This study used a multiple probes design across teachers and a delayed multiple baseline design across teaching assistants to evaluate the effects of video models, role play and feedback on teachers’ accuracy in implementing behavioral skills training and on teaching assistants’ accuracy in implementing five applied behavior analytic procedures (i.e. stimulus-stimulus pairing, multiple stimulus without replacement preference assessment, mand training, discrete trial teaching, and graphing discrete trial data). Pyramidal training was effective in increasing first tier participants’ procedural integrity of behavioral skills training steps and in increasing second tier participants’ procedural integrity of implementing the target procedures. First tier participants required feedback to maintain training skills over time, to train procedures other than the procedure implemented during their own training and to train novel staff members. Thus, pyramidal BST required ongoing supervision by a behavior analyst to effectively disseminate multiple ABA skills to a variety of staff members over time.
ACKNOWLEDGEMENTS

Thank you to my husband, Rafael, who tirelessly supported me throughout this process.

Thank you to my daughters, Ivy and Violet, for being my little lights at the end of a very long tunnel.

Thank you to my advisor, Peter Sturmey, who has inspired me more than he could possibly know to keep helping as many people as I can as quickly as possible.

Thank you to my current and former committee members, Emily Jones, Joshua Jessel, Daniel Fienup, Alicia Alvero, Florence DiGennaro Reed and Nancy Neef for their invaluable feedback and guidance.

Thank you to my friend and mentor Gina Feliciano for believing in me even during the times that I almost stopped believing in myself.

Thank you to the amazing staff and students at all of the QSAC schools for participating in this process with me and for being an incredible research team. You all make me so proud every single day.
# Table of Contents

Title Page..................................................................................................................i

Copyright Page............................................................................................................ii

Approval Page.............................................................................................................iii

Abstract.........................................................................................................................iv

Acknowledgements.....................................................................................................v

Table of Contents.......................................................................................................vi

List of Tables...............................................................................................................vii

List of Figures..............................................................................................................viii

Introduction..................................................................................................................1

Method..........................................................................................................................9

Results..........................................................................................................................32

Discussion....................................................................................................................35

Tables..........................................................................................................................41

Figures.........................................................................................................................50

Appendices..................................................................................................................58

References...................................................................................................................62
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>Conditioning Instructor Task Analysis</td>
<td>41</td>
</tr>
<tr>
<td>Table 2</td>
<td>Preference Assessment Task Analysis</td>
<td>42</td>
</tr>
<tr>
<td>Table 3</td>
<td>Mand Training Task Analysis</td>
<td>43</td>
</tr>
<tr>
<td>Table 4</td>
<td>Discrete Trial Teaching Task Analysis</td>
<td>45</td>
</tr>
<tr>
<td>Table 5</td>
<td>Discrete Trial Graphing Task Analysis</td>
<td>46</td>
</tr>
<tr>
<td>Table 6</td>
<td>Behavioral Skills Training Task Analysis</td>
<td>47</td>
</tr>
<tr>
<td>Table 7</td>
<td>Sequence of Participant Sessions</td>
<td>48</td>
</tr>
<tr>
<td>Table 8</td>
<td>BST Details for First Tier Participants</td>
<td>49</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page Number</td>
</tr>
<tr>
<td>---------</td>
<td>--------------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Figure 1</td>
<td>Pyramid of Training</td>
<td>50</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Procedural Integrity</td>
<td>51</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Percentage Correct BST Tasks</td>
<td>52</td>
</tr>
<tr>
<td>Figure 4</td>
<td>Number Correct BST Tasks</td>
<td>53</td>
</tr>
<tr>
<td>Figure 5</td>
<td>All Second Tier Procedures</td>
<td>54</td>
</tr>
<tr>
<td>Figure 6</td>
<td>Trained Second Tier Procedures</td>
<td>55</td>
</tr>
<tr>
<td>Figure 7</td>
<td>Improvement Rate Differences</td>
<td>56</td>
</tr>
<tr>
<td>Figure 8</td>
<td>Social Validity Responses</td>
<td>57</td>
</tr>
</tbody>
</table>
Introduction

Direct care staff members serving people with intellectual and developmental disabilities are often required to implement several behavior analytic procedures with only limited training soon after being hired. Often, agencies that employ direct care staff have few qualified staff trainers, high rates of absenteeism and frequent staff turnover (Seavey, 2004; Zaharia & Baumeister, 1979). These factors necessitate that agencies use effective training models to ensure the competence of their employees in applying behavior analytic methods. During pyramidal training, clinical supervisors, consultants or other experienced behavior analysts train a first tier of direct care workers in agency-required skills and then train these direct care workers to implement training packages targeting the same skills for a second tier of trainee co-workers (Jones, Fremouw & Carpes, 1977; Page, Iwata & Reid, 1982).

The Pyramidal training model has several advantages. It reduces the burden on clinical supervisors of having to train every newly hired staff member and eliminates the need to hire additional staff trainers and consultants (Finn & Sturmey, 2009; Jones et al. 1977; Nigro-Bruzzi, 2010; Page et al. 1982; Schlosser, Walker & Sigafos, 2006; Shore, Iwata, Vollmer, Lerman & Zarcone, 1995). Pyramidal training also allows staff members to train their peers in the general teaching environment (with materials and students that will remain with the staff member post training) which aids in the maintenance and generality of acquired skills (Parsons, Rollyson & Reid, 2013; Pence, St. Peter & Tetreault, 2012; Van den Pol, Reid & Fuqua, 1983). The low cost of pyramidal training makes the model an ideal choice for agencies employing large numbers of frequently changing employees.

Pyramidal training is well evidenced as an effective training model. Andzik and Cannella-Malone (2017) reviewed the efficacy of pyramidal training in a systematic review of 14
articles. The authors used visual analysis to evaluate changes in level, trend, and stability across study data to evaluate if experimental effects were obtained or absent in a given study and whether effects were replicated within a given study. Pyramidal training improved the performance of 100% of first tier participants and 83% of second tier participants.

Behavioral Skills Training (BST) is an evidenced based staff training package. Behavioral skills training (BST) generally includes four components: (a) instructions, (b) modeling (c) rehearsal and (d) feedback. Maffei-Almodovar and Sturmey (2018) calculated percentages of non-overlapping data (PND) and improvement rate difference (IRD) effect sizes to evaluate the efficacy of 18 different types of training in improving instructor implementation of discrete trial teaching (DTT), preference assessment, and mand training. PND and IRD are two means of calculating the effect sizes of small n studies. PND is a calculation of the percentage of data points in an intervention phase that do not overlap with data points in the baseline phase and are in the desired therapeutic direction (Scruggs & Mastropieri, 1998). PND calculations of 90% or above indicate highly effective treatments. IRD is a calculation of the difference between the baseline and intervention proportions of improved to unimproved data points (Parker & Vannest, 2009). IRD calculations between .9 and 1 indicate large effects. The results of the meta-analysis identified BST as the best evidenced training model for disseminating the included behavior analytic procedures, followed by pyramidal BST. The mean PND for BST effects was 98.8% (range 27-100%; highly effective) and the omnibus IRD was .97 CI95 (.96, .99; large effects). The mean PND for pyramidal BST was 98.8% (range 83.3-100%; highly effective) and the omnibus IRD was .97 CI95 (.96, 1; large effects).

The modeling and feedback components of BST are particularly effective in increasing staff accuracy of implementation of behavior analytic procedures. Brock and Carter (2017)
evaluated the efficacy of the components included in training packages to improve staff implementation of interventions for students with disabilities. The authors evaluated group experimental designs using meta-regression analysis. Modeling and performance feedback were the two package components associated with improved trainee implementation of interventions regardless of the intervention targeted for training. Other components included in the review included instructions, rationale for the procedures, and fidelity checklists. In an earlier systematic review, Brock and Carter (2013) reviewed the prevalence of components of training packages and determined that modeling and performance feedback were common components for training paraprofessionals to implement educational practices.

Pyramidal training is versatile and effective for disseminating a variety of behavior analytic procedures. For example, Jones et al. (1977) trained three elementary school teachers to use a classroom management package and then to conduct role playing and give corrective feedback in order to train 12 other elementary school teachers which decreased student problem behavior.

Pyramidal training has also improved the application of DTT procedures. Page et al. (1982) used written instructions, verbal instructions, discussion, rehearsal, verbal feedback and graphic feedback to train three institutional supervisors to discriminate correct teaching behaviors by direct care staff and to instruct, prompt, and praise those behaviors. Improvements in 45 direct care staff members’ teaching behavior were a function of training and feedback provided to their supervisors.

Shore et al. (1995) used instructions, rehearsal and feedback to train one supervisor to implement client behavior treatment programs, to train seven direct care staff members in
program implementation, and to monitor the programs. Staff implementation of client treatment procedures and client behavior improved as a result of supervisor training.

Mand training has also been improved as a function of pyramidal training. Schlosser et al. (2006) used a one day workshop consisting of presentation, modeling, coaching, rehearsal and feedback to train three direct care staff members to provide opportunities for communication to children with developmental disabilities who were non-vocal by a) identifying communication environments, b) selecting vocabulary and symbols for different activities, c) creating opportunities for requesting by using missing-item, interrupted-chain, and delayed assistance procedures, and d) using delayed prompting. These three staff members then trained five additional staff members in the above procedures. Staff members increased the number of opportunities for communication provided to their students as a result of the pyramidal training and student prompted and unprompted communications increased.

Pyramidal training has also been used to increase positive social interactions between direct care workers and clients. Finn and Sturmey (2009) used peer-to-peer BST to train four habilitation specialists to train four of their co-workers to provide positive vocal, gesture, manual sign and physical interactions to adults with psychiatric disorders and developmental disabilities in a day habilitation setting and increased the proportion of positive interactions between trainees and clients.

Pence et al. (2012) conducted two experiments to evaluate the use of pyramidal training to increase accurate staff implementation of multiple stimulus without replacement (MSWO), paired stimulus, and free operant preference assessments. In their first experiment, the authors used written and verbal instructions and a question/answer session to train three teachers (first tier) to use instructions, modeling, role play, and feedback to train implementation of the three
preference assessments. The three teachers then trained five teachers and one clinician (second tier) in the preference assessments. All six trainees acquired the preference assessment skills within a 60-90 minute training session with a confederate and then used the assessments correctly with students in their classrooms. In the second experiment, five of the trainees from the first experiment (second tier) trained an additional 18 teachers (third tier) in the preference assessments using the same trainer and trainee procedures described for the first experiment. Pyramidal training increased the accuracy of implementation of the MSWO, paired-choice, and free-operant preference assessments for the 18 third tier teachers trained during the second experiment. These participants also implemented the preference assessments correctly with students outside of the training context.

Parsons et al. (2013) used BST consisting of instructions, modeling, role play with a confederate, and feedback to train three groups of first tier practitioners to use BST to improve second tier staff members’ use of embedded teaching strategies and to conduct MSWO preference assessments. Each group of participants improved their implementation of the BST package following participation in the training program. In addition, the first tier trainers used BST correctly to train second tier staff in procedures not targeted during trainer training, including most-to-least prompting, backward chaining, teaching manual signs, graduated physical guidance, giving feedback, least-to-most prompting, and teaching jewelry making. All staff trained by the participants improved their performance of the target skills from pre to post training assessments and used the skills acquired during role play sessions correctly with their classroom students.

There are limitations to these studies. First, no studies presented procedural integrity data for experimenter implementation of pyramidal training. Second, only one study presented social
validity data (Parsons et al., 2013). Third, none of the studies presented individual session-by-session data for first or second tier participants’ correct implementation of staff training. Fourth, no studies included a visual graphic display to present individual first or second tier participants’ training performance.

Pence et al., (2012) presented procedural integrity data as overall means for first and second tier trainers’ correct use of delayed feedback during second and third tier training, but did not present data on first or second tier participants’ use of other components of training (e.g. modeling). The authors presented post training procedural integrity data only for first and second tier trainers and, therefore, did not manipulate correct training implementation as a dependent variable. Another limitation to the Pence et al., (2012) study was the range of procedural integrity with which second tier trainers trained third tier trainees. The reported procedural integrity mean for the first tier participants during second tier training was 95% (range, 91% to 98%). The reported mean procedural integrity for second tier trainers during third tier training was 83%, but the range of integrity varied from 53% to 99%. The authors reported that the low procedural integrity could be attributed to one second tier participant’s failure to provide praise for correctly performed steps, although the participant provided corrective feedback after 100% of errors.

Subsequently, Parsons et al. (2013) addressed a weakness of Pence et al., (2012) in that the authors manipulated first tier accuracy of BST as a dependent variable and reported collecting data on trainers’ correct use of eight total steps of BST instead of only on the use of delayed feedback. The authors presented group mean performance data for first tier participants in a multiple baseline across groups display. The average performances per session of groups 1, 2 and 3 comprised the first, second and third legs of the multiple baseline design. Therefore, one
could not determine through visual analysis of the figure whether there was a functional relation between BST and individual participants’ correct use of BST. There were also limitations in the procedures as first tier participants did not have access to any information about BST, such as instructions or a task analysis, during first tier baseline measurement. Also, second tier participants did not have access to any information on the procedures targeted for training during second tier baseline measurement and only accessed the relevant instructions during training. Therefore, post-training measurement for each tier may represent changes in performance related to access to instructions, rather than the full BST package. Although instructions are typically not a sufficient training component, they can improve performance and produce an exaggerated effect of training when absent from baseline (Maffei-Almodovar & Sturmey, 2018; Ward-Horner & Sturmey, 2012).

Parsons et al. (2013) was also limited in terms of the procedural integrity with which first tier participants implemented BST and in assessment of maintenance of BST and the generality of participant-implemented BST to train varied procedures and staff members. The authors presented overall session means for each first tier participant’s implementation of BST for baseline and post-training and the percentage of accuracy for one “on-the-job” probe per participant. Overall post-training means ranged from 71% to 100% and, similar to Pence et al. (2012), individual participant ranges presented evidence of some occurrences of low procedural integrity during post training (range, 50% to 100%). Importantly, experimenters provided feedback to trainers on their performance after every post-training session. Therefore, trainer performance data did not demonstrate the effects of the passage of time after the removal of feedback during post-training measurement. There were also limited data on first tier participants’ correct use of BST when training actual staff because baseline and post-training
sessions consisted only of role play. Each first tier participant implemented BST with only one staff member for one procedure during single on-the-job probes to determine maintenance of the skill and generality to untrained procedures and staff. On-the-job probes for percentage of accuracy of BST implementation ranged from 88%-100%. However, first tier trainers selected the staff member and procedure they trained during the on-the-job probes which may have limited the possibility of low trainer procedural integrity due to a difficulty with a particular staff member or procedure. Therefore, the purpose of the current study was to evaluate to what extent classroom teachers’ (first tier) implementation of a BST package (including video modeling, role play and feedback) would increase teaching assistants’ (second tier) accuracy in performing five behavior analytic procedures.

The current study addressed the weaknesses in the pyramidal training literature. First, the authors of the current study evaluated the correct implementation of a 14 step BST package by first tier participants as a dependent variable and presented individual first tier participant performance data in a multiple probes across participants graphic display. Second, the current study included extensive follow-up data collection for first tier participants with a minimum interval of four months from the last session of first tier training to the last session of first tier follow-up data collection. Third, first tier participants had to implement BST training for four target procedures that were all different than the procedure for which they implemented BST during training. Fourth, all first tier participants demonstrated BST with at least two different staff members during the follow-up phase. Fifth, in order to control for the effects of instructions, all first tier participants had access to the instructions for BST and all second tier participants had access to the instructions for target procedures during baseline measurement as well as during training and follow-up. Finally, the current study also extended previous research because the
second tier evaluated whether pyramidal training was sufficient to improve trainee performance of stimulus-stimulus pairing, MSWO preference assessment, mand training, DTT, and percentage graphing. Pyramidal staff training research is limited in this aspect, as only Pence et al. (2012) targeted more than one procedure per second tier participant. The author selected the first four procedures because they are evidence-based interventions that have been established to increase skill acquisition and decrease problem behavior (NSP, 2009; The National Professional Development Center, 2010) and most ABA educational programs use them in order to ensure effective instruction for individuals with developmental disabilities. The author targeted skill 5 because behavior analysts are required to monitor student progress through the frequent visual analysis of graphic data displays (BACB, 2014).

The current study also included IRD effect size calculations to evaluate whether pyramidal BST was effective in increasing second tier participant performance of the above targeted behavior analytic procedures. Finally, the study included procedural integrity measures for the experimenter and first tier trainers during all BST implementation and measures of social validity for all participants. The BST package for the first and second tiers included video modeling in place of live models during BST implementation because it has been shown to be an effective method of training staff to implement behavior analytic procedures when combined with performance feedback, requires less expertise to implement than a live model and simplifies instruction standardization.

Method

Participants

Three ABA classroom teachers served as first tier participants and three teaching assistants served as second tier participants. The teacher participants were referred to as Anne,
Betty, and Carol and the teaching assistant participants were referred to as Xylia, Zoe, and Yolanda throughout the experiment.

Inclusion criteria for first tier participants dictated that each participant a) held a master’s degree, b) had at least one year of teaching experience in an ABA setting, and c) demonstrated competence in the five target ABA procedures prior to baseline measurement in BST. First tier participants would be excluded only if they demonstrated greater than 65% mean accuracy in BST during baseline measurement. Participants Anne, Betty, and Carol had been working as classroom teachers at the study site for three years, two years and four months, respectively, at the start of the study. All three teachers were enrolled in BCBA certificate course tracks at the time of the study. Anne, Betty, and Carol taught in classrooms with students aged 19-21 years, 16-19 years and 8-11 years, respectively. Prior to the start of the study all participants had been trained and demonstrated competency in the five target behavior analytic procedures. All first tier participants demonstrated less than 65% mean accuracy in implementing BST for the targeted ABA procedures during baseline measurement. Therefore, no first tier participants were excluded from participation as all met the criterion required to receive training.

Inclusion criteria for second tier participants dictated that each a) held a high school diploma, b) had little or no training in ABA and c) would be excluded from participation only if they performed all five target ABA procedures at above a 65% mean percentage correct during baseline measurement. Xylia and Yolanda both possessed high school diplomas and had been working as teaching assistants at the study site for two weeks and three years, respectively at the start of the study. Zoe possessed a Bachelor of Arts degree and began working at the study site two months after the start of the study. Xylia, taught in a classroom with students aged 17-19 years, while Yolanda and Zoe both taught students aged 19-21 years. All three second tier
participants demonstrated less than 65% mean accuracy in implementing at least one of the targeted ABA procedures during baseline measurement. Three additional second tier participants had initiated participation in the study but dropped out during baseline measurement. One participant resigned due to child care difficulties; another was hired from a waiting list by the city department of education and a third transferred to a different school within the same agency.

Teaching assistants who were not second tier participants volunteered to receive training in a targeted skill when a second tier participant did not require training in the skill during baseline, training, and follow-up sessions. None had received prior training in the skill that they volunteered to learn. Figure 1 presents the pyramid of training for the total number of staff members involved in the study and the pyramid of training for skills disseminated to the first and second tier participants. Anne trained Yolanda and one additional staff member, Betty trained Zoe and four additional staff members and Carol trained Xylia and five additional staff members during the course of the study. Volunteer staff members acted as confederate students during role play sessions for first tier participants. The experimenter acted as the confederate trainee for first tier participants during role play sessions and as the confederate student during all second tier participant role play sessions, except for two baseline sessions for Zoe (one stimulus-stimulus pairing session and one vocal mand training session) during which Carol acted as the confederate student.

Setting

The study took place in a school for students with Autism spectrum disorders aged 5 to 21 years during school hours when students were present, before students arrived in the morning and after students had left for the day in either the participants’ regular classrooms or in other common areas of the school (e.g. the administrative office, another teacher’s classroom, etc.).
Other staff members and students that were not part of the study were sometimes present during sessions. Sessions to allow for maintenance and generality measurement took place either in the participants’ regular classrooms or in another classroom (if the targeted procedure was not appropriate for students in their classroom).

Classrooms contained tables, chairs, shelves, academic and leisure materials, and computers. Certain classrooms also contained additional electronic equipment as determined by student need such as Smartphones, tablets, or augmentative and alternative communication (AAC) devices. Classroom staff and students not involved in the study were also present during sessions. Each classroom contained up to six students and up to four staff members, including one teacher and three teaching assistants.

It is important to note that this study took place in a functioning ABA school program in which the experiment was secondary to the regular activities of the participating staff members. Regular administrative meetings along with impromptu meetings related to staff or student performance took precedence over meetings related to the experiment. Staff members were often not able to meet with the experimenter for longer than 15 min at a time on a given day and were not available to meet every day. Therefore, this experiment took place over nine months, a time period longer than one might see in other studies on pyramidal training in settings where research and practice obligations overlap.

**Materials**

The experimenter and Anne, Betty, and Carol used a computer, tablet, or Sharp Board® to play video models; a Samsung® phone or tablet to video record all participant sessions; and datasheets and task analyses printed on standard printer paper (21.59 cm by 27.94 cm). Tables 1-6 present the task analyses for all targeted ABA procedures and for BST. Participants also used
any academic and leisure materials, electronic equipment or student-preferred items necessary to complete instructional procedures. For example, when conducting a MSWO preference assessment, the staff member used five preferred items which may have included toys (e.g. a ball, a car or a doll), edibles (e.g. M&M's, sunchips or apple slices), electronic devices (e.g. an iPad®, or iPhone®) or other items (e.g. a string, a marker or a book). The experimenter also used scripts printed on standard printer paper (21.59 cm by 27.94 cm) for sessions that required a confederate trainee and/or a confederate student during baseline and training sessions. Confederate trainee scripts consisted of the task analysis for a given targeted behavior analytic procedure and included between two and five highlighted programmed errors to be made by the trainee. The student scripts detailed how often the confederate student had to respond correctly, make an error, or engage in a particular problem behavior during the procedure targeted for training (a full explanation of the confederate scripts is given on page 24).

**Experimental Design**

In a multiple probes design (MPD), experimenters measure the percentage of accuracy for a given task sequence intermittently during baseline, continuously during intervention, and then intermittently during the follow-up phase (Horner & Baer, 1978). When experimenters apply the independent variable to a given leg of the MPD, baseline measures continue in subsequent legs to serve as controls. Visual analysis of a MPD demonstrates experimental control if only the application of the independent variable corresponds with behavior change in the first leg and across at least 2 replications. The first tier of this study used a MPD across participants because continuous measurement of the dependent variables during baseline and follow-up was impractical.
A delayed multiple baseline design (MBD) allows the addition of new participants as a study progresses because experimenters can add subsequent baselines after the start of an initial baseline and intervention (Heward, 1978). Delayed MBD is limited when compared with a full scale MBD, because one cannot verify predictions for earlier legs of the design according to subsequent baselines begun after the application of interventions in earlier legs. Subsequent legs, therefore, serve only as replications of the initial leg. The second tier of this study used a delayed MBD across participants because frequent turnover of newly hired staff members at the study site made a full scale MBD impractical.

The first and second tier designs both included four parts: 1) baseline, 2) training, 3) generality and 4) follow-up.

**Dependent Variables**

For first tier participants, the dependent variables were the percentage of correctly performed BST steps per session and the total number of correct BST tasks compared with the total BST task opportunities for a given session. During each first tier participant session, the experimenter observed the percentage of correct BST steps and the number of correctly performed BST tasks and the total number of BST task opportunities. A correct step/task was a step/task performed correctly and in the correct order according to the task analysis for the BST package. An incorrect step/task was a step/task that deviated from the task analysis in any way other than a warranted omission. Warranted omissions occurred because the BST package did not require each of its steps to be completed during every session. For example, trainers only had to show video models to trainees during the first session of a given training package. Additionally, trainers only had to perform steps related to feedback and review of errors when
trainees made errors while performing target procedures. If a trainee performed a target procedure at 100% correct, the trainer did not need to perform these steps. The experimenter calculated the percentage of correct steps for each first tier session by dividing the number of correct steps by the total number of steps during each session and multiplying by 100. For example, if the participant performed 8 correct and 2 incorrect steps when all 10 steps of the task analysis were required according to staff or student performance, his or her percentage of correct steps for the session was 80% (i.e., 8/([8+2])*100). If the participant performed 8 correct and 1 incorrect steps and correctly omitted 1 step, the correctly omitted step was not applicable (NA) and counted as correct and his or her percentage of correct steps for the session was 90% (i.e., (9/([9+1])*100). If the participant performed 8 correct and 1 incorrect steps and incorrectly omitted 1 step, his or her percentage of correct steps for the session was 80% (i.e., (8/([8+2])*100).

The experimenter determined the number of total BST task opportunities for a given session by counting the number of trainer response opportunities for each BST step and adding these numbers to equal a sum total. The number of BST tasks that a trainer could demonstrate varied according to the numbers of steps in the targeted procedure and the trainee errors made during a given session. Most steps of the BST task analysis allowed the trainer to perform a stable number of BST tasks per session. Specifically, steps 1-7, and 13 included only one BST task whereas steps 12 and 14 required two BST tasks, regardless of the number of steps required for the target procedure or the number of errors made by the trainee during a session; however, steps 8-11 varied in the number of BST tasks required per session based on the number of steps in the procedure and the number of errors that the trainee made. Therefore, the number of correctly performed BST tasks compared with the total number of BST task opportunities
presented during a given session varied according to the target procedure and trainee performance.

For second tier participants, the dependent variable was the percentage of correctly performed steps of: 1) stimulus-stimulus pairing (conditioning the instructor as a secondary reinforcer), 2) MSWO preference assessments, 3) mand training, 4) DTT, and 5) Graphing DTT data. During each second tier participant session, the experimenter observed the percentage of correct steps of one targeted behavior analytic procedure. A correct step was a step performed according to the task analysis for the targeted ABA procedure. An incorrect step was a step that deviated from the task analysis in any way other than a warranted omission. A warranted omission included any instances where a second tier participant correctly omitted a conditional step based on student performance. For instance, during a MSWO preference assessment, one of the steps dictated what to do if a student did not choose an item from the array after being given the instruction to "pick one." If the student consistently picked an item when given the instruction, the second tier participant did not need to perform this step.

The experimenter calculated the percentage of correct steps for each second tier session by dividing the number of correct steps by the total number of steps during each session and multiplying by 100%. For example, if the participant performed 8 correct and 2 incorrect steps when all 10 steps of the task analysis were required according to staff or student performance, his or her percentage of correct steps for the session was 80% (i.e., \((8/(8+2))\times100\)). If the participant performed 8 correct and 1 incorrect steps and correctly omitted 1 step, the correctly omitted step was marked as not applicable (NA) and counted as correct and his or her percentage of correct steps for the session was 90% (i.e., \((9/(9+1))\times100\)). If the participant performed 8
correct and 1 incorrect steps and incorrectly omitted 1 step, his or her percentage of correct steps for the session was 80% (i.e., \(\frac{8}{8+2}\ast 100\)).

**Tier 1 Procedure**

**Training in target skills.** Prior to the onset of baseline and training sessions for the second tier participants, the experimenter assessed the first tier participants’ performance on all ABA procedures targeted for second tier participant training with students and student data in their classrooms to ensure that the first tier participants performed each procedure at criterion. Tables 1-5 present the criteria for mastery for each procedure. If the first tier participant made an error during any of the procedures, the experimenter provided immediate feedback by stating the correct performance requirement for the incorrectly performed step according to the task analysis (e.g. “be sure to say 'pick one' to begin each trial”). Rehearsal and feedback sessions repeated for each target skill until the first tier participant met the criterion for mastery.

**Baseline for BST.** During baseline, the experimenter observed each first tier participants’ correct implementation of the BST package by asking each first tier participant to implement it for a targeted procedure with the experimenter acting as a confederate trainee, and a staff member acting as a confederate student (see page 24 for full explanation of confederate scripts). Each baseline session consisted of the first tier participant implementing the necessary steps to complete one BST session. The initial BST session for a given procedure always required the first tier participant to complete all 14 steps of BST. All subsequent sessions of the BST package for the same procedure only required the first tier participant to complete steps 6-14 of the package (roleplay and feedback or rehearsal and feedback). During all baseline sessions, the experimenter provided the first tier participant with a typed task analysis for the BST package. During the first baseline session for BST in a given procedure, the experimenter said, “Please
read these instructions and tell me when you are finished.” During subsequent sessions, the experimenter always gave the instructions to the participant and allowed the participant time to read the instructions; however, the participants sometimes elected not to read the instructions. When the first tier participant told the experimenter that he or she was finished or elected not to read the instructions, the experimenter said “Based on the instructions, please implement behavioral skills training to the best of your ability.” The experimenter answered questions regarding BST from the first tier participant during baseline only by referring to the text of the instructions.

Sessions to assess generality also took place during baseline and included untrained staff members and students in the school. The procedure for generality sessions was the same as in other baseline sessions, except that the first tier participant performed the training with staff members that were not participants and were not previously trained in the target procedure instead of confederate trainees. Generality sessions also included students from the school during the final rehearsal and feedback component of the target procedure instead of confederate students.

A random number generator determined the sequence that the first tier participant implemented each training package during baseline and whether the session included confederate trainees and students, or actual trainees and students. The first tier participants trained different procedures in different orders according to the random assignment in order to control for confounds of time related to passage of the school year and increased experience in the study. Sessions including confederate trainees and confederate students for the training package for each procedure were assigned numbers 1-5: 1) stimulus-stimulus pairing, 2) MSWO, 3) mand training, 4) DTT, and 5) graphing. Sessions including actual trainees and students for the training
package for each procedure were assigned numbers 6-10: 6) conditioning the instructor, 7) MSWO, 8) mand training, 9) DTT and 10) graphing. These numbers were input into the online random number generator The Research Randomizer™. The generator output was used to order the sequence of sessions during baseline using block randomization to ensure that each procedure appeared only once for each participant. Table 7 presents the sequence of baseline sessions for each first tier participant. Anne, Betty, and Carol completed 8, 15 and 15 baseline sessions, respectively.

**Training in BST.** During training, the experimenter used BST to train the first tier participants to use BST to train a trainee in one randomly selected procedure. The experimenter entered the numbers 1-5 into the Research Randomizer™ for each first tier participant and trained the participant in the procedure number output by the generator. Constrained random selection ensured that the experimenter trained each first tier participant to implement BST for a different procedure. Table 8 presents the details of training for each first tier participant including the procedure targeted during BST training, the length of the video model, the errors made by the confederate trainee or volunteer staff member trainee, and the total training times. The experimenter trained Anne to implement BST for MSWO, Betty to implement BST for stimulus-stimulus pairing, and Carol to implement BST for mand training.

The first component of BST training was video modeling. The participant viewed a video model of a trainer using BST to train a trainee in the selected ABA procedure. The task analysis for BST is presented in Table 6 and is the same for each ABA procedure. In the video models, trainers completed BST in three sessions. The first session included steps 1-14 of the task analysis and required trainers to deliver the task analysis for the procedure to the trainee, show the video model of the target procedure, observe the trainee practicing the procedure with a
confederate student, deliver immediate feedback for errors and then review errors and progress toward criterion with the trainee. The subsequent sessions included only steps 6-14 of the task analysis and required the trainer to observe the trainee practicing the procedure with either a confederate or actual student, deliver immediate feedback and then review errors and progress toward criterion with the trainee.

During the video modeling component of training, the experimenter gave the BST task analysis to the first tier participant and instructed her to watch the video model and check off the steps in the task analysis as she saw them performed in the video. The experimenter then inspected the task analysis to see if the trainee checked all relevant steps. If the trainee did not check all relevant steps, the experimenter stated the step requirement of any missed steps, referred to the part of the video model where the trainer demonstrated the step, and played the relevant part of the video model, if necessary. Once the trainee indicated that she saw the step performed in the video, the experimenter directed her to insert a check for the step on the task analysis.

The second component of BST was role play. The first tier participant practiced implementing the BST package to train a confederate trainee in the selected procedure. The experimenter acted as the confederate trainee and a volunteer staff member acted as the confederate student. The first tier participant practiced implementing the training package until he or she met criterion for mastery, 87.5% correct implementation (28 correct of 32 total steps) across the implementation of two consecutive BST packages. Each BST package included three sessions. For the role play of the video modeling component, the first tier participant practiced a) showing the video model for the target procedure to the confederate trainee, b) providing the correct task analysis for the targeted procedure, c) instructing the confederate trainee to check off
the steps of the targeted procedure on the task analysis as he or she saw them performed in the video model, d) ensuring that the video model was playing before walking away from the confederate and e) visually inspecting the task analysis to ensure that the trainee checked off all of the relevant steps for a targeted procedure.

For role play of the role play and feedback component, the first tier participant practiced a) observing the confederate trainee implementing the targeted procedure, b) scoring his or her performance on the relevant task analysis, c) providing immediate feedback for any errors that the trainee made while implementing the procedure, and d) reviewing errors and progress toward criterion. The experimenter followed a script while acting as the confederate trainee implementing the procedure selected for training. The script indicated 2-5 steps that the confederate trainee performed incorrectly while implementing the procedure. The Research Randomizer™ determined errors to be performed by the confederate trainee prior to each session. The experimenter entered numbers 2-5 into the randomizer and the output determined the number of errors to be made for the session. The experimenter then entered the number of steps for the target procedure and the number of errors to be made and the output determined which specific steps the confederate trainee would perform incorrectly during the session. The staff member acting as a confederate student also followed a script. The script indicated errors or interfering behavior that the confederate student would emit during the selected procedure (e.g. the student often attempts to select two items instead of one during the MSWO preference assessment). For the confederate students, the experimenter created three scripts for each of the five procedures. Each script contained three aspects of student behavior that varied (e.g. for stimulus-stimulus pairing, the number of times the student left the instructional area, the number of times the student approached the instructor and the number of times that the student engaged
in problem behavior all varied according to the student script selected). The experimenter entered the available number of student scripts for a given procedure into the random number generator to identify the script that the confederate student followed during each role play and feedback repetition and constrained randomization so that steps selected for the staff member to perform incorrectly and scripts selected for the confederate student were different for each role play and feedback repetition. The experimenter allowed the first tier participant to review the student scripts during role play sessions, but not during the final session of a given BST package as the final session was meant to simulate training a staff member with an actual student.

The fourth component of BST was feedback. The experimenter provided immediate feedback to the first tier participant for any errors that she made during role play sessions. Feedback consisted of the experimenter stating the step, or a part of the step from the BST package task analysis that the first tier participant either omitted or performed incorrectly. For example, the first step of the task analysis for the MSWO preference assessment states that the trainee should place five items in a row on a table or on the floor in front of the student and remove all other potentially distracting items. If the confederate trainee correctly placed five items on the table in front of the student, but left another toy close by, the first tier participant should have provided immediate feedback for this error by saying, “Be sure to remove all potentially distracting items that you are not using for the preference assessment from the area.” If the first tier participant did not provide this feedback within 10 sec of the trainee error, the experimenter provided immediate feedback to the first tier participant by referring to step 8 of the task analysis for BST (e.g. “Be sure to give immediate feedback when the trainee makes an error”) and also referring to the step of the selected procedure that the
confederate trainee erred on or omitted (e.g. “The trainee should remove potentially distracting items from the preference assessment area.”)

To deliver feedback on the first tier participant’s scoring of the confederate trainee’s performance the experimenter calculated the percentage of agreement of the first tier participant’s scoring of the trainee with the confederate trainee script and stated the percentage of agreement with the first tier participant. Step 13 of BST required 80% agreement between the trainee performance data collected by first tier participant and the confederate script.

**Generality to staff.** Once a first tier participant met criterion for implementing a training package for a particular procedure with a confederate trainee, the experimenter directed her to train a volunteer staff member not previously trained in the procedure and to include a student in the final rehearsal and feedback session for the trainee. The experimenter provided immediate feedback for any errors that the first tier participant made while training the staff member. If the first tier participant implemented the BST package with the staff member with at least 80% accuracy, training concluded. All three first tier participants performed the BST package with a staff member with at least 80% accuracy to conclude training.

**Follow-up/Generality across skills/trainees.** The experimenter observed the first tier participants’ BST implementation periodically for a minimum of four months after training in BST. The experimenter observed Anne for five months, and Betty and Carol for four months following training. During follow-up, the experimenter assessed whether improvements in the implementation of the BST package maintained over time and whether participants implemented BST correctly for procedures other than those targeted during their training and with novel trainees. During follow-up sessions, the experimenter asked the first tier participant to implement BST for one of the remaining four behavior analytic procedures with either a volunteer staff
member or with a second tier participant and observed the accuracy with which the first tier participant implemented the BST package. For each follow-up session, if the first tier participant implemented the BST package with errors on four or fewer steps for the targeted procedure (10 of 14 correct steps for initial BST sessions, or 5 of 9 correct steps for subsequent sessions), the experimenter did not provide any feedback and continued to the next session until the BST package for the selected procedure was complete. If the first tier participant made errors on more than four steps during any session, the experimenter conducted a feedback session prior to conducting any additional sessions. Feedback sessions lasted 5-10 minutes, during which, the experimenter met with the first tier participant and provided the first tier participant with her scored BST task analysis and the scored task analysis of the selected procedure for her trainee. The experimenter then verbally reviewed all of the errors performed by both the first tier participant and her trainee. The feedback sessions provided during follow-up measurement were not part of the initial research plan for this study, but were added when the experimenter observed decreases in first tier participants’ performances in order to prevent compromised procedural integrity of the independent variable for the second tier of the study. The experimenter ordered the sequence of the procedures observed during follow-up using the Research Randomizer™ as was done for baseline sessions. Table 7 presents the sequence of follow-up sessions for each first tier participant.

Tier 2 Procedure

Baseline. During baseline sessions, the experimenter or a first tier participant provided the second tier participant with a typed task analysis for a procedure and the second tier participant implemented the procedure one time. The experimenter or first tier participant said, “Please read these instructions and tell me when you are finished.” When the second tier
participant told the experimenter or the first tier participant that she was finished, the experimenter or first tier participant said “Based on the instructions that you read, please implement procedure x to the best of your ability”. The experimenter or first tier participant then answered questions from the second tier participant during baseline sessions only by referring to the text of the instructions for the procedure. The experimenter or first tier participant video recorded each baseline session by setting up a Samsung® tablet or phone within 1 m of the second tier participant. The experimenter or first tier participant then acted as a confederate student by following a randomly assigned student script for the targeted procedure. First tier participants conducted baseline sessions only with second tier participants that they were not assigned to train. The experimenter also observed sessions conducted with actual students once for each of the five procedures during baseline for each of the second tier participants. These sessions were the same as other baseline sessions, except that the experimenter observed the second tier participant conduct the selected procedure with a student in the classroom instead of a confederate student.

A random number generator determined the order in which the second tier participant implemented the procedures and whether the session included confederate students, or actual students. The experimenter assigned sessions including confederate students numbers 1-5: 1) stimulus-stimulus pairing, 2) MSWO, 3) mand training, 4) DTT, 5) graphing, and assigned sessions including students numbers 6-10: 6) conditioning the instructor, 7) MSWO, 8) mand training, 9) DTT, 10) graphing. The experimenter then input the numbers 1-10 into the Research Randomizer™ and used the output to order the sequence of sessions during baseline using block randomization so that each procedure appeared only once in each block set of five consecutive sessions. Table 7 presents the sequence of sessions for baseline data collection for each second
tier participant. Block randomization ensured that the participant implemented each procedure the same number of times and in varying orders in order to control for the effects of the passage of time and experience in the study.

**Training.** Xylia received training in mand training, DTT and graphing, participant Y received training in conditioning the instructor and DTT, and participant Z received training in DTT. The first training session for second tier participants consisted of the first tier participant delivering the video modeling component (steps 1-5 of BST), and then delivering the role play and feedback components with a confederate student (steps 6-14 of BST). During subsequent sessions, the first tier participant repeated the role play and feedback components until the second tier participant met criterion for role play (two consecutive sessions at the predetermined mastery criterion for the target procedure). The first tier participant then delivered rehearsal and feedback where the second tier participant practiced the skill with an actual student and repeated the component until the second tier participant met the criterion for rehearsal and feedback (one session at the predetermined criterion). Thus, the first tier participant delivered the video model component only once for each procedure, but delivered the components for role play and feedback and rehearsal and feedback continuously until the second tier participant achieved mastery criterion. Criterion for mastery of a given behavior analytic procedure varied according to the number of steps in the procedure, but generally allowed for no more than 1-2 errors on a given procedure across consecutive sessions.

During video modeling, the first tier participant exposed the second tier participant to the video model for a target procedure and instructed her to check off the steps of a given task analysis while watching the video model by inserting a check mark in the same row, but in the column to the right of the relevant step of a given task analysis. The first tier participant then
inspected the task analysis to ensure that the second tier participant inserted a check for all relevant steps of the procedure accurately.

During role play and feedback, the second tier participant practiced implementing the procedure with a confederate student. The second tier participant practiced implementing the procedure during consecutive sessions until meeting the predetermined criterion for mastery for the target procedure. Each session included one implementation of the selected procedure. The first tier participant observed the second tier participant implementing the procedure and provided immediate spoken feedback for any errors that the second tier participant made while implementing the procedure by making a positive statement referring to the step in the task analysis that the trainee either omitted or performed incorrectly within 10 seconds of the error. The first tier participant recorded either a plus or a minus for each step of the task analysis relevant to the procedure according to second tier participant performance during the role-play session. After the second tier participant completed the role play of the procedure, the first tier participant made a positive statement about the second tier participant’s performance, reviewed any errors that the second tier participant made, made another positive statement about the second tier participant’s performance and then informed the second tier participant of whether she had met the mastery criterion for training and what the next training session would entail (e.g. role play or rehearsal with an actual student).

**Generality to students.** Once a second tier participant met criterion for implementing a targeted procedure in role play, the first tier participant asked her to rehearse the procedure with a student. Other than the replacement of a confederate student with an actual student, rehearsal sessions were identical to role play sessions. The second tier participant practiced implementing
the procedure with a student until she met the predetermined criterion for mastery for the target procedure.

**Follow-up.** During follow-up, the experimenter observed the second tier participants implement each of the five target ABA procedures once per week with students. Follow-up sessions were the same as student sessions during baseline. The experimenter observed each second tier participant every week for three weeks (Xydia), four weeks (Zoe) or five weeks (Yolanda) after her last training session. The experimenter ordered the sequence of the procedures observed each week during follow-up using the Research Randomizer™ in the same fashion as in baseline. Table 7 presents the sequence of follow-up sessions for each second tier participant.

**Inter-Observer Agreement**

The experimenter scored all first tier participant sessions during assessment of the targeted behavior analytic procedures prior to the onset of the first tier of the study. One ABA doctoral candidate and one doctoral level behavior analyst also scored 30% of the tier one participants’ sessions during the assessment period by watching video recordings of the sessions to score inter observer agreement (IOA). The experimenter assigned each session a number and used the Research Randomizer™ to select participant sessions for IOA measurement. IOA calculation included dividing the number of agreements by the total number of agreements plus disagreements and multiplying by 100%. An agreement was an instance where both the experimenter and observer marked a plus, minus or NA for a given step of the task analysis for the procedure. A disagreement was an instance where the experimenter and the observer marked different scores for the same task analysis step. Mean IOA for first tier participants’ performance of the targeted behavior analytic skills during assessment was 93.4% (range 86%-100%).
The experimenter scored all first and second tier participants’ sessions during baseline, training, generalization, and all volunteer staff trainee sessions and confederate trainee sessions. One undergraduate psychology student observer also scored 30% of sessions by watching videos of the completed sessions. The experimenter assigned each session a number and used the Research Randomizer™ to determine the sessions that the student observer scored. The mean IOA for first tier participants was 96.5% (range 88.5%-100%) and for second tier participants, confederate staff trainees and volunteer staff trainees was 95.8% (range 82.6%-100%). The mean IOA calculations for conditioning the instructor, MSWO, mand training, DTT and graphing were 96% (range 85.1%-100%), 94.7% (range 86%-100%), 95.4% (range 82.6%-100%), 96.1% (range 86%-100%), and 97.1% (87.5%-100%), respectively.

Procedural Integrity

One undergraduate psychology student observer also scored experimenter implementation of the training procedures for first tier participants for 30% of sessions by watching recorded videos of the sessions. The experimenter assigned each training session a number and used the Research Randomizer™ to determine the sessions that the observer scored. The observer scored experimenter procedural integrity using the BST task analysis. The observer marked a plus for each step correctly performed according to the task analysis, a minus for each step incorrectly performed or omitted and NA for correctly omitted steps. The observer calculated procedural integrity by dividing the total number of pluses by the sum of pluses, minuses and NA’s and multiplying by 100%. The mean procedural integrity for the experimenter’s implementation of BST during first tier training was 93.8% (range 81.3%-96.9%).
The experimenter measured procedural integrity for first tier participants’ implementation of BST during second tier participant training for 100% of training sessions during the second tier of the study. Figure 2 presents first tier participants’ procedural integrity and second tier participants’ accuracy in implementing a target procedure during each session of second tier training. Mean procedural integrity per session for Anne, Betty and Carol in implementation of BST during the second tier participant training was 77.9% (range 44.4%-100%), 92.9% (range 77.8%-100%), and 85.5% (range 66.7%-100%), respectively. The overall mean procedural integrity per session for the second tier of the study was 85.5% (range 44.4%-100%). Mean procedural integrity per BST package for Anne, Betty and Carol during second tier participant training was 79.1% (range 72.8%-85.4%), 90.6%, and 85% (range 78%-94%), respectively. The overall mean procedural integrity per BST package for the second tier of the study was 83.4% (range 72.8%-94%).

The experimenter also measured procedural integrity for script following for 100% of sessions for confederate trainees and confederate students via second tier participant, volunteer staff trainee and confederate trainee performance data. Performance data that the experimenter and independent observer scored for a given trainee session indicated whether a trainee performed correctly according to student behavior. For instance, observers scored a trainee’s performance as a plus during MSWO if she correctly blocked access to items and repeated the trial when the confederate student attempted to choose two items given the direction, “pick one.”(step 9). Observers scored this step as a minus if the trainee allowed access to the items or ended the assessment without repeating the trial. If the confederate student never attempted to select two items, observers scored the trainee performance for this step as ‘NA.’ If the confederate trainee omitted this step or did not perform the step correctly as indicated on the
confederate trainee script and as indicated by a minus on her performance data for the procedure, then the confederate trainee followed the script correctly.

Similarly, if the confederate student script for a particular session indicated that the student should attempt to select two items simultaneously “often” during the MSWO and at least three of five trials of the trainee’s performance data indicated that the confederate student attempted to select two items, then the confederate student correctly followed the script. Mean procedural integrity for confederate trainees and confederate students was 97.3% (range 88.2%-100%) and 97.1% (range 50%-100%), respectively.

Social Validity

The experimenter assessed whether first and second tier participants preferred and found their training to be effective by giving them a survey once at the end of the study. The participants circled the numbers one through five on each of several statements if they “highly disagreed,” “sort of disagreed,” “neither agree or disagreed,” “sort of agreed,” or “highly agreed,” with the statement (e.g. “The videos I watched were useful.”). Appendix A presents the social validity surveys for the first and second tier participants, respectively.

Effect Size Calculation

As in Maffei-Almodovar and Sturmey (2018), the first author used free online software (http://www.vassarstats.net/prop2_ind.html) to calculate IRDs and confidence intervals for the effects of BST on second tier participants’ correct implementation of each target ABA procedure and used WinPepi (http://www.biomedcentral.com/1742-5573/content/1/1/6) to calculate omnibus IRDs (Parker, Vannest & Davis, 2011). The unit of analysis was the target ABA procedure. The omnibus IRDs reflected an aggregation of data points across all procedures trained using BST compared with all procedures left untrained.
Results

Tier 1

All three first tier participants increased their percentage of accuracy of implementation of the BST steps across all of the target ABA procedures from baseline to follow-up (see Figure 3). During baseline, the mean of correctly performed BST steps for Anne, Betty and Carol was 39.1%, 40.1%, and 37.1% respectively. During training the mean of correctly performed BST steps for Anne, Betty and Carol was 85.3%, 87.4%, and 87% respectively. During follow up the mean of correctly performed BST steps for Anne, Betty and Carol was 82.6%, 86.2%, and 86.6% respectively. All tier one participants also implemented the training package across at least two untrained procedures at mastery criterion and only required brief feedback sessions to implement BST for other untrained procedures and with novel staff at the mastery criterion during follow-up. Anne required one feedback session to implement BST for stimulus-stimulus pairing and two feedback sessions to implement BST for DTT. Betty required one feedback session to implement BST for both MSWO and mand training and Carol required one feedback session to implement BST for DTT.

All first tier participants also increased their correct BST tasks compared with total BST task opportunities after training (see Figure 4). During baseline, Anne correctly completed 30 out of 53, 30 out of 76 and 40 out of 75 BST tasks while implementing BST for MSWO, DTT and stimulus-stimulus pairing respectively. During training she correctly completed 258 out of 282 tasks during BST implementation for MSWO. During follow-up, Anne correctly completed 69 out of 80 and 52 out of 56 BST tasks correctly while implementing BST for mand training and graphing, respectively with the same staff member that she had trained during baseline and during her training (S1). During follow-up she also correctly completed 66 out of 75 and 116 out
of 147 BST tasks during implementation of BST for stimulus-stimulus pairing and DTT respectively with a novel staff member (Xylia).

During baseline, Betty correctly completed 40 out of 61, 49 out of 78, 55 out of 93, 54 out of 120 and 53 out of 95 BST tasks while implementing BST for graphing, MSWO, stimulus-stimulus pairing, mand training and DTT respectively. During training she correctly completed 231 out of 254 tasks during BST implementation for stimulus-stimulus pairing. During follow-up, Betty correctly completed 84 out of 98 BST tasks while implementing BST for MSWO with a staff member that she had trained during baseline, but not during her training (S2). She also correctly completed 111 out of 137, 49 out of 54 and 60 out of 64 BST tasks respectively while implementing BST for mand training, DTT and graphing with novel staff members (S4, Zoe and S4).

During baseline, Carol correctly completed 63 out of 104, 61 out of 115, 41 out of 93, 43 out of 65 and 33 out of 89 BST tasks while implementing BST for mand training, stimulus-stimulus pairing, DTT, graphing and MSWO respectively. During training she correctly completed 316 out of 342 tasks during BST implementation for mand training. During follow-up, Carol correctly completed 100 out of 109, 132 out of 140, 72 out of 81, 82 out of 115 and 111 out of 121 BST tasks respectively as she implemented BST for stimulus-stimulus pairing, mand training, graphing, DTT and MSWO with novel staff members (S8, Yolanda, and S9).

Anne spent a total of 2.75 hrs in training for BST and a total of 8.18 hrs participating in the study, Betty spent 2.75 hrs in training and 8.78 hrs in participation, and Carol spent 3.75 hrs in training and 9.65 hrs in participation.

Tier 2
All three second tier participants increased their accuracy of implementation of the trained behavior analytic procedures from baseline to follow-up (see Figures 5 and 6).

During baseline the mean percentages of correctly performed steps for Xylia for stimulus-stimulus pairing and DTT were 55.3%, and 58% respectively. During follow-up the mean percentages of correctly performed steps for Xylia for the same two procedures were 88.2% and 93.3% respectively. Xylia spent a total of 32.48 min, and 1.12 hrs in training for the two procedures, respectively and a total of 5.85 hrs participating in the study.

During baseline the mean percentages of correctly performed steps for participant Yolanda for mand training, DTT and graphing were, 24.3%, 44.3%, and 55.5%, respectively. During follow-up the mean percentages of correctly performed steps for Yolanda for the same three procedures were 82%, 80%, and 90% respectively. Yolanda spent a total of 1.1 hrs, 53.28 min, and 27.08 min in training for the three procedures, respectively and a total of 6.27 hrs participating in the study.

During baseline, the mean percentage of correctly performed steps for Zoe for discrete trial was 40%. During follow-up the mean percentage of correctly performed steps for Zoe for the same procedure was 85%. Zoe spent a total of 49 min in training for DTT and a total of 5.87 hrs participating in the study.

**Social Validity**

Figure 8 presents the social validity responses for the first and second tier participants. Social validity measure responses ranged from 1-5 (from highly disagree to highly agree) for all participants. All first tier participants highly agreed with the statements: “I learned to implement video modeling with a staff member,” “I learned to implement role playing with a staff member,” “I learned to implement feedback with a staff member,” “The feedback I received was
useful,” “Overall, I liked the training I received,” and, “I would recommend this training to other teachers.” Two second tier participants highly agreed and one participant neither agreed nor disagreed with the statements, “I learned to perform S-S pairing with a student,” “I learned to perform MSWO preference assessment with a student,” “I learned to perform mand training with a student,” “I learned to perform DTT with a student,” and “I learned to graph DTT data.” All second tier participants highly agreed with the statements, “The videos I watched were useful,” “Overall, I liked the training I received,” and “I would recommend this training to other entry-level staff.”

**Effect Sizes**

Figure 7 shows the IRDs for each target procedure trained during the second tier and the omnibus IRDs for procedures trained through BST and procedures that were left untrained. The omnibus IRD for trained procedures was .94 CI<sub>95</sub> (.83, 1; large effects). The omnibus IRD for untrained procedures was -.11 CI<sub>95</sub> (-.25, .02; no effects).

**Discussion**

This study added to the current research base in pyramidal staff training by demonstrating that implementation of BST by first tier participants led to improvements in second tier participants’ performance of the targeted ABA procedures, replicating the effects of both Pence et al., (2012) and Parsons et al. (2013). This study demonstrated that the experimenter and classroom teachers used an effective BST package including video modeling, role play and feedback to accomplish socially valid staff training effects. Specifically, results demonstrated were representative of the real world setting, in that classroom teachers used video modeling, role play and feedback to effectively train teaching assistants to implement the target ABA
procedures. Pyramidal training also occurred over an extended time period, and the generality of BST allowed dissemination of multiple ABA procedures across multiple staff members.

This study extended Pence et al., (2012) by demonstrating that the training model allowed classroom teachers to train the teaching assistants to accurately implement several different evidence-based procedures, rather than only preference assessments. The study also extended Parsons et al. (2013) by demonstrating that first tier participants a) used skills acquired during BST of one target skill with one staff member to train novel staff members in at least four other procedures, and b) maintained BST skills for four to five months after training with access to 1-3 brief feedback sessions during follow-up.

Further, the current study addressed the weaknesses in the pyramidal staff training literature by evaluating the correct implementation of a 14 step BST package by first tier participants as a dependent variable and presenting individual first tier participant performance data in a multiple probes across participants graphic display. The study controlled for the effects of instructions because all first tier participants had access to the instructions for BST and all second tier participants had access to the instructions for target procedures during baseline measurement as well as during training and follow-up. This study also allowed for some degradation of the first tier participants’ performance over time during follow-up, because trainers only accessed feedback from the experimenter after a session of BST if they made errors on more than 4 steps of BST during a given session allowing for a more realistic demonstration of maintenance and generality of the skill over time compared with Parsons et al., (2013) where trainers accessed feedback from the experimenter after every session during the post training phase.
Experimenter, first tier participants and script following integrity strengthened the internal validity of the current study compared with previous studies. Further, the training model was socially valid because it addressed programmatic needs for training of new staff, saved resources by eliminating the need for additional staff trainers, and social validity survey data demonstrated the acceptability of the study methods and outcomes.

The current study also presented with several limitations. One limitation was the use of roleplay sessions in place of sessions with staff member trainees and actual students during training for first and second tier participants. According to Maffe-Almodovar and Sturmey (2018), the use of roleplay may limit the generality of training results because training effect sizes for preference assessments, mand training and DTT were smaller when the training context differed from the follow-up context. This limitation was ameliorated for first tier staff to the extent to which first tier participant accuracy during implementation of BST in roleplay sessions was comparable to sessions with actual staff members and students. The limitation was also ameliorated for the second tier staff because Xylia and Zoe accurately implemented target procedures with actual students after training with confederate students; however, Yolanda’s accuracy in implementing mand training and DTT both decreased when she attempted the procedures with actual students after completing training in role play and required additional training to meet mastery criterion with students for both procedures.

Another limitation was that the BST package implemented during the course of this study included 14 detailed steps rather than the 8 general steps of the task analysis used by Parsons et al. (2013). In short, our version of BST may have been more difficult to implement than the version trained by Parsons et al. (2013). For instance, in order for immediate feedback to be considered correct, the first tier participant was required to deliver the feedback within 10 s of a
trainee’s error (step 8). Therefore, if a participant delivered feedback at 12 or 13 s, her performance in delivering feedback was incorrect. In addition, the requirement to deliver feedback “using language from the task analysis” (step 9) may have led to instances where the first tier participant delivered effective feedback, but did not perform correctly according to the task analysis requirement. These and other errors on several other steps were common during baseline; however, these two steps were often the only steps that participants performed incorrectly during follow-up sessions after training. The number of steps of a given target ABA procedure (DTT has 10 steps, whereas vocal mand training has 22 steps) or the number of errors made by a given trainee could also have affected accuracy in performing the BST steps.

Appendix B presents two data sheets for BST of DTT. The first data sheet shows the BST tasks to be scored for the first tier participant when the trainee did not make any errors while performing the procedure. The second data sheet shows the BST tasks to be scored for the first tier participant when the trainee made many errors. The comparison of these two data sheets demonstrates how the required tasks for trainers vary according to trainee performance.

A third limitation was that first tier participant procedural integrity of BST during follow-up was also quite variable, but comparable to those obtained by Pence et al., and Parsons et al., (2013). There were a few outlier sessions where procedural integrity for a given session was quite low. For instance, Anne completed BST with only 44.4% procedural integrity during one session of BST in DTT during follow-up and 55.5% procedural integrity for another session of BST in the same procedure. Anne made errors on steps 8, 9, 11, 13 and 14 (21 correct out of 34 total BST tasks) during one session and on steps 8, 9, 11 and 14 (16 correct out of 21 total BST tasks) during the other session. Betty and Carol also presented with at least one session of low procedural integrity for BST during follow up. Betty implemented one session of BST for
MSWO with only 55.5% accuracy and Carol implemented one session of BST for DTT with 66.6% accuracy. It is difficult to parse out the reasons for these instances of low procedural integrity because the experimenter exposed the first tier participants to novel staff and novel procedures concurrently during follow-up. Thus, it is unclear as to whether degradation of performance occurred as a result of changing staff members, changing procedures or simply the passage of time. Feedback delivered to first tier participants during follow-up also interfered with accurate maintenance measurement for first tier participants. On a more positive note, experimenter feedback improved the performance of all of the first tier trainers and their trainees all met criterion for the target skill and maintained the skill for up to three to five weeks after training without any feedback.

For procedural integrity measurement, Pence et al., (2014) and Parsons et al., (2013) defined a session as the implementation of an entire BST package. Procedural integrity for the current study defined a session as one meeting where the trainee roleplayed or rehearsed the target procedure only one time. Therefore, the completion of an entire BST package for the current study required three or more sessions. Although Anne performed poorly during two sessions of implementation of BST for DTT (44.4% and 55.5%), her overall percentage of accuracy for implementation of the package was 72.8% (116 correct out of 147 total BST tasks). Therefore, the range of procedural integrity for first tier participants’ in their implementation of BST packages for this study (range 72.8% to 94%) had a higher low end than those reported by both Pence et al., (2014) and Parsons et al., (2013). This discrepancy does not negate the procedural integrity problem and it is clear from the current study performance data that all first tier participants required at least some delayed feedback, if only periodically, to accurately implement BST across staff and procedures and over time. This information applied to a
functioning pyramidal training model highlights the importance of oversight and supervision by a qualified behavior analyst to ensure that trainers maintain skills and deliver accurate BST after training when programmatic needs require training of varied procedures with varied staff. The time required for the behavior analyst to oversee and positively affect BST implementation was minimal in light of the current study results as Anne required only 15 total min of feedback over five months post BST training, while Betty and Carol required 10 min and 5 min respectively over four months.

The results of this study add to the current pyramidal training research base and also present opportunities for further research. Future studies should evaluate the feedback component of BST more thoroughly. For instance, parametric analyses of performance feedback may shed some light on whether the latency to feedback after a trainee error or the extent to which language used during feedback matches the task analysis for a given procedure affects feedback effectiveness during BST. Future research might also systematically introduce first tier participants to novel staff and novel procedures post-training in BST to determine whether each aspect affects accurate implementation of the skill. Researchers might also systematically manipulate the number of confederate trainee errors made by trainees during sessions in order to assess whether the number of trainee errors affect the accuracy of trainer BST implementation. Finally, future studies may reveal more complete information regarding the maintenance of BST skills over time by allowing performances to degrade without feedback during post-training follow-up measurement.
Table 1

*Task Analysis for Conditioning Instructors as Secondary Reinforcers*

1. Sanitized environment: Potential preferred items should be secured in clear bins, on high shelves or in clear bags so that student cannot access items without instructor directly providing them.

2. Had items from most recent preference assessment (MSWO/free operant). Also, may have opened a few bins or bags in order to continue to assess which items student chose to interact with.

3. Set timer for 5 minutes.

4. Started timer.

5. Provided first preferred item to student by saying student’s name and/or naming and delivering preferred item by either placing it in student’s hand or mouth, or in front of student (on the table or floor).

6. *Continued to provide preferred items according to rate dictated by student (provided an item as soon as student finished consuming previous item or while student was still interacting with previous item). A maximum of 3 seconds elapsed between completed consumption* of an item and delivery of another item.

7. *Placed demands on student (e.g. instruction, blocking responses or removing potentially dangerous items or school property at risk of being destroyed) only due to problem behavior and allowed more experienced staff member to intervene if potentially dangerous behavior continued for more than 5 seconds.

8. *If student engaged in problem behavior, staff member did not deliver preferred items until problem behavior ceased for at least 3 seconds.

9. *Approached student to deliver item if student did not remain close to staff member.

10. After delivering items for 5 minutes (signaled by timer beep) walked at least 3 feet away from student.

11. Set timer for 5 minutes.

12. Started timer.

13. *Provided a preferred item if student approached and continued to walk 3 feet away from student after delivering each item for next 5 minutes provided that student continued to approach.

14. *Approached student each time student did not approach within 5 seconds and delivered items for next 5 minutes.

15. *Collected data on frequency of student approach by depressing button on a tally counter (preset to zero) each time student approached during final 5 minutes of the session.

16. Ended session when timer beeped.

*Passing criteria is 87% for 3 consecutive sessions*
Table 2

Task Analysis for MSWO Preference Assessment

1. Placed 5 items/edibles in a row on a table or floor in front of student (remove all other potentially distracting items).

2. Ensured appropriate attending prior to presenting preference assessment choices. Attending behavior includes: 1) student is seated, 2) student is oriented-shoulders pointed-toward instructor 3) absence of problem behavior (e.g. crying, aggression, attempting to leave area).

3. Said, “Pick one” to begin trial. Each trial consists of presenting items, allowing student access to a selected item and then removing item.

4. If student failed to select an item within 5 s, randomized placement of items by placing at least half of the remaining objects in different positions and repeated Step 2.

5. If student selected an item within 5 s, gave student up to 30 s of access to item within 5 s of selection. Selection may consist of student touching, pointing to or vocalizing for item.

6. Listed the selected item on the data form in order of selection during student’s access to item.

7. Randomized placement of remaining items/edibles during student’s access to selected item.

8. Removed selected item after providing student up to 30 s access to item and placed it behind him/her or in a container so that it was out of sight of student.

9. If the student selected more than one item at a time or if the student attempted to select a new object when he or she had access to an object, blocked the response and repeated Step 2 if the student did not already have access to an item. If student already had access to an item and selected another item, redirected the student to play with the item he or she already had, by blocking access to other items and saying, “you picked ____”

10. Began next trial by saying “pick one” with remaining items available within 5 s of removing previously selected item.

11. Continued steps 2-9 until all five items/food items were selected or until last 2 items were left without being selected.

12. If no selection was made after 3 consecutive presentations of step 3, ended preference assessment.

Passing criteria is 83% for 3 consecutive sessions
Table 3

Task Analysis for Mand Training - Vocal

1. Selected items for mand training using preference assessments and verbal behavior inventories.

Assessing MO:

2. *Placed item targeted for mand training where student could see it.

3. *Once student attended to item, waited for him/her to demonstrate MO by reaching or leaning toward item.

4. *If student did not attend/demonstrate MO toward item within 5 seconds, manipulated item, moved it about or touched student with it (e.g. roll a car on student’s leg). If student still did not attend/demonstrate MO toward item, conducted additional preference assessment and began again once a preferred item was identified.

Echoic-Mand training (Level I-Level V):

5. Once student demonstrated MO toward item, modeled target word/phrase.

If student responded with word/phrase after only one model:

6. Delivered item and scored a + in E column.

7. Allowed consumption of item (e.g. edible, bubbles, etc), or access to non-consumable (toy, iPhone etc.) for up to 30 sec

8. Removed non-consumables after a maximum of 30 sec by saying, “my turn” and gently removing from student.


If student did not echo target word/phrase after one model within 5 seconds:

10. Did not deliver item, scored a – in E column and begin again at step 5

11. If student echoed target word/phrase after first model on 3 consecutive trials, moved to transfer mand procedure.

12. If student failed to echo target word/phrase on 3 consecutive trials, but continued to demonstrate MO toward item, moved to pairing (bottom of page)

Transfer Mand Procedure (Level I-Level V):

13. Once student demonstrated MO toward item, waited 5 seconds for student to emit target mand.

If the student emitted the target mand:
14. Delivered item and scored a + in the M column.

15. Allowed consumption of item, or access to a non-consumable for up to 30 sec

16. Removed non-consumables after a maximum of 30 sec

**If student did not emit target mand:**

17. Did not deliver item and scored a – in M column.

18. Modeled target word/phrase and delivered item, only when student correctly echoed target word/phrase

19. Removed any non-consumable reinforcers after a maximum of 30 sec by saying, “my turn” and gently removing item from student.

20. Began sequence again starting with step 13

21. If student failed to emit target mand over 3 consecutive trials, returned to echoic-mand training.

22. Once 3 consecutive echoics were emitted (and reinforced), moved back to transfer mand procedure.

---

*Passing criteria is 90% for 3 consecutive sessions*
Task Analysis for DTT

1. Ensured appropriate attending prior to presenting antecedent. Attending behavior includes: 1) student is seated, 2) student is oriented-shoulders pointed-toward instructor 3) absence of problem behavior (e.g. crying, aggression, attempting to leave area).

2. Ensured that student was looking at instructor or task materials before giving any verbal instructions with clear articulation and neutral tone.

3. Delivered antecedent that matched designated program.

4. Provided predetermined prompt designated for each program within 1 s after delivering antecedent.

If response was correct:

5. Delivered praise within 1 s.

6. Delivered reinforcer within 1 s of praise according to Sr+ schedule.

7. Recorded a plus on data sheet.

If response was incorrect:

8. Delivered corrective prompt within 1 s following an incorrect response paired with corrective feedback (e.g. “This is tapping the table.”).

9. Recorded a minus on datasheet.

10. Waited until student consumed reinforcer (e.g. edible, bubbles, etc.) or removed non-consumable (e.g. toy, iPhone etc.) after designated reinforcer interval by saying, “my turn” and gently removing the item from the student and began the next trial.

Passing criteria is 80% for 3 consecutive sessions
Table 5

Task Analysis for Calculating and Graphing Discrete Trial Data

1. Located correct graph within student’s data binder or clipboard.

2. Calculated percentage correct only if there were a minimum of five student responses by:

3. Counting the number of correct responses the student made as indicated by a (+).

4. Counting the number of total responses the student made as indicated by (+) and (-).

5. Dividing the number of correct responses by the number of total responses and then multiplying by 100.

6. Placed a dot on the graph that corresponded to the final percentage calculated by following the horizontal gridline from the corresponding percentage number on the Y axis (left side) to the first available vertical gridline where no dots were currently plotted.

7. If the data sheet indicated that the prompt level or target skill had not changed since the previous dot was plotted on the graph (the last dot before the current dot), used a ruler to draw a solid line from the previous dot to the current dot.

8. If the data sheet indicated that the prompt level or target skill had changed since the last dot on the graph, drew a phase line by:

9. Using a ruler to draw a dashed vertical line between the last data point of the previous sequence and the current data point, extending from at least 1 cm above the top of the Y axis down to the X axis of the graph

10. Writing the new skill or prompt level after the phase line and on top of the graph

11. Did not connect dots across the phase line.

12. Filled in all other necessary information (e.g., date, initials, etc.)

Passing criteria is 83.3% for 3 consecutive sessions
Table 6

**Task Analysis for BST: Video Model, Role Play and Feedback**

**Model**
1. Gave task analysis for procedure to trainee. The trainee should be seated at a computer, iPad or similar device.
2. Told the trainee to watch the video model for conducting the procedure
3. Told the trainee to check off the steps of the task analysis that he/she was given as he/she sees them.
4. Ensured that that the video model began to play before walking away from the trainee.
5. Inspected the task analysis to ensure that the trainee checked off all relevant steps.

**Role play/rehearsal & Feedback***
6. Instructed the trainee to, based on the video model, conduct the procedure to the best of their ability.
7. Collected data on every step of the task analysis on the trainer roleplay/rehearsal and feedback data form by the time the trainee had finished roleplaying/rehearsing the procedure. The trainer must have all data collected on the correct data sheet prior to the final review of errors with the trainee.
8. Provided immediate feedback, anytime the trainee made an error while conducting the procedure.
9. Used language from the task analysis to refer to correct behavior while giving feedback for error
10. When the trainee performed a step correctly, either did not provide feedback, or acknowledged the performance as correct.
11. Reviewed all errors with the trainee after roleplay/rehearsal was complete
12. Made a minimum of 2 positive statements about the trainee’s performance (at least start and end).

**Student session**- the staff member rehearses the protocol with a staff member acting as a student confederate (no student).

**Training to Criterion**
13. Scored steps performed by trainee correctly (with at least 80% IOA with experimenter)
14. Correctly informed the trainee as to whether they have met criterion (at least two consecutive role play sessions followed by one student session at criterion) for the procedure and what the next session will entail (e.g. "you did not meet criterion for training yet, so we will continue with another role play session")

*Criterion= 87.5% over 2 R and F sessions*
### Table 7

**Sequence of sessions for first and second tier participants**

<table>
<thead>
<tr>
<th>Participant</th>
<th>Sequence of procedures during baseline</th>
<th>Sequence of procedures during follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A</strong></td>
<td>2, 9, 5</td>
<td>8, 10, 6, 9</td>
</tr>
<tr>
<td><strong>B</strong></td>
<td>10, 2, 1, 8, 4</td>
<td>7, 8, 9, 10</td>
</tr>
<tr>
<td><strong>C</strong></td>
<td>3, 6, 4, 5, 7</td>
<td>6, 8, 10, 9, 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Participant</th>
<th>Sequence of procedures during baseline</th>
<th>Sequence of procedures during follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>X</strong></td>
<td>2, 1, 3, 4, 5, 5, 3, 1, 7, 4, 2, 5, 4, 2, 8, 1, 3, 5, 2, 1, 3, 6, 4, 9, 10</td>
<td>7, 8, 9, 6, 10, 9, 7, 6, 10, 8, 8, 6, 9, 7, 10</td>
</tr>
<tr>
<td><strong>Y</strong></td>
<td>2, 1, 10, 8, 9, 4, 5, 7, 6, 3, 5, 2, 4, 1, 3</td>
<td>7, 9, 6, 8, 10, 9, 7, 10, 8, 6, 7, 9, 10, 8, 10, 9, 6, 7, 8, 10, 8, 6, 9, 7</td>
</tr>
<tr>
<td><strong>Z</strong></td>
<td>1, 2, 3, 4, 5, 5, 4, 3, 2, 1, 2, 3, 5, 1, 7, 9, 6, 4, 10, 8</td>
<td>7, 10, 6, 9, 8, 6, 9, 8, 10, 7, 8, 7, 9, 6, 10, 8, 7, 10, 6, 9</td>
</tr>
</tbody>
</table>
Table 8

**BST details for first tier participants**

<table>
<thead>
<tr>
<th>Participant</th>
<th>Target Procedure</th>
<th>Length of Video Model</th>
<th>Incorrect Steps Performed by Trainee</th>
<th>Incorrect Steps Performed by Trainee</th>
<th>Incorrect Steps Performed by Trainee</th>
<th>Incorrect Steps Performed by Trainee</th>
<th>Total Training Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anne</td>
<td>MSWO</td>
<td>14.82 min</td>
<td>Session 1 8, 9</td>
<td>Session 2 3, 10, 12</td>
<td>Session 3 2, 8</td>
<td>Session 4 1, 4, 12</td>
<td>2.75 hrs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Session 1 6, 7</td>
<td>Session 2 5, 8</td>
<td>Session 3 6, 9</td>
<td>Session 4 8, 9</td>
<td></td>
</tr>
<tr>
<td>Betty</td>
<td>S-S Pairing</td>
<td>14.15 min</td>
<td>Session 1 4, 7</td>
<td>Session 2 10, 14</td>
<td>Session 3 3, 9</td>
<td>Session 4 NA</td>
<td>2.75 hrs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Session 1 6, 8</td>
<td>Session 2 NA</td>
<td>Session 3 15, 16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carol</td>
<td>Mand Training</td>
<td>13.28 min</td>
<td>Session 1 4, 9, 6</td>
<td>Session 2 1, 10</td>
<td>Session 3 2, 5</td>
<td>Session 4 4, 5, 9, 10</td>
<td>3.75 hrs</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Session 1 2, 11</td>
<td>Session 2 4, 5</td>
<td>Session 3 1, 9</td>
<td>Session 4 4, 5, 9, 10</td>
<td></td>
</tr>
</tbody>
</table>


Figure 1. The top panel shows the pyramid of training for the total number of staff members that volunteered to be trained during the study. The bottom panel shows the pyramid of skills disseminated by the experimenter to the first tier participants and by the first tier participants to the second tier participants.
Figure 2. The graph shows each first tier participant’s percentage of procedural integrity of BST and each second tier participant’s percentage of accurately performed steps on trained procedures. The open symbols represent sessions with a confederate student. The closed symbols represent sessions with a student. Underlines under a data point indicate that the experimenter provided feedback to the trainer after the session.
Figure 3. The graph shows percentage of correct BST steps. Open symbols present sessions conducted with a particular staff volunteer (S1-S9) or with a Tier 2 confederate staff member and confederate students. Closed symbols present sessions conducted with actual staff and students. Brackets present sessions completed with a particular staff volunteer (S1-S9) or with a Tier 2 confederate staff member and confederate students. Underlines under a data point indicate that the experimenter provided feedback to the trainer after the participant.
Figure 4. The graph shows the number of correct BST tasks compared with total BST tasks required for a session. Open symbols present sessions conducted with actual staff trainees and students. Closed symbols present sessions conducted with a particular staff volunteer (S1-S9). Brackets present sessions conducted with specific staff members and confederate students. Closed symbols present sessions conducted with confederate staff members and confederate students. Open symbols present sessions conducted with a Tier 2 participant. Underlines under a data point indicate that the experimenter provided feedback to the trainer after the session.
Figure 5. The graph shows each second tier participant’s percentage of accurately performed steps on each procedure across the phases of the experiment. The open symbols represent sessions with a confederate student. The closed symbols represent sessions with a student. Underlines under a data point indicate that the experimenter provided feedback to the trainer after the session.
Figure 6. The graph shows each second-tier participant’s percentage of accurately performed steps on trained procedures only across the phases of the experiment. The open symbols represent sessions with a confederate student. The closed symbols represent sessions with a student. The underlines under a data point indicate that the experimenter provided feedback to the trainer after the session.
Figure 7. The graph shows the IRDs for each procedure trained with BST (top left panel) or left untrained (bottom left panel) during the second tier of the study and the omnibus IRDs for procedures trained through BST versus procedures that were left untrained (right panel).
Figure 8. The graphs show social validity responses for the first and second tier participants.
Appendix A

Social Validity Survey for First Tier Participants

Please circle 1-5 for the following statements.
1= highly disagree, 2= sort of disagree, 3= neither agree or disagree, 4= sort of agree, 5= highly agree

1. I learned to implement video modeling with a staff member.
   1 2 3 4 5

2. I learned to implement role playing with a staff member.
   1 2 3 4 5

3. I learned to implement feedback with a staff member.
   1 2 3 4 5

4. The training I received taught me how to work with staff members in the classroom
   1 2 3 4 5

5. The videos I watched were useful.
   1 2 3 4 5

6. I would have liked to watch the videos more often.
   1 2 3 4 5

7. The feedback I received was useful.
   1 2 3 4 5

8. Overall, I liked the training I received.
   1 2 3 4 5

9. The training was easy.
   1 2 3 4 5

10. The training was short.
    1 2 3 4 5

11. The training could be improved. (Please tell us how it could be improved.)
    1 2 3 4 5

12. I would recommend this training to other teachers.
    1 2 3 4 5

13. Additional comments:
Social Validity Survey for Second Tier Participants

Please circle 1-5 for the following statements.

1= highly disagree, 2= sort of disagree, 3= neither agree or disagree, 4= sort of agree, 5= highly agree

1. I learned to perform stimulus-stimulus pairing (conditioning myself as a reinforcer) with a student.
   1 2 3 4 5

2. I learned to perform MSWO preference assessment with a student.
   1 2 3 4 5

3. I learned to perform mand training with a student.
   1 2 3 4 5

4. I learned to perform discrete trial teaching (DTT) training with a student.
   1 2 3 4 5

5. I learned to graph DTT data.
   1 2 3 4 5

6. The training I received taught me how to work with the students in the classroom.
   1 2 3 4 5

7. The videos I watched were useful.
   1 2 3 4 5

8. I would have liked to watch the videos more often.
   1 2 3 4 5

9. The feedback I received was useful.
   1 2 3 4 5

10. Overall, I liked the training I received.
    1 2 3 4 5

11. The training was easy.
    1 2 3 4 5

12. The training was short.
    1 2 3 4 5

13. The training could be improved. (Please tell us how it could be improved.)
    1 2 3 4 5

14. I would recommend this training to other entry-level staff.
    1 2 3 4 5

15. Additional comments
## BST for DTT Data Sheet with No Trainee Errors

**Staff Name:** Betty  
**Trainer:** LM  
**Date:** 10/30-11/2/17  
**Component:** *Rehearsal and feedback*

<table>
<thead>
<tr>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Gave task analysis for procedure to trainee. The trainee should be seated at a computer, iPad or similar device.</td>
</tr>
<tr>
<td>2. Told the trainee to watch the video model for conducting the procedure</td>
</tr>
<tr>
<td>3. Told the trainee to check off the steps of the task analysis that he/she was given as he/she sees them.</td>
</tr>
<tr>
<td>4. Ensured that the video model began to play before walking away from the trainee.</td>
</tr>
<tr>
<td>5. Inspected the task analysis to ensure that the trainee checked off all relevant steps.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Role play/rehearsal &amp; Feedback*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Role play session</strong>- the staff member rehearses the protocol with a staff member acting as a student confederate (no student).</td>
</tr>
<tr>
<td><strong>Student session</strong>- the staff member rehearses the protocol with a student from the classroom.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Role play/rehearsal &amp; Feedback*</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Instructed the trainee to, based on the video model, conduct the procedure.</td>
</tr>
<tr>
<td>7. Collected data on every step of the task analysis on the trainer roleplay/rehearsal and feedback data form by the time the trainee had finished roleplaying/rehearsing the procedure. The trainer must have all data collected on the correct data sheet prior to the final review of errors with the trainee.</td>
</tr>
<tr>
<td>8. Provided immediate feedback, anytime the trainee made an error while conducting the procedure.</td>
</tr>
<tr>
<td>9. Used language from the task analysis to refer to correct behavior while giving feedback for error</td>
</tr>
<tr>
<td>10. When the trainee performed a step correctly, either did not provide feedback, or acknowledged the performance as correct.</td>
</tr>
<tr>
<td>11. Reviewed all errors with the trainee after roleplay/rehearsal was complete</td>
</tr>
<tr>
<td>12. Made a minimum of 2 positive statements about the trainee’s performance (at least start and end).</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Training to Criterion</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. Scored steps performed by trainee correctly (with at least 80% IOA with experimenter)</td>
</tr>
<tr>
<td>14. Correctly informed the trainee as to whether they have met criterion (at least two consecutive role play sessions followed by one student session at criterion) for the procedure and what the next session will entail (e.g. &quot;you did not meet criterion for training yet, so we will continue with another role play session&quot;)</td>
</tr>
</tbody>
</table>

**Correct(+)**  
**Incorrect(−)**  
**11/2/17 (R+F 3)**  
**Beginning**  
**End**  
**Criterion**  
**Next Session**
## BST for DTT Data Sheet with Many Trainee Errors

### Staff Name: Carol  
Trainer: LM  
Date: 4/28-5/26/17  
Component: **Rehearsal and feedback**  
Correct=(+)  
Incorrect=(−)  
5/26/17-1 (R+F 2)

### Model
1. Gave task analysis for procedure to trainee. The trainee should be seated at a computer, iPad or similar device.

2. Told the trainee to watch the video model for conducting the procedure.

3. Told the trainee to check off the steps of the task analysis that he/she was given as he/she sees them.

4. Ensured that that the video model began to play before walking away from the trainee.

5. Inspected the task analysis to ensure that the trainee checked off all relevant steps.

### Role play/rehearsal & Feedback*

#### Role play session- the staff member rehearses the protocol with a staff member acting as a student confederate (no student).

6. Instructed the trainee to, based on the video model, conduct the procedure.

7. Collected data on every step of the task analysis on the trainer roleplay/rehearsal and feedback data form by the time the trainee had finished roleplaying/rehearsing the procedure. The trainer must have all data collected on the correct data sheet prior to the final review of errors with the trainee.

8. Provided immediate feedback, anytime the trainee made an error while conducting the procedure.

9. Used language from the task analysis to refer to correct behavior while giving feedback for error

10. When the trainee performed a step correctly, either did not provide feedback, or acknowledged the performance as correct.

11. Reviewed all errors with the trainee after roleplay/rehearsal was complete

12. Made a minimum of 2 positive statements about the trainee’s performance (at least start and end).

### Training to Criterion
13. Scored steps performed by trainee correctly (with at least 80% IOA with experimenter)

14. Correctly informed the trainee as to whether they have met criterion (at least two consecutive role play sessions followed by one student session at criterion) for the procedure and what the next session will entail (e.g. "you did not meet criterion for training yet, so we will continue with another role play session")

<table>
<thead>
<tr>
<th>Step</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>7</td>
<td>9</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

### Beginning  
End

### Correct=(+)  
Incorrect=(−)  
5/26/17-1 (R+F 2)
References


Nosik, M. R., Williams, W., Garrido, N., & Lee, S. (2013). Comparison of computer based instruction to behavior skills training for teaching staff implementation of


