Scaling Up Behavioral Skills Training: Effectiveness of Large-Scale and Multiskill Trainings

Andrea B. Courtemanche¹ · Laura B. Turner¹ · John D. Molteni^{1,2} · Nicole C. Groskreutz^{1,3}

Published online: 2 November 2020

© Association for Behavior Analysis International 2020

Abstract

We used behavioral skills training (BST) to teach multiple skills to 2 cohorts of 18 participants. BST consisted of the standard 4 components: (a) didactic instruction, (b) modeling, (c) role-play, and (d) feedback, modified to be delivered in a large-group format. All components were provided by 1 trainer, simultaneously to all participants, with peers delivering feedback during role-plays. Across 4 targeted skills (e.g., discrete-trial teaching), the average performance of Cohort 1 improved from less than 60% correct implementation in baseline to a performance of between 85% and 100% correct, across participants, following BST. We used social validity data collected from Cohort 1 to modify the length of instruction across skills for Cohort 2. BST was similarly effective for Cohort 2, with a decrease in the additional training required for trainees to demonstrate the skill in a novel role-play scenario or with a client. Implications for effectively scaling up BST are discussed.

Keywords Behavioral skills training · Group training · Peer feedback · Social validity · Staff training

Applied behavior-analytic interventions, in both research and practice, are commonly implemented within a hierarchical model in which those providing the direct, day-to-day service (e.g., paraprofessionals, students, caregivers, and Registered

Behavior Technicians; RBTs) often have not received formal or advanced education in applied behavior analysis (Romanczyk, Callahan, Turner, & Cavalari, 2014). Therefore, the initial and ongoing job training and supervision of these “service providers” is crucial when implementing behavior-analytic interventions (Behavior Analyst Certification Board [BACB], 2014). As a result, behavior-analytic researchers and practitioners have focused on identifying effective training methods (e.g., Page, Iwata, & Reid, 1982).

The training method that has received the most attention and empirical support is behavioral skills training (BST). BST includes an expert trainer who provides implementers with (a) a rationale regarding the importance of the skill, (b) vocal and written instructions on how to implement the skill (e.g., didactic instruction), (c) a model of the skill, (d) the opportunity to practice the skill in role-play situations, and (e) performance feedback regarding their use of the skill. This sequence of training steps is repeated until the trainees meet mastery criteria (BACB, 2018; Parsons, Rollyson, & Reid, 2013). Recent research also suggests that trainees rate BST as an acceptable training procedure (e.g., Fetherston & Sturmey, 2014; Hassan et al., 2018), with several of the individual components preferred by trainees. Specifically, trainees prefer the use of modeling or demonstrations, practice opportunities, and performance feedback (Reid, Green, Parsons, & Rotholz, 2018; Strohmeier, Mule, & Luiselli, 2014).

Research Highlights

- We scaled up behavioral skills training (BST), effectively teaching two cohorts of 18 participants multiple skills, delivering all components of BST in an efficient large-group format.
- We used social validity data to modify the length of time spent on the instruction of various skills, with the performance and social validity ratings of the second cohort supporting these modifications.
- Across large groups of trainees (18 participants), we evaluated the use of trained skills in a novel context (either during a role-play scenario or with an actual client).
- During role-plays, instead of trainers providing all feedback, peers (i.e., fellow participants) delivered feedback and determined when they felt their groups had acquired the trained skills, increasing training efficiency.

✉ Andrea B. Courtemanche
acourtemanche@usj.edu

¹ Department of Counseling and Applied Behavioral Studies, University of Saint Joseph, 1678 Asylum Dr., West Hartford, CT 06117, USA

² PRISM Autism Center, Farmington, CT, USA

³ Peak Behavioral Services, LLC, Jackson, WY, USA

BST has been empirically evaluated across a variety of implementer populations and skills. For example, using BST, undergraduates have learned to conduct functional analyses (Iwata et al., 2000), special education teachers have learned to implement discrete-trial teaching (Sarokoff & Sturmey, 2004), and teaching assistants have learned to systematically assess student preferences (Lavie & Sturmey, 2002). Researchers have also effectively used BST to teach adults with autism spectrum disorder (ASD) to implement discrete-trial and incidental teaching procedures with young children with ASD (Lerman, Hawkins, Hillman, Shireman, & Nissen, 2015).

BST has typically been delivered and systematically evaluated in a one-on-one format, meaning that an expert trainer teaches a single implementer (Crockett, Fleming, Doepke, & Stevens, 2007; Fetherston & Sturmey, 2014; Homlitas, Rosales, & Candel, 2014; Iwata et al., 2000; Lafaskakis & Sturmey, 2007; Lambert, Bloom, Kunnavatana, Collins, & Clay, 2013; Lavie & Sturmey, 2002; Lerman et al., 2015; Madzharova, Sturmey, & Jones, 2012; Miles & Wilder, 2009; Nigro-Bruzzi & Sturmey, 2010; Rosales, Stone, & Rehfeldt, 2009; Roscoe & Fisher, 2008; Sarokoff & Sturmey, 2004). However, BST has also been effectively delivered in a group format. For example, Iwata et al. (2000) delivered instructions and modeling in a group format to 11 undergraduates, while still providing the interactive components of role-play, practice, and feedback in a one-on-one format. Parsons, Rollyson, and Reid (2012, 2013) also incorporated group delivery of instruction, with two expert trainers successfully implementing the *entire* BST process in small groups of eight, and three to four participants concurrently. In both studies, Parsons et al. (2012, 2013) had expert trainers deliver feedback to participants while they role-played and practiced skills with each other. Importantly, although the delivery of feedback by expert trainers is considered effective and best practice, it is also a resource-intensive procedure.

Many studies on BST focus on teaching one or two skills to implementers (e.g., Fetherston & Sturmey, 2014; Homlitas et al., 2014; Iwata et al., 2000; Lafasakis & Sturmey, 2007; Lambert et al., 2013; Lavie & Sturmey, 2002; Lerman et al., 2015; Madzharova et al., 2012; Nigro-Bruzzi & Sturmey, 2010; Rosales et al., 2009; Roscoe & Fisher, 2008; Sarokoff & Sturmey, 2004). More recently, Sawyer et al. (2017) evaluated the effectiveness of BST in teaching multiple skills to a larger group of individuals simultaneously. These researchers trained a small group of seven undergraduate, preservice teachers several behavior-analytic skills (e.g., multiple-stimulus without-replacement preference assessment, least-to-most prompting, response interruption and redirection). Similar to Parsons et al. (2012, 2013), the participants in Sawyer et al. (2017) role-played and practiced with each other, as well as with the instructor, who provided additional modeling, corrective feedback, and praise before each participant's competency check.

Although the literature evaluating BST is growing, it is still limited regarding information on effectively and efficiently scaling up the procedures (in terms of both teaching multiple skills and teaching multiple individuals simultaneously). Additional outcome evaluations regarding large-scale BST are important because many behavior-analytic practice organizations provide initial and ongoing training to large groups with the intended outcome of performance-based competency on several skills for assessing, increasing, and decreasing behaviors.¹ For example, organizations that employ RBTs are required to utilize BST to teach a range of content on the *RBT Task List*, Second Edition (BACB, 2019). Given the effectiveness of BST procedures when implemented with small groups and individuals, it is reasonable to hypothesize that BST can also be effective when used to teach multiple skills to larger groups of service providers. Empirical demonstrations are needed to help support organizations when implementing best practices at a larger group level.

Therefore, the purpose of this study was to respond to the call of other researchers to continue to scale up BST (e.g., Sawyer et al., 2017), while preserving both the effectiveness and the efficiency of the procedures. To accomplish this, we systematically evaluated the acquisition and maintenance of multiple skills (in trained and novel contexts) taught using a resource-friendly, peer-feedback BST model with two large groups (i.e., 18 trainees per group with one expert trainer). We further utilized these two case examples to highlight the use of social validity assessments not only as standard outcome measures (e.g., Fetherston & Sturmey, 2014; Parsons et al., 2013) but also as data to support modifications to our training procedures to improve the effectiveness and acceptability of those procedures across cohorts.

Case Demonstration 1: Increasing the Scale of BST

Method

We provided BST to a group of employees at a private, state-approved, special education laboratory school located on a university campus. School administrators were seeking to transition their school away from using an eclectic educational approach toward a greater focus on applied behavior analysis (see Howard, Stanislaus, Green, Sparkman, & Cohen, 2014). Administrators approached the trainers seeking a large-scale behavior-analytic training that covered multiple topic areas and would be appropriate for the majority of their staff.

¹ For demonstrations of knowledge-based competencies across large groups, see Luiselli, St. Amand, MaGee, and Sperry (2008) and Luiselli, Bass, and Whitcomb (2010).

Participants and setting School administrators identified employees who would benefit from the training, but participation in the training was voluntary. The special education school held weekly, half-day professional development trainings. The training outlined in this study occurred during those times. Participants received their hourly pay during training hours. Twenty individuals initially consented to participate and filled out a demographic questionnaire. Two participants dropped out of the training. Cohort 1 included 18 individuals (13 females and 5 males) ranging from 23 to 52 years of age. All participants were employed as either classroom teachers, related service personnel, or paraprofessionals who provided direct services to school-aged individuals with intellectual and developmental disabilities. Participants' duration of employment at the school ranged from 2 months to 15 years. All participants held a high school diploma, and more than half ($n = 11$) held a bachelor's degree. A university institutional review board approved all study procedures. Participants provided informed consent for us to videotape during observations and share their de-identified data. All of the didactic and role-play portions of the training were conducted in a classroom setting on a university campus. The in vivo observations were conducted at the participants' place of employment during the times when they were providing direct services to a familiar client with a disability in the client's natural classroom setting.

Trainer qualifications The trainers consisted of three doctoral-level Board Certified Behavior Analysts who were full-time faculty members at a university with a verified applied behavior analysis course sequence. Each of the trainers worked independently when training the participant group, so that the trainer-to-trainee ratio remained consistently at 1:18, with each of the three trainers providing different portions of the training.

Dependent measures and data collection To identify socially significant skills, we provided school administrators where the trainees worked with the RBT *Task List* (BACB, 2013) and assisted them in identifying which skills were most relevant. We systematically assessed four performance-based skills (i.e., prior to and following BST on the particular skill, and after a delay during a follow-up assessment) to evaluate the trainees' acquisition and maintenance of these skills. The four targeted skills included (a) writing an objective session note, (b) conducting a paired-stimulus preference assessment, (c) using discrete-trial teaching (DTT), and (d) implementing a differential reinforcement procedure to replace a problem behavior with an appropriate alternative behavior (DRA). We chose these four skills because they were the most relevant skills based on the job requirements and current practices within the school where training took place.

We task analyzed all four skills into component steps that represented the correct implementation of the skills (see Appendix 1 for task analyses of the correct implementation of each skill). For each skill, observers recorded whether the participant correctly implemented each component step. For example, during DTT, if the participant provided the discriminative stimulus once and only once per trial, he or she received a "+" for that step. At the end of each observation, we calculated the percentage of steps implemented correctly by the participant (i.e., the total number of steps completed correctly divided by the total number of steps, then converted to a percentage).

Interobserver agreement On 22% of sessions, a second independent observer recorded the participant's performance of the targeted skill (either live or via videotape). A point-by-point comparison between the data collected by the primary and secondary observers was conducted for each step in the skill. The number of agreements was divided by the number of agreements plus disagreements and then converted to a percentage. Across participants, interobserver agreement (IOA) was 94.6% (range 75%–100%), 96.5% (range 71%–100%), 95.3% (range 76%–100%), and 95.2% (range 83.3%–100%) for writing an objective note, conducting a paired-stimulus preference assessment, implementing a DRA procedure, and using DTT, respectively. There were several IOA scores that were below 80%, especially for the skill of writing an objective note. There were only four steps associated with this skill. As such, if there was a disagreement on one step of the skill, it resulted in a score of 75% agreement. Anytime an unacceptable level of IOA occurred, we conducted retraining of the second observers using videos of participants.

Experimental design We used a multiple-probe design across skills, replicated across all 18 participants, to evaluate the effectiveness of the group BST procedure on acquisition and maintenance of the four targeted skills. See Table 1 for a timeline of training and assessment activities for Cohort 1. Each participant's performance on the four skills was evaluated at four different times. First, all four skills were assessed at the beginning of the study. Second, each skill was individually assessed again, prior to the onset of BST for a targeted competency (except writing an objective note, as this was the first skill taught). Third, each targeted competency was individually assessed immediately after BST for that particular skill where participants demonstrated the specific skill during role-play situations with a trainer. The first three assessments took place across 8 weeks.

The fourth assessment of skills occurred after the 8-week training was completed (i.e., maintenance). During maintenance testing, participants demonstrated the skill of conducting a paired-stimulus preference assessment directly with a client with a disability with whom they typically

Table 1. Training Schedule for Cohort 1

Week 1: Training Day 1	Week 2: Training Day 2	Week 3: Training Day 3	Week 4: Training Day 4	Week 5: Training Day 5
1. Baseline assessment of all skills 2. BST on writing an objective note 3. Acquisition assessment of writing an objective note	1. Maintenance assessment of writing an objective note 2. BST on PA (Part 1)	1. Baseline assessment of PA 2. BST on PA (Part 1)	1. BST on PA (Part 2) 2. Acquisition assessment of PA	1. Baseline assessment of DRA 2. BST on DRA (Part 1)
	Week 6: Training Day 6	Week 7: Training Day 7	Week 8: Training Day 8	Weeks 9–12: Maintenance Assessments
	1. BST on DRA (Part 2) 2. Acquisition assessment of DRA	1. Baseline assessment of DTT 2. BST on DTT (Part 1)	1. BST on DTT (Part 2) 2. Acquisition assessment of DTT	1. Maintenance assessments of DTT, DRA, and writing an objective note in novel role-play 2. Naturalistic observations of PA

Note. BST = behavioral skills training; PA = preference assessment; DTT = discrete-trial teaching; DRA = differential reinforcement of alternative behavior

worked. We assessed performance of the remaining three skills during a novel role-play situation with one of the trainers, in a classroom on the university campus where the training occurred.

Procedures

Baseline Prior to beginning the training, we asked participants to demonstrate each of the four skills. For the skill of writing an objective note, participants all watched the same 5-min video, at the same time, and then attempted the skill of writing an objective note. We also individually assessed the performance of all 18 trainees on the remaining three skills in separate role-play scenarios. Each role-play situation was designed to take approximately 5 min. The participant was given a description of the role-play scenario, data sheets for all three skills, a lesson plan for DTT, and behavior guidelines for the DRA procedure. Participants were given approximately 1–2 min to review all of the written materials. The trainer answered trainees' questions with "Do your best." During the baseline assessments, the trainer acted as a client and followed a script with a predefined response pattern (including correct, incorrect, and no-response trials). The trainer did not give participants any feedback on their performance. We conducted at least one baseline probe on each of the four skills prior to beginning the training.

BST Training consisted of four major components: (a) didactic instruction with rationales, (b) modeling the skill, (c) role-play practice with peers, and (d) peer feedback. We delivered all training components in a large-group format (i.e., to all 18 participants simultaneously).

Instructions and rationales. Group didactic instruction with rationales lasted 10–300 min (see Table 2), depending on the complexity of the skill targeted. The trainer used PowerPoint presentations that we developed. Presentations included (a) a general description of each skill, (b) rationales as to why each skill was important, and (c) detailed descriptions of the steps required to complete each skill. During didactic instruction, we discussed how the participants could use these skills in their daily work. Participants received copies of the PowerPoint slides and the task analyses that outlined the correct implementation of each skill (see Appendix 1). For applicable skills, participants also received sample data sheets. For example, when teaching the skill of conducting a paired-stimulus preference assessment, participants were given a data sheet (see Appendix 2) to aid them in conducting the assessment.

Modeling. We then modeled the relevant skill to the group using the exact steps on the task analysis of the skill (see Appendix 1). The trainer always modeled the skill live,

recruiting a participant to play the role of the client. We asked the participant to engage in correct and incorrect responses so that the trainer was able to demonstrate all components of the skill. We also supplemented live demonstrations with videos of the skill that accurately depicted all of the steps of the task analysis. For example, for DTT, the trainer modeled the skill and showed participants a video of a teacher using DTT. For all skills, live and video demonstrations combined lasted less than 5 min in duration.

Role-play with peer feedback. After we modeled each skill, participants practiced in role-play situations in small groups of two to three participants. Within their groups, participants rotated through roles to (a) implement the target skill and (b) collect procedural fidelity data on the steps of the task analyses and provide feedback to the peer practicing the skill. Participants continued to role-play each skill with peer feedback until each group self-determined that they implemented the relevant steps correctly. During role-play practice with peer feedback, the trainer monitored participants and answered participants' questions. The trainer did not systematically engage in role-play practice or provide feedback to each participant.

Acquisition assessment Immediately after completing the role-play practice with peers, we assessed each participant's performance of the targeted skill. These acquisition assessment procedures were similar to baseline procedures but differed in two main ways. First, during the acquisition assessment of writing an objective note, participants watched a different video from the one in baseline. Second, after each participant demonstrated the targeted skill in the role-play with a trainer, the trainer delivered individualized positive and corrective verbal feedback based on the participant's performance. After the trainer delivered feedback, the participant was able to leave the training for that day.

Maintenance assessment in trained and novel contexts The entire BST training on all skills lasted 8 weeks. After the 8-week training, we assessed maintenance of all skills using either a novel role-play situation or in vivo observation of the participant

performing the skill with a client (see Table 1 for a detailed schedule of assessments). The participants conducted the paired-stimulus preference assessment with a client who was enrolled at the special education school where participants worked. Participants chose a familiar client with whom they regularly worked. Participants demonstrated the remaining three skills within novel role-play scenarios (i.e., scenarios different from those used previously) with one of the trainers (randomly assigned) in a university classroom. Immediately following the participant's performance of each skill, the trainer provided both positive and corrective feedback to the participant. If a participant did not perform a skill to competency (i.e., with at least 90% of steps correctly implemented), the participant was required to demonstrate the skill again at least 24 hr later. We repeated this process until the participant performed the skill to competency.

Treatment integrity We videotaped all training sessions and randomly selected 50% of skills (i.e., DTT and DRA) to assess our fidelity to BST. We used treatment integrity checklists to assess whether the trainer (a) gave a description of the skill to be targeted, (b) provided a rationale as to why the skill was important, (c) provided a detailed description of the steps of the targeted skill, (d) modeled the targeted skill according to the task analysis, and (e) had the trainees practice the skill in role-play situations. Our integrity to BST was 100% across sessions and trainers.

Results

The average performance of Cohort 1 across the four skills during baseline, acquisition, and maintenance probes is depicted in the top panel of Fig. 1. Two of the 18 participants were not present for the baseline assessment of the skill of using DRA (although both participants were present for the BST training and postassessments on DRA). As a result, only 16 participants' data are presented in Fig. 1 for DRA. During baseline, participants' correct implementation of each skill was 42% (range 0%–75%) for writing an objective note, 48% (range 15%–63%) for conducting a paired-stimulus preference assessment, 67% (range 49%–82%) for the DRA procedure, and 57% (range 41%–68%) for DTT. Following BST for each of the four targeted skills, participants' average

Table 2. Scheduled Minutes Spent on Behavioral Skills Training Across Skills and Cohorts

Skill	Rationale/Instructions/Modeling		Peer Practice and Feedback		Total Time of BST	
	Cohort 1 ^a	Cohort 2 ^b	Cohort 1	Cohort 2	Cohort 1	Cohort 2
Writing an objective note	10 min	10 min	15 min	15 min	25 min	25 min
Paired-stimulus preference assessment	60 min	60 min	60 min	50 min	120 min	110 min
Differential reinforcement of alternative behavior	120 min	120 min	60 min	105 min	180 min	225 min
Discrete-trial teaching	300 min	300 min	60 min	105 min	360 min	405 min

^a $n = 18$; ^b $n = 18$

Fig. 1 Data represent the average group (solid line) and individual performance (circles) of Cohorts 1 and 2 during baseline, acquisition, and the first attempt of the skill during the maintenance assessment. The circles represent the performance of each participant immediately before BST, immediately after BST, and during the maintenance assessment. The solid line represents the mean performance of the group immediately before BST, immediately after BST, and during the first attempt of the skill during the maintenance assessment. DTT = discrete-trial teaching; DRA = differential reinforcement of alternative behavior; BL = baseline; ACQ = acquisition; MNT = maintenance

performance increased to 85%–96% correct implementation: 85% (range 50%–100%) for writing an objective note, 93% (range 73%–100%) for conducting a paired-stimulus preference assessment, 93% (range 55%–100%) for the DRA procedure, and 96% (range 80%–100%) for DTT. The low scores of some participants are described further in what follows. Figure 1 also depicts the individual performance of each participant immediately before BST, immediately after BST, and during the first attempt at performing the skill during the maintenance assessment. Fifteen of the 18 participants improved their performance of all four skills immediately after BST compared to baseline. Three participants improved their performance in three of the four skills, but failed to improve in either writing an objective note or DRA.

Although performance within individual participants cannot be evaluated in Fig. 1, we saw several common patterns of responding. For example, Participant 13's (P13) results are displayed in Fig. 2, which is a pattern of responding we saw in 8 of the 18 participants (44%). During baseline, P13 had low to moderately stable levels of correct implementation of the four skills. Following BST, P13 performed each skill at 100% correct implementation, which maintained across time and was demonstrated in a new context (for the paired-

stimulus preference assessment) with an actual client without any additional training or feedback.

A slightly different pattern of responding that we observed across 6 of the 18 participants (33%) is displayed in Fig. 3. These participants failed to maintain the use of one of the four skills during a novel role-play situation or when demonstrating the skill with a client with disabilities. As an example of this pattern of responding, during baseline, P14's performance on the skills was generally low to moderate and stable. The exception was P14's implementation of the DRA procedure at 83% correct on the first baseline probe, but it was on a decreasing trend across baseline. After BST, P14's performance of each skill improved substantially, ranging from 75% to 100% across skills. P14 also met the 90% criterion for writing an objective note, using DRA, and DTT when provided with a novel role-play situation, but P14 did not meet the 90% correct criterion when asked to demonstrate the paired-stimulus preference assessment with a client with a disability. After the trainer provided feedback and waited 24 hr to retest, P14 implemented the paired-stimulus preference assessment at 100% accuracy with a client.

The remaining four participants (22%) were also similar to P14, except that they failed to maintain their use of two or

Fig. 2 Data represent Participant 13's performance (Cohort 1) during baseline, acquisition, and maintenance assessments. Participant 13's data are representative of those trainees who acquired and maintained their use of each skill learned using BST without requiring any additional feedback during training sessions. DTT = discrete-trial teaching; DRA = differential reinforcement of alternative behavior

more of the four targeted skills, requiring feedback and additional assessments on those skills. All of these participants met the 90% criterion on all skills during their second attempt.

Case Demonstration 2: Using Participant Feedback to Refine BST

Method

One way to evaluate whether a large-scale training was successful is, of course, to assess whether each of the trainees acquired the targeted skills. An additional variable of importance, however, relates to the trainees' assessment of the

acceptability of the training methods (Reid et al., 2018). Participants' acceptability of training procedures is thought to be related to the training procedures' overall effectiveness. Selecting training procedures that appeal to trainees may increase the acceptability of those trainings by trainees and administrators (Strohmeier et al., 2014). We asked all 18 participants from Cohort 1 to complete an anonymous paper-and-pencil social validity survey rating the acceptability of the training program goals, training procedures, and outcomes. Twelve participants completed and returned the survey, which consisted of 10 statements to rate using a 5-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*).

Table 3 summarizes the results of the social validity survey from Cohort 1. In general, trainees found the training to be

Fig. 3 Data represent Participant 14's performance during baseline, acquisition, and maintenance assessments. Participant 14's data are representative of those trainees who acquired each skill during BST but required additional feedback to correctly use the skill when working with a client with a disability or within a novel role-play situation. DTT = discrete-trial teaching; DRA = differential reinforcement of alternative behavior

Table 3. Average Rating (With Ranges) for Each Social Validity Questionnaire Item Using a 5-Point Scale

Statement	Cohort 1 ^a	Cohort 2 ^b
1. This was an important training for me to attend.	4.5 (4–5)	4.7 (3–5)
2. I enjoyed the training methods (i.e. lectures, models, role-plays).	3.7 (2–5)	4.6 (4–5)
3. There was enough time allocated to each component of the program.	3.4 (2–5)	4.8 (4–5)
4. The feedback I received improved my use of behavior-analytic skills.	4.5 (3–5)	4.8 (4–5)
5. I liked the way in which the feedback was delivered.	4.2 (3–5)	4.9 (4–5)
6. The training methods were effective in teaching me behavior-analytic skills.	3.8 (3–5)	4.6 (4–5)
7. This training will have a positive impact on my organization.	4.2 (3–5)	4.8 (4–5)
8. I feel comfortable implementing behavior-analytic techniques.	4.1 (3–5)	4.5 (4–5)
9. The training personnel are sufficiently knowledgeable and professional.	4.6 (4–5)	5.0
10. I would recommend this training to a coworker.	4.0 (2–5)	4.9 (4–5)

Note. The scale ranged from 1 (*strongly disagree*) to 5 (*strongly agree*)

^a $n = 12$; ^b $n = 13$

important and felt that the training would have a positive impact on their organization. Additionally, on average, the trainees agreed that they enjoyed the training methods. However, the trainees also reported that there needed to be more training time allocated toward specific skills, as evidenced by an average rating of 3.4 in response to the statement “There was enough time allocated to each component of the program.” Participants were also asked to rank order the four skills from easiest to most difficult to acquire. Trainees reported that DTT and DRA were more difficult to acquire, whereas conducting a paired-stimulus preference assessment and writing an objective note were easier to acquire and required less teaching time. Based on the results of the social validity assessment for Cohort 1, we replicated the procedures detailed previously with an additional cohort of participants (i.e., Cohort 2), adjusting the amount of time allocated to each skill based on the results of the social validity assessment.

Participants and setting Cohort 2 included 18 participants (17 females and 1 male) from the same special education setting as Cohort 1, ranging in age from 21 to 59 years. All participants held a high school diploma and more than half of participants ($n = 12$) held at least a bachelor’s degree. The trainers were the same individuals described previously.

Dependent measures and IOA Measurement of all dependent variables remained consistent with Case Demonstration 1. IOA was collected on 22% of sessions. IOA for writing an objective note, conducting a paired-stimulus preference assessment, implementing a DRA procedure, and using DTT was 95.8% (range 75%–100%), 97.2% (range 85%–100%), 97.1% (range 86.3%–100%), and 94.9% (range 83.0%–100%), respectively.

Experimental design and procedures Cohort 2 participated in BST using the same procedures outlined for Cohort 1, using the same experimental design (i.e., multiple-probe design across skills). The only difference was the amount of time allocated to each skill, based on the social validity surveys. Trainers allocated an additional 1.25 hr to teach DRA (an increase in practice and peer feedback) and an additional 0.75 hr to teach DTT (an increase in practice and peer feedback). The trainers also spent less time on the skill of conducting a paired-stimulus preference assessment, removing 10 min of role-play and practice. The amount of time spent on BST for each skill for Cohorts 1 and 2 is listed in Table 2.

Results

The average performance of Cohort 2 across the four selected skills is depicted in the bottom panel of Fig. 1. Similar to Cohort 1, 2 of the 18 participants in Cohort 2 were not present for the baseline assessment of DRA. As a result, only 16

participants’ data are presented in Fig. 1 for DRA. During baseline, participants’ correct implementation of each skill was 24% (range 0%–75%) for writing an objective note, 31% (range 0%–90%) for conducting a paired-stimulus preference assessment, 71% (range 14%–100%) for the DRA procedure, and 59% (range 0%–92%) for DTT. Following BST for each of the four targeted skills, participants’ average performance increased to 75%–98% correct implementation: 75% (range 25%–100%) for writing an objective note, 98% (range 81%–100%) for conducting a paired-stimulus preference assessment, 98% (range 88%–100%) for the DRA procedure, and 98% (range 94%–100%) for DTT. Sixteen of the 18 participants improved their performance of all four skills immediately after BST compared to baseline. Two participants did not improve their performance on writing an objective note.

Although the average group performance across Cohorts 1 and 2 was similar during the acquisition assessment, the range in scores was smaller for Cohort 2 for the majority of skills. For example, the range of performance of implementing DRA posttraining for Cohort 1 was 55%–100% correct implementation, whereas the range for Cohort 2 was 87.5%–100%. Similar results were found for DTT, with the range of performance on the acquisition assessment decreasing from 83%–100% for Cohort 1 to 94%–100% for Cohort 2.

In contrast to Cohort 1, in which only 44% (8 of 18) of participants met the criterion of at least 90% correct implementation during the first maintenance assessment, in Cohort 2, 16 of 18 (89%) participants achieved this criterion during the first maintenance assessment across all four skills. In Cohort 2, only two participants (11%) had to retest one of the targeted skills 24 hr after receiving feedback (33% for Cohort 1), and no participant had to retest two or more skills (22% for Cohort 1).

Cohort 2 completed the same social validity survey as Cohort 1 ($n = 13$). As is evident in Table 3, the social validity ratings for Cohort 2 were higher across all 10 statements, as compared to Cohort 1’s ratings. Most notably, Cohort 2’s ratings of the statements relating to the enjoyment and effectiveness of large-scale BST were substantially higher compared to Cohort 1. In addition, Cohort 1’s average rating of the statement “There was enough time allocated to each component of the program” was 3.4, whereas Cohort 2 strongly agreed with this statement, with an average rating of 4.8, indicating that there was enough training time allocated to each of the targeted skills.

General Discussion

Competency- and performance-based training is increasingly identified as a model of training for service providers (Parsons et al., 2012). The results of this study support the

use of large-scale BST to teach multiple skills to groups of trainees with a high trainee-to-expert trainer ratio (i.e., 18:1). Trainers utilized didactic instruction, modeling, role-play practice, and performance feedback from peers to instruct two large cohorts of participants of various academic and professional backgrounds on a variety of skills. The majority of participants achieved the prescribed level of competency (i.e., 90% correct implementation or higher) on targeted skills during acquisition and maintenance assessments. We also extended previous research on large-scale BST (e.g., Iwata et al., 2000; Sawyer et al., 2017) by assessing the maintenance of learned skills during a novel role-play or an in vivo assessment in the workplace. Finally, we assessed the acceptability of our training procedures and used the results of those social validity assessments to make adjustments to the training of a future cohort.

A significant contribution of the current study was that we modified BST procedures to accommodate large groups of trainees with only one trainer by substituting peer role-play and feedback for that provided by an expert trainer. When small-scale BST is appropriate for a particular clinical setting (e.g., teaching a small number of skills to a small number of trainees), all components can be implemented as originally prescribed. For example, trainers can deliver brief verbal instructions and check for trainee understanding frequently and efficiently. In addition, expert trainers can easily deliver feedback to each individual trainee during role-play practice. In contrast, implementing BST on a large scale requires trainers to modify some aspect of their delivery of didactic instruction, modeling, behavioral rehearsal, and/or performance feedback to increase the efficiency of the procedures. For example, during the didactic components of training, we utilized instructional methods aimed at increasing engagement, such as providing frequent opportunities for participants to respond in large-group and small-group discussions (e.g., by having participants provide examples from work and answering simple comprehension questions regarding the content) and handing out PowerPoint slides allowing them to follow along and take notes.

Further, conducting role-play practice and providing performance feedback are arguably the most time-consuming BST components, especially for large groups with a high trainee-to-trainer ratio. To address this issue, we had trainees role-play in small groups and provide both positive and corrective feedback to each other during role-plays. Although it may not be considered best practice to have novice trainees provide feedback (as opposed to an expert trainer), our data support trainees' acquisition and maintenance of skills when using this approach. To increase the likelihood that the performance feedback delivered by trainees was accurate, we provided all trainees with a copy of the steps of the task-analyzed skill that matched the demonstration given by the expert

trainer. Based on the results of this study, practitioners who conduct large staff trainings could supplement their didactic trainings with modeling of targeted skills and incorporating opportunities for peers to practice and provide feedback using structured treatment integrity forms.

However, because we did not require all trainees to perform at a certain criterion during peer role-plays, and because we did not monitor the accuracy and quality of feedback that peers delivered, we cannot conclude that all trainees had the same BST experience. For example, there were 5 out of the 36 participants who did not improve their performance of one particular skill after BST. Because we did not monitor how often these five participants were attending during the instructions and whether they had sufficient role-play and feedback opportunities during practice for that particular skill, we cannot identify why their performance did not increase, and it is not known whether they would have increased their performance on this skill if the BST was conducted in a one-on-one setting. Future research evaluating the use of peer feedback during role-plays could incorporate integrity checks on BST components to ensure that all required components occur during training.

In addition to teaching multiple skills to a large trainee audience, we assessed whether trainees could perform their newly acquired skills in a novel role-play situation or when working with a client with a disability. Only a small number of BST demonstrations in the current literature have evaluated whether trainees were able to perform newly trained skills in novel situations (e.g., Fetherston & Sturmey, 2014; Hassan et al., 2018; Homlitas et al., 2014; Lafasakis & Sturmey, 2007; Lerman et al., 2015). Many of the trainees across Cohorts 1 and 2 (67% of trainees) correctly used their newly taught skills in a novel role-play situation or when working with a client. Some trainees, however, did not correctly implement the skill on their first attempt and required feedback before meeting the required performance criterion (12 out of 36 total participants). These differences in trainee performance continue to highlight the need to address individualized trainee needs so that targeted skills will maintain across contexts, while also balancing the needs of the entire trainee group. An additional consideration is that all participants voluntarily consented to the training, perhaps speaking to their motivation to learn new skills that would help their job performance. We do not know how those who did not volunteer would have performed in this hybrid BST model that incorporated both expert trainer and peer feedback. In the future, researchers could evaluate whether a group booster session would result in improved performance across a number of participants who did not demonstrate maintenance of a trained skill.

A limitation of the current study is that we did not have baseline measures on trainee performance of the novel role-play situations or while trainees were working with their clients. As such, we do not know how trainees would have performed the targeted skills within these novel role-plays or with actual

clients prior to training. However, because trainees' performances during the baseline role-play probes were consistently low, it is unlikely that trainees would have performed the skills in these alternate contexts at the performance criterion prior to BST sessions. An additional limitation of our maintenance data is that we do not have any data on the clients' behavior during the paired-stimulus preference assessment demonstrations. Future studies evaluating the effects of BST and generalization of skills taught should include measurement of client behavior and ensure that trainees have opportunities to display all skill components (i.e., responding to both correct and incorrect client responses) during demonstrations.

Lastly, an important component of the case presentations described here included the use of social validity measures to assess the goals, procedures, and outcomes of the training. Based on the results of Cohort 1's social validity assessment, we modified the length of time spent teaching the targeted skills for Cohort 2, and we found that there were substantial differences in the training outcomes of Cohorts 1 and 2. Not only did Cohort 2 rate the acceptability of training procedures higher than Cohort 1 on the social validity assessment, but the lowest performers in Cohort 2 performed better than the lowest performers in Cohort 1. In addition, there was a decreased need for additional training for Cohort 2 participants during the maintenance assessment. These findings support the importance of soliciting trainee feedback to drive training modifications. Specifically, it may be important to survey trainees to identify which skills they find to be the most difficult to acquire so that adjustments can be made to the time allocated to each skill in future trainings. A limitation of the current case presentations, however, is that we do not know whether Cohort 2's performance was better than Cohort 1's as a result of the modifications we made to the training based on Cohort 1's social validity assessment, or perhaps because the trainers got better at training their second time through the procedures. Despite this limitation, Cohort 2's social validity

assessment indicated that there was enough teaching time allocated to each skill, that participants enjoyed the BST training methods, that the training components were effective, and that they would be very likely to recommend this training to their coworkers.

This study adds to the limited literature on the effectiveness of BST on such a large scale (i.e., 18:1 trainee-to-trainer ratio). The results of this study show how a relatively simple modification to the delivery of performance feedback can maintain the initial effectiveness, efficiency, and social validity of BST when transitioning from small-scale to large-scale trainings. In addition to conducting posttraining social validity measures, we would also suggest that trainers survey trainees and administrators to identify which skills are most relevant to the trainees and their everyday work. In the present study, we provided administrators of the school where the trainees worked with the *RBT Task List* (BACB, 2013) and worked with them to identify which skills were most relevant. Identifying skills that are relevant to the trainees' everyday work responsibilities may increase their motivation to acquire these skills more rapidly, especially during sometimes-lengthy large-scale trainings.

Author Note The authors would like to acknowledge Nicole St. Hill for her efforts in helping coordinate training sessions.

Compliance with Ethical Standards

Conflict of Interest The authors have no conflicts of interest to disclose.

Ethical Approval All procedures were reviewed and approved by a university institutional review board.

Informed Consent All participants provided informed consent to be video recorded and for their de-identified data to be used for presentations and publications.

Appendix 1

Task analysis of writing an objective session note:

Step:	Performed correctly mark “+” Performed incorrectly or omitted mark “-”
1. Trainee identifies setting and/or time of day	
2. Trainee identifies individuals present during observation	
3. Trainee objectively describes at least one behavior observed in the video using measurable terms.	
4. Trainee refrains from noting subjective information such as feelings, thoughts, opinions, reports	
Percentage of Steps Implemented Correctly:	

Task analysis of paired stimulus preference assessment:

	Trials Correct (+); Incorrect or omitted (-); Not Applicable (NA)				
	1	2	3	4	5
<i>1. Allows learner to sample items prior to beginning</i>					
2. Establishes attending					
3. Offers a choice of 2 items (at least 6 inches apart) and says “pick one”					
4. If learner reaches for both items, blocks learner’s reach					
4a. Represents choice.					
5. If learner does not respond within 10s, reestablishes attending					
5a. Represents choice					
6. Allows access to chosen item for a few seconds (<1 min).					
<i>7. Identifies item chosen most frequently.</i>					
Percentage of Steps Implemented Correctly:					

Task analysis of differential reinforcement of alternative behavior

Trial	Step	Correct (+); Incorrect or omitted (-); Not Applicable (NA)
1	Delivers reinforcement contingent on the functionally alternative behavior within 5 seconds.	
	Withholds reinforcement contingent on maladaptive behavior	
2	Delivers reinforcement contingent on the functionally alternative behavior within 5 seconds.	
	Withholds reinforcement contingent on maladaptive behavior	
3	Delivers reinforcement contingent on the functionally alternative behavior within 5 seconds.	
	Withholds reinforcement contingent on maladaptive behavior	
4	Delivers reinforcement contingent on the functionally alternative behavior within 5 seconds.	
	Withholds reinforcement contingent on maladaptive behavior	
5	Delivers reinforcement contingent on the functionally alternative behavior within 5 seconds.	
	Withholds reinforcement contingent on maladaptive behavior	
	Percentage of Steps Implemented Correctly:	

Task analysis of discrete trial teaching for receptive identification

Step:	Trials					
	Correct (+); Incorrect or omitted (-); Not Applicable (NA)					
	1	2	3	4	5	6
1. Structures learning environment appropriately						
2. Establish attending						
3. Place 3 stimuli in field						
4. Rotates stimuli before presenting Sd						
Sd						
1. Uses correct Sd (CAN BE VARIED)						
2. Provides clear and concise Sd in statement rather than question format						
3. Provides Sd only once per trial						
Prompts and Prompt Fading						
1. Uses correct prompt identified on lesson plan						
2. Prompt comes after Sd						
Feedback						
1. Delivers immediate reinforcement for correct responses						
2. Provides verbal praise with enthusiasm						
3. Uses correction system identified on lesson plan for incorrect responses						
4. Provides corrective feedback with neutral tone of voice						
5. Records data correctly.						
Other:						
1. Inter-trial interval is 3-5 sec						
Percentage of Steps Implemented Correctly:						

Appendix 2

Paired Stimulus Preference Assessment Datasheet

Item:

1
2
3
4

Trial	Left	Right
1	1	4
2	2	3
3	3	2
4	4	1
5	1	3
6	2	4
7	3	1
8	4	2
9	1	2
10	2	1
11	3	4
12	4	3

Most preferred item: _____

References

- Behavior Analyst Certification Board. (2013). *Registered Behavior Technician (RBT®) task list*. Littleton, CO: Author.
- Behavior Analyst Certification Board. (2014). *Professional and ethical compliance code for behavior analysts*. Littleton, CO: Author.
- Behavior Analyst Certification Board. (2018). *Supervisor training curriculum outline: 2.0*. Littleton, CO: Author.
- Behavior Analyst Certification Board. (2019). *Registered Behavior Technician (RBT®) task list*. Littleton, CO: Author.
- Crockett, J. L., Fleming, R. K., Doepke, K. J., & Stevens, J. S. (2007). Parent training: Acquisition and generalization of discrete trial teaching with parents of children with autism. *Research in Developmental Disabilities*, 28(1), 23–36. <https://doi.org/10.1016/j.ridd.2005.10.003>.
- Fetherston, A. M., & Sturmey, P. (2014). The effects of behavioral skills training on instructor and learner behavior across responses and skill sets. *Research in Developmental Disabilities*, 35(2), 541–562. <https://doi.org/10.1016/j.ridd.2013.11.006>.
- Hassan, M., Simpson, A., Danaher, K., Haesen, J., Makela, T., & Thomson, K. (2018). An evaluation of behavioral skills training for teaching caregivers how to support social skill development in their child with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 48, 1957–1970. <https://doi.org/10.1007/s10803-017-3455-z>.
- Homlitas, C., Rosales, R., & Candel, L. (2014). A further evaluation of behavioral skills training for implementation of the picture exchange communication system. *Journal of Applied Behavior Analysis*, 47, 198–203. <https://doi.org/10.1002/jaba.99>.
- Howard, J. S., Stanislaus, H., Green, G., Sparkman, C. R., & Cohen, H. G. (2014). Comparison of behavior analytic and eclectic interventions for young children with autism after three years. *Research in Developmental Disabilities*, 35(12), 3324–3344. <https://doi.org/10.1016/j.ridd.2014.08.021>.
- Iwata, B., Wallace, M., Kahng, S., Lindberg, J., Roscoe, E., Connors, J., et al. (2000). Skill acquisition in the implementation of functional analysis methodology. *Journal of Applied Behavior Analysis*, 33, 181–194. <https://doi.org/10.1901/jaba.2000.33-181>.
- Lafasakis, M., & Sturmey, P. (2007). Training parent implementation of discrete-trial-teaching: Effects on generalization of parent teaching and child correct responding. *Journal of Applied Behavior Analysis*, 40, 685–689. <https://doi.org/10.1901/jaba.2007.685-689>.
- Lambert, J. M., Bloom, S. E., Kunnavatana, S. S., Collins, S. D., & Clay, C. J. (2013). Training residential staff to conduct trial-based functional analyses. *Journal of Applied Behavior Analysis*, 46, 296–300. <https://doi.org/10.1002/jaba.17>.
- Lavie, T., & Sturmey, P. (2002). Training staff to conduct a paired-stimulus preference assessment. *Journal of Applied Behavior Analysis*, 35, 209–211. <https://doi.org/10.1901/jaba.2002.35-209>.
- Lerman, D. C., Hawkins, L., Hillman, C., Shireman, M., & Nissen, M. A. (2015). Adults with autism spectrum disorder as behavior technicians for young children with autism: Outcomes of a behavioral skills training program. *Journal of Applied Behavior Analysis*, 48, 1–24. <https://doi.org/10.1002/jaba.196>.
- Luiselli, J. K., Bass, J. D., & Whitcomb, S. A. (2010). Teaching applied behavior analysis knowledge competencies to direct-care service providers: Outcome assessment and social validation of a training program. *Behavior Modification*, 34(5), 403–414. <https://doi.org/10.1177/0145445510383526>.
- Luiselli, J. K., St. Amand, C. A., MaGee, C., & Sperry, J. M. (2008). Group training of applied behavior analysis (ABA) competencies to community-based service providers for adults with developmental disabilities. *International Journal of Behavioral Consultation and Therapy*, 4(1), 41–47.
- Madzharova, M. S., Sturmey, P., & Jones, E. A. (2012). Training staff to increase manding in students with autism: Two preliminary case studies. *Behavioral Interventions*, 27(4), 224–235. <https://doi.org/10.1002/bin.1349>.
- Miles, N., & Wilder, D. A. (2009). The effects of behavioral skills training on caregiver implementation of guided compliance. *Journal of Applied Behavior Analysis*, 42, 405–410. <https://doi.org/10.1901/jaba.2009.42-405>.
- Nigro-Bruzzi, D., & Sturmey, P. (2010). The effects of behavioral skills training on mand training by staff and unprompted mands by children. *Journal of Applied Behavior Analysis*, 43, 757–761. <https://doi.org/10.1901/jaba.2010.43-757>.

- Page, T., Iwata, B., & Reid, D. (1982). Pyramidal training: A large-scale application with institutional staff. *Journal of Applied Behavior Analysis*, 15, 335–351.
- Parsons, M. B., Rollyson, J. H., & Reid, D. H. (2012). Evidence-based staff training: A guide for practitioners. *Behavior Analysis in Practice*, 5(2), 2–11. <https://doi.org/10.1007/BF03391819>.
- Parsons, M. B., Rollyson, J. H., & Reid, D. H. (2013). Teaching practitioners to conduct behavioral skills training: A pyramidal approach for training multiple human service staff. *Behavior Analysis in Practice*, 6(2), 4–16. <https://doi.org/10.1007/BF03391798>.
- Reid, D. H., Green, C. W., Parsons, M. B., & Rotholz, D. A. (2018). The best and worst things staff report about behavioral training workshops: A large-scale evaluation. *Behavior Analysis in Practice*. <https://doi.org/10.1007/s40617-018-00297-3>.
- Romanczyk, R. G., Callahan, E. H., Turner, L. B., & Cavalari, R. N. S. (2014). Efficacy of behavioral interventions for young children with autism spectrum disorders: Public policy, the evidence base, and implementation parameters. *Review Journal of Autism and Developmental Disorders*, 1, 276–326. <https://doi.org/10.1007/s40489-014-0025-6>.
- Rosales, R., Stone, K., & Rehfeldt, R. A. (2009). The effects of behavioral skills training on implementation of the picture exchange communication system. *Journal of Applied Behavior Analysis*, 42, 541–549. <https://doi.org/10.1901/jaba.2009.42-541>.
- Roscoe, E. M., & Fisher, W. W. (2008). Evaluation of an efficient method for training staff to implement stimulus preference assessments. *Journal of Applied Behavior Analysis*, 41, 249–254. <https://doi.org/10.1901/jaba.2008.41-249>.
- Sarokoff, R. A., & Sturmey, P. (2004). The effects of behavioral skills training on staff implementation of discrete trial training. *Journal of Applied Behavior Analysis*, 37, 535–538. <https://doi.org/10.1901/jaba.2004.37-535>.
- Sawyer, M. R., Andzik, N. R., Kranak, M. P., Willke, C. P., Curiel, E. S. L., Hensley, L. E., & Neef, N. A. (2017). Improving pre-service teachers' performance skills through behavioral skills training. *Behavior Analysis in Practice*, 10(3), 296–300. <https://doi.org/10.1007/s40617-017-0198-4>.
- Strohmeier, C., Mule, C., & Luiselli, J. K. (2014). Social validity assessment of training methods to improve treatment integrity of special education service providers. *Behavior Analysis in Practice*, 7(1), 15–20. <https://doi.org/10.1007/s40617-014-0004-5>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.