The Psychological Record
https://doi.org/10.1007/s40732-021-00497-4

ORIGINAL ARTICLE

Training Staff to Create Equivalence-Based Instruction Materials in Qualtrics

Kathleen E. Marano-Frezza[1] · Jason C. Vladescu[1] · Kenneth F. Reeve[1] · Tina M. Sidener[1] · David J. Cox[2]

Accepted: 4 November 2021

Abstract
Staff training is an important line of research to ensure that clinicians in the field of applied behavior analysis provide quality services. One approach to providing training involves the use of asynchronous training materials in which the trainer and trainee do not need to be physically present at the same time. This allows for training despite limited numbers of trainers or geographic restrictions. The purpose of this study was to train participants to create equivalence-based instruction (EBI) materials in Qualtrics, a commercial survey software package. In the first phase of the study, participants experienced a training package consisting of a video model and task analysis that described how to create EBI materials. Phase 2, evaluated whether the EBI materials created in Phase 1 led to equivalence-based responding. Results indicated that the training was effective for training staff and that staff found the training socially acceptable. Results also demonstrated that the EBI materials resulted in the learning of the trained and derived relations.

Keywords Equivalence-based instruction · Staff training · Task analysis · Video modeling

Staff trainers in the field of applied behavior analysis can choose from a wide variety of training procedures when teaching skills to staff. When the training procedure requires that the trainer and trainee be present at the same time, the procedure is considered synchronous. However, there are relatively few staff trainers in rural areas and areas outside of the United States, which may make synchronous training difficult or impossible (Gerencser et al., 2019). There is also a relative deficit of trainers compared to the number of staff who require training (Karsten et al., 2015). In addition, the COVID-19 pandemic has caused many service providers to make the transition to telehealth services (Cox et al., 2020), which necessitated the use of more technology-based training procedures. Therefore, synchronous procedures may have limited accessibility for some trainees.

Asynchronous training is an alternative procedure that does not require trainers and trainees to be present at the same time. A variety of asynchronous training procedures are available, including video modeling, computer-based instruction (CBI), and self-instruction in the form of manuals or task analyses (e.g., Gerencser et al., 2019; Gutierrez et al., 2019; Tyner & Fienup, 2016). Some asynchronous training procedures involve a combination of training strategies, including using both video models and self-instructional materials. A review of the available literature on asynchronous training strategies found that self-instructional packages are one of the most effective methods for training staff across a wide variety of skills (Marano et al., 2020).

The existing literature can serve as a guide for researchers because they can apply the techniques to train other skills that have not yet been evaluated in the published literature. For example, the literature on training graph creation can provide direction for training individuals to perform other skills on a computer. Mitteer et al. (2018) taught behavior technicians to create publication-quality graphs using a video model. A subsequent study also demonstrated that video modeling effectively trained participants to input data

This study was conducted in partial fulfillment of a Ph.D. in Applied Behavior Analysis from Caldwell University by the first author, under the supervision of the second author. The authors thank Joseph Marano-Frezza for his help with various aspects of this manuscript.

Kathleen E. Marano-Frezza
katiemarano1@gmail.com

1 Department of Applied Behavior Analysis, Caldwell University, 120 Bloomfield Avenue, Caldwell, NJ 07006, USA

2 Behavioral Health Center of Excellence, Los Angeles, CA, USA

Published online: 27 June 2022

and create graphs (Mitteer et al., 2020). Tyner and Fienup (2015) also contributed to this line of research by comparing video modeling and text-based instruction on the creation of multiple-baseline design graphs. Results showed that video modeling was the more effective technique; text-based instruction in the form of a task analysis also improved performance. Berkman et al. (2019) extended the literature by comparing video modeling with voiceover instructions and enhanced written instructions for training graphing skills. Results indicated that both training procedures were effective. These findings suggest that video models, task analyses, or potentially a combination of both may be effective for training computerized skills (i.e., graph creation). Therefore, similar training techniques may be effective for training other skills that are performed on a computer.

Despite the efficacy of various training strategies for teaching a range of skills that behavior analysts are required to learn, there are some skills that staff training research has not targeted. For example, equivalence-based instruction (EBI) refers to a teaching procedure designed to increase instructional efficiency by incorporating the principles of stimulus equivalence (Stromer et al., 1992). EBI incorporates the principles of stimulus equivalence by teaching individuals to treat physically disparate stimuli as functionally interchangeable. During EBI, the instructor designs training so that some relations are directly taught (e.g., A goes with B), whereas other derived relations emerge without direct teaching (e.g., B goes with A). This makes EBI an economical and generative procedure because some responses emerge in the absence of direct teaching due to the programmed contingencies inherent to EBI (Brodsky & Fienup, 2018). EBI is important to learn because the procedure can lead to expanded repertoires without direct teaching of all components (e.g., learning less vs. more, learning additional language skills; Cooper et al., 2020). Researchers have effectively used EBI to train a variety of skills to adults of typical development, including statistics (Critchfield & Fienup, 2010), interpreting operant functions (Albright et al., 2016), and portion-size estimation (Vladescu et al., 2020). EBI is a viable strategy for teaching individuals with developmental disabilities because it has the potential to optimize instructional efficiency (Arntzen et al., 2010; Rehfeldt, 2011).

Despite the potential benefits of EBI, it does not appear to be widely adopted by clinicians. According to Critchfield et al. (2018), one possible reason EBI is not used is that the majority of EBI research is basic in nature and conducted in laboratory settings targeting skills that are of limited social significance. In addition, the terminology and symbolic notation system used to describe EBI and define equivalence classes may also be unfamiliar to behavior analysts who have limited training in this area. Moreover, the extant EBI literature spans approximately 50 years (originating with Sidman, 1971), which may present a formidable challenge for practitioners to consume in light of numerous competing contingencies. Taken together, these barriers may make it difficult for staff to design EBI procedures themselves, let alone train other staff to use EBI. However, the *Behavior Analyst Certification Board's Fifth Edition Task List* (Behavior Analyst Certification Board, 2017) requires that behavior analysts seeking certification can define and provide examples of derived stimulus relations. This makes EBI an ideal skill for researchers to teach as a way of extending the literature on staff training.

There is currently no research investigating directly training staff to create or administer EBI. However, Blair and Shawler (2020) addressed the lack of training materials available for teaching staff to design EBI by creating a task analysis that describes how to design emergent responding training. Due to the ease of providing instruction using computers, the authors determined that computerized programs offer a viable method for conducting emergent responding training. However, clinicians do not often use computerized EBI because it (1) typically requires substantial monetary and technology-based resources, (2) the technology used to conduct EBI in published research is not widely available, and (3) many instructors lack the skills to program software (Blair & Shawler, 2020). In an attempt to ameliorate these difficulties, Blair and Shawler created a task analysis and tutorial that uses free and easily accessible computer programs. However, they did not provide data on the efficacy of this task analysis. Likewise, Cummings and Saunders (2019) and Cariveau et al. (2020) created tutorials for creating match-to-sample preparations in PowerPoint on laptops and iPads, respectively. Although these procedures may be similar to those required to design EBI materials, no data were provided to demonstrate the efficacy of these tutorials. Additional research is warranted that provides a tutorial for designing EBI materials that also provide efficacy data.

Therefore, the purpose of this study was to extend the staff training literature by teaching adults of typical development to program Qualtrics software (Qualtrics, 2020) to conduct EBI. Qualtrics is a user-friendly computer program that allows researchers and clinicians to create modules that include text, videos, and competency checks during training. The software can also record participant responses and score data as computerized training is delivered. Given the favorable results of Mitteer et al. (2018) and Tyner and Fienup (2015), participants received training via a self-instructional package consisting of a video model and corresponding task analysis. In Phase 1, participants viewed a video model and task analysis that described each step necessary for designing EBI and created EBI materials. In Phase 2, participants experienced EBI using the materials created in Phase 1 to evaluate whether the training materials led to equivalence-based responding.

Phase 1

Participants

Eight participants were recruited through word of mouth from the local community (see Electronic Supplementary Material for participant demographic information). Individuals were provided with general information about the study and interested participants were instructed to contact the experimenter if they were interested in participating. Four males and four females between the ages of 25 and 28 years old participated. All participants identified as middle class and no participants reported any prior experience with using Qualtrics or with EBI. To be eligible for participation, participants were required to have access to a computer, a mobile phone or tablet, a reliable internet connection, and the ability to use Zoom video conferencing software. Participants completed participation in the study within two weeks. Participants completed informed consent forms online using Qualtrics. All experimental procedures were approved by Caldwell University's Institutional Review Board.

Setting and Materials

The primary experimenter conducted all sessions remotely using Zoom, with participants and the primary experimenter remaining in their respective locations. The record feature of the Zoom program was used to record all sessions. During sessions, participants and the experimenter sat at a desk or workstation in a quiet area of their residences. Baseline session materials and stimuli included the participants' computer, a hardcopy of the written protocol, and a list of session stimuli (see Electronic Supplementary Material for the written protocol and EBI stimuli). The written protocol consisted of 10 steps that provided brief descriptions of how to create the EBI materials. Training and posttraining session materials and stimuli included the participants' laptops or computers, an additional tablet or cell phone that was used to view the video model, and a hardcopy of the task analysis (see Electronic Supplementary Material for the task analysis). The task analysis consisted of 277 steps that provided a list of all steps necessary to create the EBI materials. The experimenter determined whether participants had access to a printer via the participant demographic survey. If participants could not print hardcopies of necessary materials (i.e., written protocol, list of session stimuli, task analysis), the experimenter mailed paper copies of the materials to participants prior to the first baseline session.

Participants accessed relevant session materials using a folder shared via Google Drive. The Google Drive folder initially contained individual online computer folders that included the demographic questionnaire, a written protocol, a list of equivalence class stimuli, the relevant Qualtrics information, and folders containing stimuli to be used for EBI. The EBI stimuli were separated into folders for baseline and posttraining stimuli and a folder for training stimuli. Equivalence classes consisted of one picture or symbol and two sequences of letters. The experimenter labeled stimuli as "Stimulus A1," "Stimulus B1," "Stimulus C1," "Stimulus A2," Stimulus B2," Stimulus C2," "Stimulus A3," "Stimulus B3," and "Stimulus C3." Participants used three different sets of EBI stimuli for baseline and posttraining sessions, and used a different set of stimuli for training that were identical to those shown in the training video. The training stimuli consisted of relations between animals (e.g., written word "dog," written word "woof," picture of a dog). The stimuli for baseline and posttraining consisted of arbitrary relations between arbitrary symbols and consonant–vowel–consonant letter sequences consisting of different letters. These sequences were meaningless and designed to be pronounceable, but have no meaning. The purpose of using arbitrary relations between stimuli was to test the efficacy of the training and testing blocks for training adults to demonstrate equivalence-based responding in Phase 2. Using the letter sequences ensured participants had no prior history with any of the relations nor were there likely any common stimulus features that might have facilitated class formation.

Design, Measurement, and Interobserver Agreement

A nonconcurrent multiple-baseline across participants design (Watson & Workman, 1981) was used to assess the effectiveness of using a video model and task analysis to train participants to create EBI programming in Qualtrics. The primary experimenter collected data for each session using data sheets created for the study. The primary experimenter scored participant performance of each step in the task analysis. The experimenter instructed the participants to share their screens prior to using the Qualtrics software. The experimenter then scored participants' responding as they completed each step. Correct responses were defined as the participant completing all components of each step as described in the task analysis. Incorrect responses were defined as participants performing steps differently from the descriptions in the task analysis so that it did not result in the same outcome or omitting steps. The experimenter calculated the percentage of steps implemented correctly during each session by dividing the number of steps implemented correctly by the total number of steps and multiplying by 100. The experimenter also recorded whether participants used the training materials during the posttraining probes (i.e., participants followed the task analysis, or viewed or listened to the video model to complete at least one step). For example, if participants never played the video during a session, but did attend to the task analysis, that session was scored as only using the task analysis.

An independent observer scored a minimum of 33% of sessions across conditions for interobserver agreement (IOA) purposes from the screen recordings of sessions. IOA of participant performance was calculated on a step-by-step basis for each step in the task analysis. An agreement was defined as both raters scoring the performance of a step identically. A disagreement was defined as each rater scoring a step differently. IOA was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100. Mean IOA for Neil was 100%, for Victor was 99% (range: 99%–100%), for Julia was 100%, for Mike was 100%, for Lynn was 100%, for Sarah was 100%, for Jenna was 100%, and for Peter was 97.5% (range: 92.7%–100%).

Duration of training sessions was also measured. The experimenter started a timer after the instruction to begin viewing the materials was given and stopped the timer when participants indicated they finished setting up the Qualtrics software. IOA data were also collected by an independent observer for a minimum of 33% of sessions. IOA of session duration was calculated by dividing the shorter duration by the longer duration and multiplying by 100. IOA of session duration indicated 100% agreement across participants.

Setting Up Equivalence-Based Instruction in Qualtrics

Stimulus Preassessment

To ensure that none of the arbitrary stimuli had any unidentified common features that may affect stimulus class formation, 10 adults who did not participate in the study completed a stimulus preassessment. The adults were recruited from an applied behavior analytic graduate program and were different from those who participated in the study to ensure exposure to the stimuli would not influence responding during the study. For each set of arbitrary stimuli, individuals were presented with the nine stimuli and asked to sort the stimuli into groups. The individuals did not receive any information regarding the number of groups or number of stimuli in each group. If 30% of the individuals had grouped multiple stimuli together, those stimuli would have been replaced. This did not occur, resulting in all 27 arbitrary stimuli meeting the criteria for inclusion in the study. In addition, individuals were asked to tact each arbitrary symbol to test for prior familiarity (i.e., individuals were asked "What does this look like to you?"). If 30% of more of the individuals provided the same tact for a stimulus, that stimulus would have been replaced. Because this did not occur, all nine arbitrary symbols in the preassessment were included in the study.

EBI

Participants programmed the Qualtrics program for EBI with three equivalence classes containing three members each. EBI was designed to provide training according to a one-to-many training structure (Fields et al., 1999; Fields & Verhave, 1987) and a simple-to-complex protocol (Albright et al., 2016; Green & Saunders, 1998). The one-to-many training structure and simple-to-complex training protocols were selected because they are among the most effective methods for EBI (Adams et al., 1993; Arntzen et al., 2010). The one-to-many training structure involves pairing one stimulus (e.g., the A stimuli) with other stimuli (e.g., AB, AC) and then testing the other relations. The simple-to-complex protocol requires that participants respond correctly to all trials (questions) for one trained relation (e.g., AB) before the corresponding derived relation (e.g., BA) is tested. Participants created a total of seven surveys in Qualtrics (hereafter referred to as training and testing blocks in accordance with the terminology consistent with EBI research). Because the Qualtrics software uses the terminology "surveys," "questions," and "answer choices," those terms were used for participant materials. The seven training and testing blocks consisted of 3-trial training blocks for the AB and AC relations, 3-trial testing blocks for the BA and CA equivalence relations, a 6-trial testing block that evaluated the emergence of the BC and CB relations, and 18-trial testing blocks for the pretest and posttest that evaluated all relations. For example, the AB training block consisted of three trials that paired the A stimuli with the B stimuli. The training blocks (AB and AC relations) involved programming Qualtrics to provide automated feedback for correct and incorrect selections. Participants set up the program to randomize the presentation order of questions and order of comparison stimuli (e.g., McPheters et al., 2021). That is, the order in which sample stimuli appeared throughout presentations of trials was randomized, as well as the location of the three comparison stimuli beneath the sample stimulus.

General Procedure

Prior to the first session, the experimenter emailed participants a link to the consent form and participant demographic questionnaire. After these forms were completed, the experimenter shared a link to the Qualtrics software and a link to a Google Drive folder containing necessary session materials. Then, the experimenter used the screen-sharing feature on Zoom to show the participants how to navigate the Google Drive folder that was shared, including accessing and downloading the relevant session stimuli. While clicking on the various folders, the experimenter said:

> This folder contains the materials you will need to set up the Qualtrics program. You will use specific sets of stimuli during each session. If you click on the folder for "Baseline/Posttraining Stimuli," you will see that it contains three folders of different stimuli. If you click on the folder for "Set 1," you will see files for three pictures. There is also a document that lists the stimuli that go together in equivalence classes. An equivalence class refers to stimuli that look different, but are treated the same way, meaning that these stimuli go together. In this document, you will see that all stimuli in Class 1 go together, all stimuli in Class 2 go together, and all stimuli in Class 3 go together. You will also see that all pictures are labeled as the A stimuli, the three-letter words are labeled as the B stimuli, and the four-letter words are labeled as the C stimuli. You will set up the software based on these classes. You also should have the printed sheet that lists the different classes and stimuli that go together. This information is identical to the information in the Google Drive folder. I will tell you which set of stimuli to use during each session and you can use the sheet to help you remember which stimuli go together.

Baseline sessions began with the experimenter instructing participants to share their screens while they set up EBI programming in Qualtrics. Then, participants attempted to set up the Qualtrics software to facilitate instruction based on an EBI paradigm. During baseline and posttraining sessions, the experimenter set a maximum time of 45 min for participants to program the Qualtrics software. Consistent with prior staff training research (e.g., Lipshultz et al., 2015), the time was based on the longest time it took three adults with prior training in using Qualtrics software who did not participate in the study to program the EBI materials using the same stimuli. If participants were engaging in correct responses after the time period elapsed, they were permitted to continue setting up the Qualtrics program until they indicated they were finished.

Following baseline sessions, participants practiced using the Qualtrics software while viewing a video model and task analysis that described how to complete each step. Prior to the first training session, the experimenter also shared a folder containing the task analysis and video model. The purpose of delaying sharing these folders was to prevent participants from accessing the training materials during baseline. After completing the training, participants again set up programming using the baseline stimuli.

Baseline

Baseline sessions required participants to attempt to set up EBI programming in Qualtrics using the baseline/posttraining stimuli shared with them through Google Drive and a written protocol. The experimenter said:

> Now you will use Qualtrics to set up an equivalence-based instruction program. Equivalence-based instruction, or EBI, is used to teach people that things that seem different actually go together. For example, a picture of a horse goes with the written word horse, and the sound neigh. You will set up software to train people that various stimuli go together. Please open the Google Drive folder I shared with you. Please share your screen with me by clicking on the "Share Screen" button on the bottom of the screen. Do your best to create an equivalence-based instruction program in Qualtrics using the stimuli in the "Baseline/Posttraining" folder labeled Set _____. Each class contains one picture and two written words that go together. When you set up the software, you will need to upload the pictures and type in the words that you see in the folder and on the sheet in front of you. Please hold up the written protocol you printed out so I can make sure you have the correct materials. You can also use that written protocol to help you. You can use the printed sheet that lists the different classes and stimuli that go together to help you remember which stimuli go together. Again, you will use the stimuli in the "Baseline/Posttraining" folder labeled Set _____. Can you repeat back to me which stimuli you will be using for this session? You can log into Qualtrics using the link, username, and password in the file labeled "Qualtrics Information" or written on the written protocol. Please let me know when you are finished. You will have up to 45 minutes to complete this. I cannot answer any questions at this time.

The experimenter did not provide any feedback to participants as they set up the Qualtrics software. Participants moved to the training condition after steady state responding was achieved.

Training

Following baseline, participants received training on how to use the Qualtrics software via a video model and accompanying task analysis. Each session consisted of participants completing the full task analysis. The task analysis described each step participants should complete to set up the Qualtrics software and corresponded to the steps depicted in the video model (see Electronic Supplementary Material for link to the video). Video duration was 54 min 13 s. The video included voiceover instructions that described each step in the task analysis. The voiceover and steps depicted matched the steps listed in the written protocol participants used during baseline sessions. For example, the first step required participants to upload the stimuli into Qualtrics. The task analysis listed the buttons participants should click to upload stimuli. The video showed a screen recording of the experimenter uploading the stimuli while the voiceover indicated which buttons to click. The video followed this format for all steps in the task analysis. The experimenter said:

> I have shared a new folder with you on Google Drive labeled "Training Materials." Please access the folder on both your computer and cellphone or tablet to print out the task analysis. Then, use your cellphone or tablet to access the video I emailed you. You will watch the video telling you what to do using this device. You can pause, play, and rewind the video as many times as you would like. I suggest holding your device sideways and clicking the arrows in the top left corner to expand your screen because this makes it easier to pause and play the video. You can use the corresponding task analysis and list of stimuli you printed out that describes each step. Please hold the task analysis you printed out up to the camera so I can make sure you have the correct materials. You will again use Qualtrics to set up an equivalence-based instruction program using the stimuli on Google Drive and on your sheet, but you will use the stimuli labeled "Training" this time. Each class of stimuli that go together will again consist of one picture and two words. Please share your screen with me by clicking on the "Share Screen" button on the bottom of the screen. Do your best to create an equivalence-based instruction program. Please let me know when you are finished. I cannot answer any questions about how to set up the software at this time.

Participants could ask clarification questions, but did not receive any feedback on their responding. The training condition continued until participants performed all steps correctly during one session.

Posttraining

After completing the training session, participants again used the baseline session stimuli to set up Qualtrics for EBI. Similar to Mitteer et al. (2018), participants still had access to the video model and task analysis during posttraining sessions. Therefore, posttraining sessions followed the same procedures as training sessions, but used the same stimuli as baseline sessions. In order to complete the posttraining phase, participants were required to complete all steps of the task analysis correctly with all three sets of EBI stimuli. The experimenter read the following script during each session:

> Please access the Google Drive folder on both your computer and cellphone and make sure you have a printed copy of the task analysis. You can continue to use the video in the Google Drive folder and the printed task analysis for this session if you would like. You can pause, play, and rewind the video as many times as you would like. You can also use the list of stimuli you printed out that lists all the stimuli in each class. You will again use Qualtrics to set up an equivalence-based instruction program using the stimuli on Google Drive and on your sheet, but you will use the stimuli labeled Set ____ in the "Baseline/Posttraining folder" this time. Can you repeat which stimuli you will be using? Each class of stimuli that go together will again consist of one picture and two words. Please share your screen with me by clicking on the 'Share Screen' button on the bottom of the screen. Do your best to create an equivalence-based instruction program. Please let me know when you are finished. I cannot answer any questions at this time.

Social Validity

Immediately after the completion of the posttraining phase, participants completed an adapted version of the Interventions Rating Profile-15 (Martens et al., 1985) to evaluate their satisfaction with the procedures, goals, and outcomes of the study. Participants rated their satisfaction using a 5-point Likert-type scale, in which higher scores indicated higher satisfaction. Participants completed the social validity questionnaire in Qualtrics to ensure responding was anonymous.

Procedural Integrity

A trained observer also scored procedural integrity of the experimenter's behavior during a minimum of 33% of sessions across all phases. The observer calculated the percentage of correctly implemented steps by dividing the number of steps implemented correctly by the total number of steps and multiplying by 100. All procedural integrity scores across participants indicated 100% of steps were implemented correctly. A second observer also scored procedural integrity for a minimum of 33% of procedural integrity sessions for IOA. An agreement was defined as both observers scoring a step identically, whereas a disagreement was defined as each observer scoring a step differently. Step-by-step agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100. Mean interobserver agreement was 100% across participants.

Results and Discussion

Figures 1 and 2 depict the results of Phase 1. During baseline, all participants engaged in low levels of correct responding. No participants completed more than 30% of steps correctly during any baseline sessions: Neil (*M* = 2.2%, range: 1.9%–2.7%), Victor (*M* = 1.9%, no range), Julia (*M* = 1.2%, range: 0.7%–1.5%), Mike (*M* = 0.8, range: 0.4%–1.9%), Lynn (*M* = 2.6, range: 1.9%–3%), Sarah (*M* = 1.4%, range: 0.7%–1.9%), Jenna (*M* = 26.2%, range: 11.8%–29%), Peter (*M* = 5.8%, range: 3.8%–12.2%). During baseline, all participants correctly logged into Qualtrics, but Jenna was the only participant to correctly input picture files and stimuli names into the Qualtrics software. No participants correctly formatted questions (e.g., randomizing questions) or set up scoring.

Correct responding increased during the training phase so that all participants completed all steps correctly with the stimuli that were identical to those in the training video. All participants completed the training in approximately 1 hr (*M* = 57 min 41 s, range: 51 min 5 s–60 min 7 s; (see Electronic Supplementary Material). During posttraining sessions, all participants correctly completed all steps using the stimuli that were used during baseline sessions. Following training, all participants' mean duration for creating EBI materials decreased (*M* = 35 min 42 s, range: 24 min 12 s–55 min 34 s). A calculation of quantitative effect size using the web-based single-case design hierarchical linear model application (Pustejovsky, 2016; Valentine et al., 2016) demonstrated a large effect size between baseline and posttraining performance ($d = 14.4$; 95% confidence intervals = 9.18–19.64). The effect size was calculated using a restricted maximum likelihood estimation method and level baseline and treatment time trends.

The results of the social validity survey also demonstrated that participants found the training procedure acceptable (*M* = 5) and would like to learn other skills using similar training procedures (*M* = 5). Participants also reported the time required to learn the skill was reasonable (*M* = 4.8, range: 4–5) and they were confident in their ability to create EBI materials in Qualtrics (*M* = 4.6, range: 4–5). Therefore, the results demonstrate that the video modeling and task analysis package was an effective and socially valid procedure for training participants to set up Qualtrics for EBI.

The overall mean training time of approximately 1 hr for all participants suggests that a training package consisting of a video model and written task analysis may be efficient for creating EBI materials. In addition, mean training time decreased for all participants following training. Although the training materials were not systematically faded, participants were not required to use both the video model and task analysis following training. Therefore, multiple participants chose to use only the written task analysis following training. Neil, Mike, Jenna, and Peter correctly completed all steps without using the video model. Victor and Lynn reported that they primarily used the task analysis, but used the video model for specific steps only (e.g., randomizing the order of questions in the training question blocks). Julia and Sarah continued to use both the video model and the task analysis during posttraining. This suggests that participant responding may maintain at high levels in the absence of all training materials. Future researchers may choose to systematically fade the training materials following training to evaluate the conditions under which responding maintains.

The results of Phase 1 support and extend previous staff training literature by demonstrating that an asynchronous training package is effective for training computer-based skills (i.e., creating EBI materials). Mitteer et al. (2018) evaluated the use of a video model to train staff to create graphs and found that it was effective. Tyner and Fienup (2016) compared the use of a video model and a task analysis to train staff to create graphs, and found that both techniques improved responding, but the video model was more effective. The present study extends this line of research by showing that a package consisting of a video model and task analysis is effective for training staff to create EBI materials.

This study was the first to evaluate a procedure related to training EBI skills to staff. Although the extent literature includes a tutorial for creating such materials (Blair & Shawler, 2020), no published studies have included data demonstrating the efficacy of training. Data are needed to demonstrate that a training procedure results in the creation of high-quality EBI materials. Therefore, the results of Phase 1 add to the literature by showing that the training procedure was effective for creating EBI materials. However, additional research is needed to demonstrate that the EBI materials created lead to responding indicative of equivalence class formation. This was addressed in Phase 2.

Phase 2

The purpose of Phase 2 was to evaluate whether the EBI training procedures created by participants from Phase 1 would lead to equivalence-based responding. To do so, we evaluated whether four adults of typical development demonstrated equivalence-based responding after completing the Qualtrics EBI created by a subset of participants from Phase 1.

Fig. 1 Percentage of Correctly Completed Steps (Phase 1)

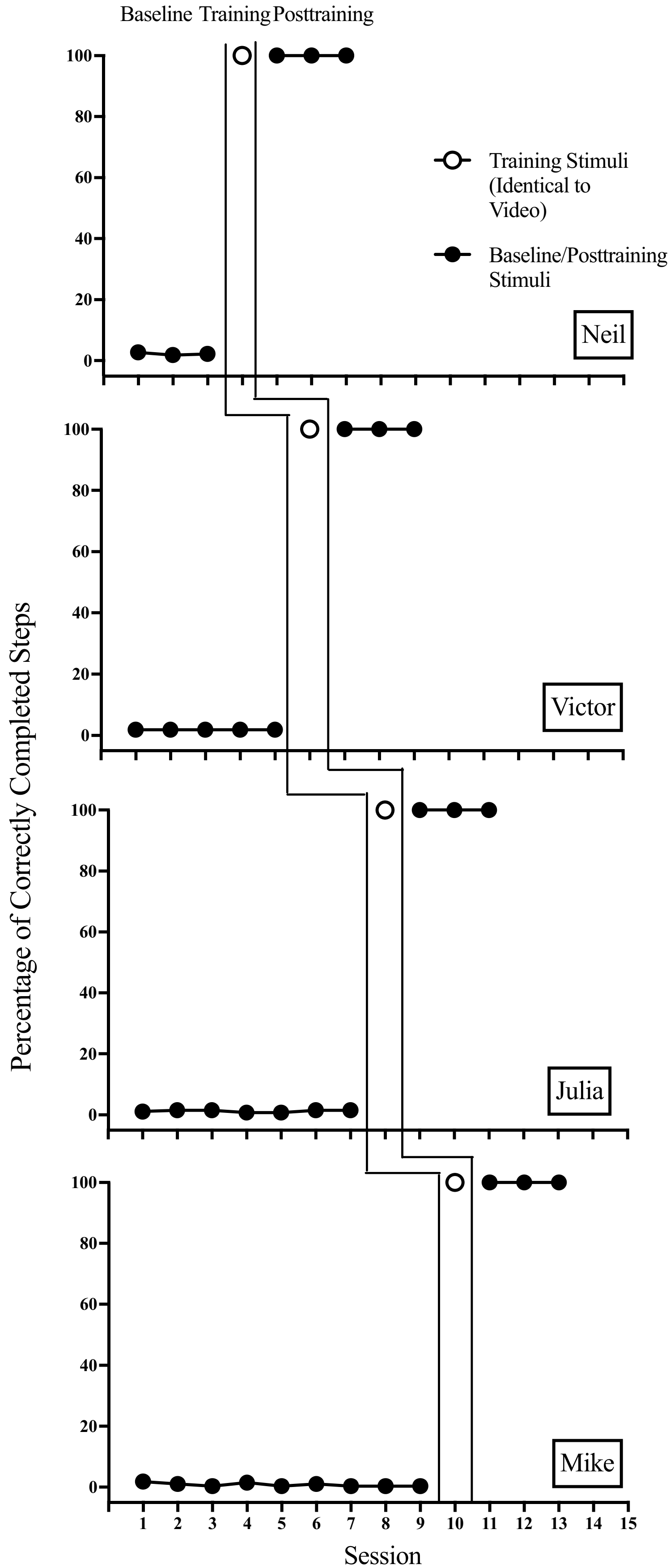
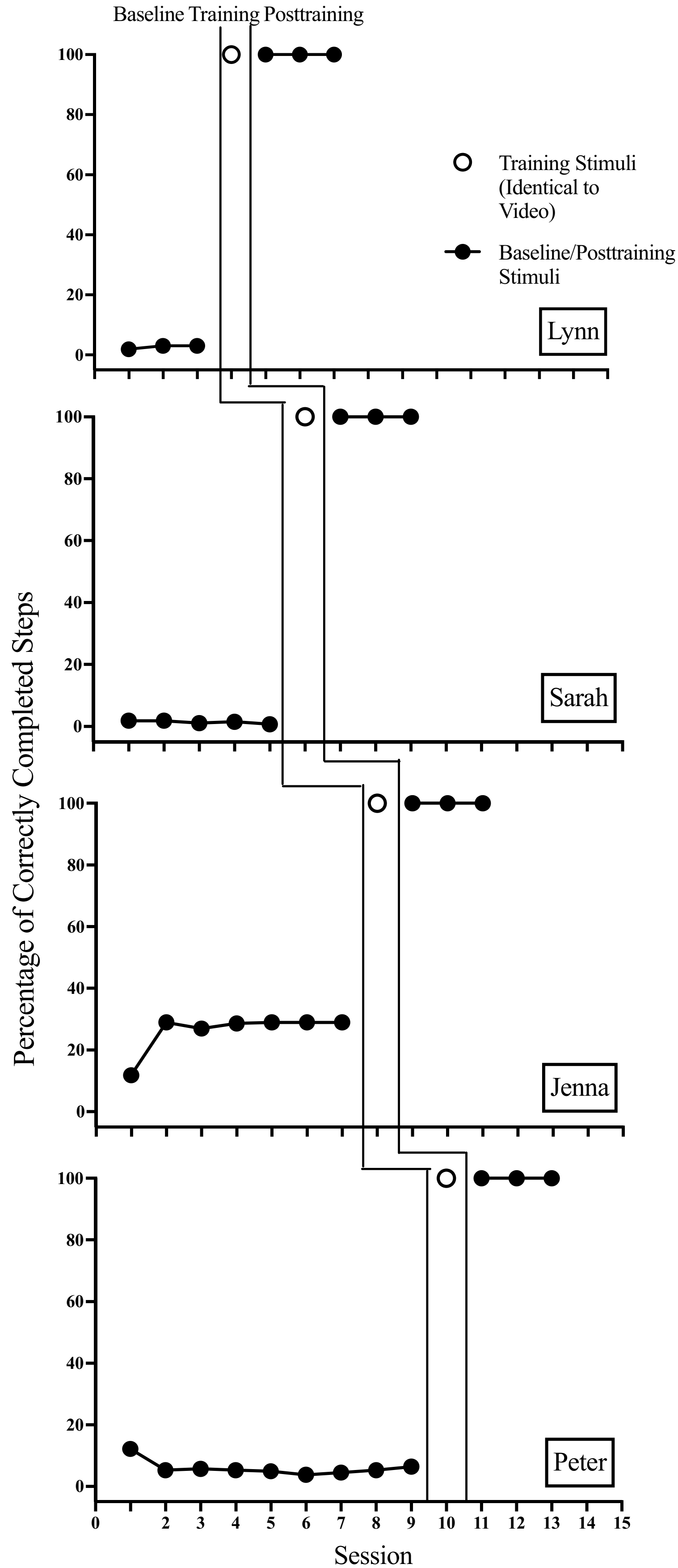


Fig. 2 Percentage of Correctly Completed Steps (Phase 1)



Participants

The training and testing blocks created by a small subset (i.e., four) of the participants in Phase 1 were used in Phase 2. Materials met the inclusionary criteria for the second phase of the study if 100% of steps were completed correctly during Phase 1. We randomly selected Mike, Victor, Neil, and Jenna's surveys using the materials from Stimuli Set 1. Participants in Phase 2 used these surveys to evaluate whether the materials led to the demonstration of equivalence-based responding.

The experimenters recruited four adults of typical development via word of mouth to participate in Phase 2 (see Electronic Supplementary Material for participant demographic information). All participants were graduate students in a Master of Arts in Applied Behavior Analysis program who identified as female. Sam was 24 years old, Caucasian, low socioeconomic status, and reported that she was familiar with EBI procedures from coursework, but had never received training regarding using EBI. Leah was 31 years old, Hispanic, low socioeconomic status, and had no prior experience with EBI. Jane was 24 years old, Hispanic, low socioeconomic status, and was familiar with EBI procedures, but never received training regarding using EBI. Jessica was 25 years old, Caucasian, middle socioeconomic status, and reported minimal experience with using EBI. Jessica was the only participant with prior Qualtrics experience, but reported that she did not feel comfortable using the software. Because the purpose of Phase 2 was to evaluate whether the training and testing blocks led to equivalence-based responding among arbitrary stimuli with which participants had no prior exposure, participants were included despite prior EBI familiarity or Qualtrics experience. Participation in the study required one session lasting approximately 1 hr. Completion of the study made participants eligible for extra credit in applicable courses.

Setting and Materials

Participants and the experimenter used their own laptops or computers and a cellphone or tablet to communicate via Google Meet from their respective locations. During sessions, participants joined the Google Meet session using their laptops and additional devices (e.g., cellphone, tablet) that were angled so that participants' hands and faces were in view. The purpose of angling the camera to allow for more visibility and joining via multiple devices was to allow the experimenter to view participants' eyes and hands to ensure they did not have access to any notes or off-screen methods to aid in responding. Participants and the experimenter sat at a desk or workstation in a quiet area of their residences and used laptops and headphones for all sessions. Participants accessed relevant session materials (i.e., the Qualtrics survey) via email. The training and testing blocks used in Phase 2 consisted of the equivalence classes from Stimuli Set 1 that participants used during the baseline and posttraining phases from Phase 1. The record feature of Google Meet was used to record sessions.

Design, Measurement, and Interobserver Agreement

A nonconcurrent multiple baseline across participants design was used to assess the effectiveness of the Qualtrics EBI program for training adults of typical development to demonstrate equivalence-based responding. Participant responses to the Qualtrics materials served as the dependent variable. The Qualtrics software automatically recorded the responses of the participants as they completed training and testing blocks. Correct responses were defined as the participants selecting stimuli that corresponded with the three experimenter-defined equivalence classes. Incorrect responses were defined as the selection of stimuli that did not correspond with the experimenter-defined equivalence classes. Data were summarized as the percentage of correct responses overall for each session, and for each relation by dividing the number of correct responses by the total number of questions and multiplying by 100. For example, if a participant responded correctly on one of the three trials that assessed the AB relation, data were summarized as 33% correct for the AB relation. IOA data were not collected because Qualtrics automatically recorded the option selected by participants. Thus, these data were not subject to an analysis of IOA.

Procedure

Baseline

During the baseline condition, participants completed the pretest section of the EBI training created in Qualtrics by engaging in selection responses with their computer mouse (see Electronic Supplementary Material for an example of a trial). Because a multiple-baseline across participants design was used, each participant completed the pretest a different number of times until steady state responding was achieved. Each session consisted of 18 trials designed to test which stimuli were part of the same classes. The experimenter instructed participants to share their screens and complete the pretest by saying:

> Please share your screen with me by clicking on the "Share Screen" button on the bottom of your screen. Please click on the Qualtrics survey labeled "Pretest." Do your best to complete the Qualtrics survey by choosing which stimuli you think go together.

> You will select which stimulus you think matches by clicking on the button next to that stimulus. Then, you will click the red box with the arrow to move to the next question. I cannot answer any questions at this time.

Participants then completed the pretest in the absence of any feedback.

EBI

After the baseline condition was complete, participants received EBI using the Qualtrics software that was programmed in Phase 1. Three equivalence classes containing three members each were taught. Training followed a one-to-many training structure and followed a simple-to-complex protocol (Green & Saunders, 1998). Participants completed the training block that trained the AB relation, which consisted of three trials (i.e., A1 as the sample stimulus, A2 as the sample stimulus, and A3 as the sample stimulus) and automatically provided feedback for all correct and incorrect responses. Next, participants completed the block of three trials that tested the BA symmetry relation in the absence of feedback. If symmetry was demonstrated, participants then completed the block of three trials that trained the AC relation, followed by the three trial CA symmetry relation test block, and the block of six trials to test for the emergence of the BC and CB equivalence relations. To move between trial blocks, participants had to respond correctly to all questions in a trial block during two consecutive sessions. If participants did not respond correctly to all questions, they repeated the training for the corresponding relation until they responded correctly to all questions two consecutive times. For example, if participants responded incorrectly during the BA trial block, they completed the AB training block again at least twice. If participants responded incorrectly during the CA trials block, they completed the AC training block again at least twice. If participants responded incorrectly during the BC and CB trial block, they repeated all previous training and testing blocks at least twice (i.e., AB, BA, AC, and CA). Participants had to correctly respond to all questions in each block twice to complete training. The experimenter said:

> Please share your screen with me by clicking on the "Share Screen" button on the bottom of your screen. Please click on the Qualtrics survey labeled ____. Do your best to select the stimulus that you think goes with the stimulus that is presented. You may not take any notes while completing this. The faster you answer all trials correctly, the faster you will complete training. I cannot answer any questions at this time.

Posttest

After participants responded correctly to all questions in the training portion of EBI, they completed the posttest. The posttest followed the same format as the pretest, in which participants responded to all questions for all relations in the absence of feedback. To ensure steady state responding, participants were required to complete the posttest three times.

Social Validity

After completing the posttest, participants completed an adapted version of the Interventions Rating Profile-15 (Martens et al., 1985) in Qualtrics to evaluate satisfaction with the training procedures and outcomes of the study. Participants rated their satisfaction using a 5-point Likert-type scale, in which higher scores indicated higher satisfaction.

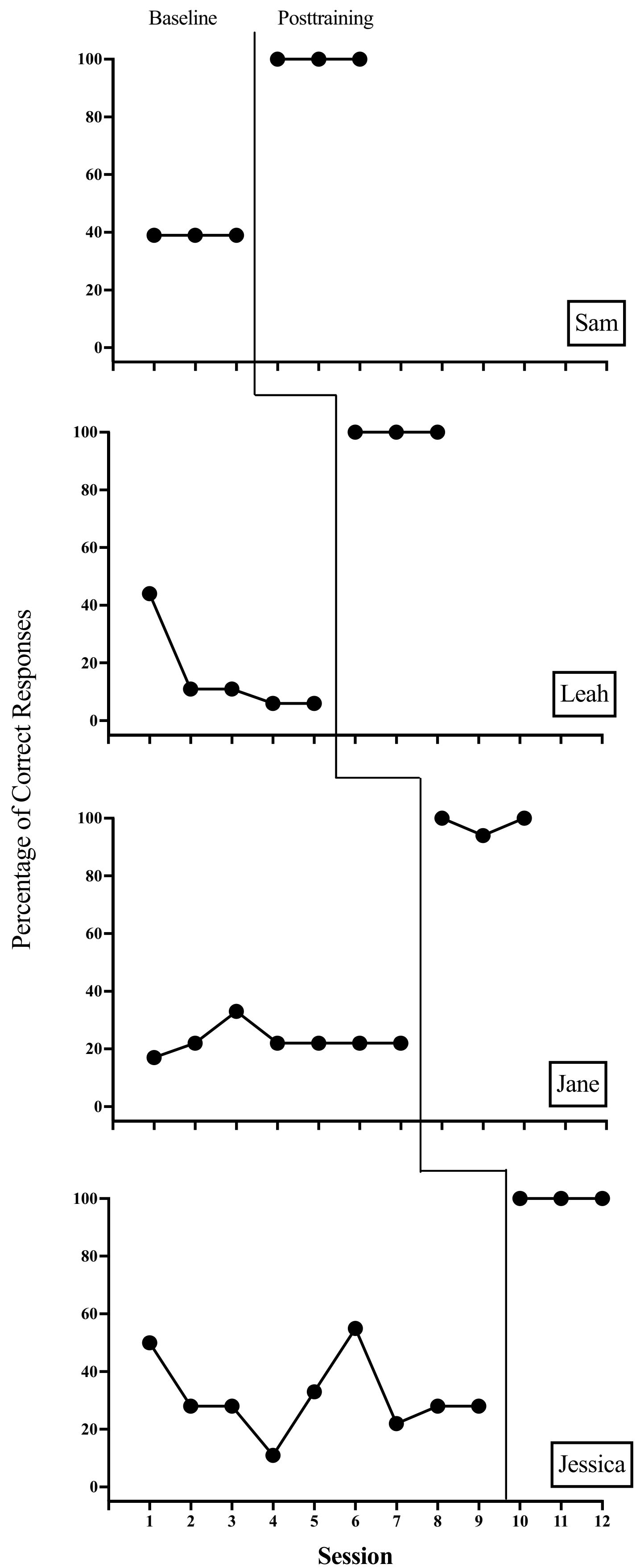
Procedural Integrity

An independent observer collected procedural integrity data on the experimenter's behavior during a minimum of 33% of sessions. The observer calculated the number of correctly implemented steps by dividing the number of correctly implemented steps by the total number of steps and multiplying by 100. Procedural integrity for all participants across all conditions was 100%. A second observer also scored procedural integrity for a minimum of 33% of procedural integrity sessions for IOA. An agreement was defined as both raters scoring a step identically, whereas a disagreement was defined as each observer scoring a step differently. Step-by-step agreement was calculated by dividing the number of agreements by the number of agreements plus disagreement and multiplying by 100. IOA data indicated 100% agreement for all participants across all conditions.

Results and Discussion

Figure 3 depicts the percentage of correct responses during baseline and posttraining for all participants. During baseline, all participants responded correctly during a low percentage of trials. Following EBI, all participants demonstrated a nearly perfect percentage of correct responding. In particular, Sam, Leah, and Jessica responded correctly to all questions in the posttest (100% correct responding), whereas Jane responded correctly during all but one trial during the posttest ($M = 98$, range: 94%–100%). A calculation of quantitative effect size using the web-based single-case design hierarchical linear model application (Pustejovsky, 2016; Valentine et al., 2016) demonstrated a large effect size between baseline and posttraining performance ($d = 6.4$; 95% confidence intervals = 4.23–8.58). The effect size was

Fig. 3 Percentage of Correct Responses (Phase 2)



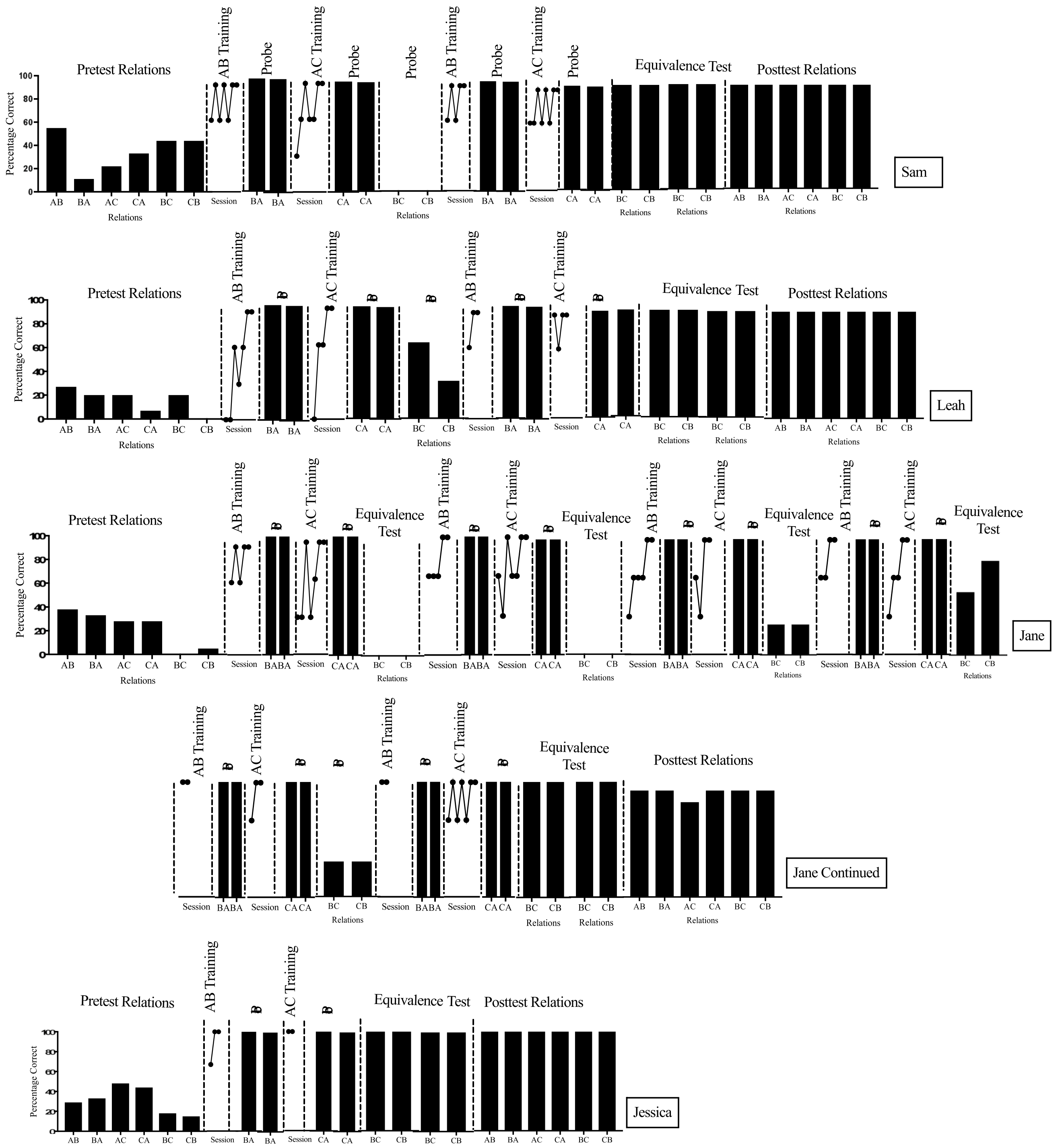


Fig. 4 EBI Graphs (Phase 2)

calculated using a restricted maximum likelihood estimation method and level baseline and treatment time trends.

During the EBI pretest, all participants responded below 60% correct responding for all relations (see Fig. 4). Participants required between zero and five rounds of remedial training with the training and testing blocks. Throughout EBI, all participants demonstrated emergence of all symmetrical relations. Sam and Leah repeated the training and testing blocks once before they demonstrated emergence of the BC and CB equivalence relations, whereas Jessica demonstrated the equivalence relations on her first attempt, and Jane repeated the training and testing block sequence five times before demonstrating emergence of the relations. During the posttest, Sam, Leah, and Jessica demonstrated 100% unprompted correct responding for all relations during all posttests, whereas Jane demonstrated 100% correct responding during two posttests (i.e., correct responding for all relations) and 94% correct responding during one posttest (i.e., correct responding for five of the six relations).

Social validity results also indicate participant satisfaction with EBI. Most participants indicated that the time required to learn the material was reasonable ($M = 5$), they reported confidence in their ability to identify the stimuli that went together ($M = 4.5$, range: 4–5), they would recommend this form of training to other people ($M = 4$, range: 1–5), and they indicated they would like to learn other skills using the procedure ($M = 3.75$, range: 1–5).

During EBI, all participants learned the symmetrical relations without direct teaching. However, responding to the equivalence relations test block and the amount of training required varied across participants. Only one participant (Jessica) responded correctly for the equivalence relations test block on her first attempt. Jessica also had the most prior experience with EBI (experience with setting up the software for matching to sample tasks). Therefore, prior learning history and exposure to EBI may have facilitated faster learning. Two participants (Sam and Leah) repeated the series of training and testing blocks twice before they responded correctly to all trials in the equivalence testing block. One participant (Jane) repeated the series of training and testing blocks six times before she responded correctly to all trials. It is possible that including additional trials to test the AB and AC relations in the absence of any feedback or additional trials within each training and testing block (e.g., including each AB relation twice within a training block) may have reduced the need for remedial training. Despite the need for remedial training, all participants completed participation in the study in 90 min or fewer, indicating that EBI is an efficient procedure due to the learning of some relations without direct training.

General Discussion

Overall, the results of Phases 1 and 2 demonstrate that adults with no prior experience with Qualtrics or EBI can learn to set up EBI, and the materials they created led to the demonstration of equivalence-based responding. Although substantial research exists on the development of stimulus equivalence classes and EBI (Fienup et al., 2015), no study to date had evaluated how to train staff to create or conduct EBI. The present study addresses this gap in the literature by training participants to create EBI materials and showing that the materials they created led to equivalence-based responding.

In addition, the social validity results also indicate that the training was satisfactory for the majority of participants. All participants in Phase 1 reported that the training was beneficial and they would like to learn other skills using similar procedures. The social validity results of Phase 2 also indicate that EBI is a favorable training procedure, although one participant provided lower ratings. That participant indicated that they would not recommend this form of training to others and would not want to learn other skills using EBI. This is likely due to the requirement to complete remedial training until mastery was achieved. However, the other participants indicated that they would recommend EBI to others and would like to learn other skills using the procedure.

The study also has potential benefits for graduate students in applied behavior analysis specifically. Given that the Behavior Analyst Certification Board®□ is including information about derived stimulus relations on future certification examinations, it is becoming increasingly important for behavior analysts to be familiar with EBI procedures (Critchfield et al., 2018). In addition, the third edition of Cooper et al. (2020) also includes a chapter related to EBI, indicating that skills related to EBI are an area of increasing attention within the field.

In addition, despite the potential for EBI to optimize instructional time by resulting in derived relations, EBI is infrequently used in clinical settings due to complicated terminology and the majority of research being basic in nature (Critchfield et al., 2018). Therefore, a possible avenue for exposing practitioners to EBI procedures is to include it within the graduate training for behavior analysts. By teaching graduate students who are studying applied behavior analysis about EBI and how to use it, they may be more likely to actually use the procedure in clinical settings. The Qualtrics software that participants learned to use during this study is also a valuable tool for graduate students, given its accessibility in university settings, ability to score data, and applicability for both research and clinical uses. Therefore, the study extends the research on EBI by evaluating a training package that staff can use to learn to create EBI.

The results also extend the previous staff training literature in a number of ways. First, the results support the previous literature demonstrating the efficacy of asynchronous training procedures for training staff to implement behavioral technologies. Conducting research on asynchronous training strategies is particularly important given the difficulty with providing training in rural areas (Gerencser et al., 2019) and lack of trainers relative to the number of staff who require training (Graff & Karsten, 2012). Gerencser et al. (2019) conducted a literature review that identified experiments that used asynchronous training techniques to train staff to implement skills relevant for human service staff working with individuals with developmental disabilities. The authors determined that video models and self-instructional packages are commonly used training procedures that are effective for training staff. In a literature review of all staff training studies using asynchronous techniques, Marano et al. (2020) found similar results. Additional studies also found that asynchronous training procedures can effectively teach computer-based skills such as graph creation (Mitteer et al., 2018; Tyner & Fienup, 2016). The present study extends the previous research by demonstrating the utility of a video-based self-instructional package for training a different skill that was not addressed in previous research.

One potential limitation of this study is that the training materials were not systematically faded during the posttraining condition in Phase 1. Although participants had continued access to the materials, only two participants chose to continue to use both the video model and task analysis. Four participants used only the task analysis and two participants used primarily the task analysis, but used the video model for specific steps. This may reflect how individuals use resources in the natural environment. For example, when completing an online task, an individual may use "Help" features of the software to read a list of steps or watch a brief video describing necessary steps. Therefore, having continued access to such materials is likely representative of real-world environments in which individuals would likely create EBI materials. In addition, fading may not be necessary because (1) trainees can access these materials without the need for a trainer and (2) trainee responding may initially be controlled by the materials through rule-governed behavior. However, behavior may become controlled by more naturally occurring contingencies as staff continue to use Qualtrics. Therefore, behavior may transition from being rule-governed to contingency-controlled.

Another potential limitation is that the training procedures were not compared to alternative methods to train staff to set up EBI (e.g., written instructions alone, BST). Therefore, the efficacy and efficiency of the procedures compared to other training procedures cannot be determined. Future researchers should consider comparing the procedures of the present studies to alternative training methods. The study also incorporated an online synchronous component in which the experimenter reviewed how to use Google Drive with the participants. It is possible that this component of the study would not be necessary due to the task analysis detailing how to navigate Google Drive and download the necessary stimuli. Future researchers may choose to omit this component of the training procedure to provide an entirely asynchronous training. In addition, the training procedure requires that participants have a wi-fi connection. This may be a limitation in that those in rural areas or those of a lower socioeconomic status may lack this resource.

The study also provides direction for future research involving EBI, as it is the first study to train a skill related to conducting EBI. Although the results demonstrate that the training package taught participants to create EBI, there are still multiple areas related to EBI that should be trained and empirically evaluated. For example, the present study trained participants to set up Qualtrics for classes consisting of three visual stimuli. Future researchers could train participants to create EBI with larger class sizes or auditory stimuli. In addition, the equivalence classes in the present study were previously constructed by the experimenters; future research could train participants to create their own equivalence classes. Participants in future research could also learn information about EBI in general. For example, the participants in the present study learned to set up Qualtrics, but did not receive direct training on the various characteristics of EBI (e.g., nodes, training structures, training protocols). It is important that behavior analysts learn about the various options for using EBI, learn to read relevant research, and learn to apply the literature to the creation of EBI procedures. Therefore, future researchers should continue to extend this line of research by evaluating packages that train other aspects of EBI.

The present study demonstrates the utility of Qualtrics software both for training staff to create EBI materials and for training equivalence classes. This study supports previous research demonstrating the utility of Qualtrics software for EBI (McPheters et al., 2021). Although the Qualtrics software is applicable for use with EBI, additional researchers should consider evaluating similar procedures with alternative popular quiz software (e.g., Quizzizz or Survey Monkey). In addition, researchers may evaluate different EBI training structures (e.g., many-to-one) and training protocols (e.g., complex-to-simple). The majority of participants in Phase 2 required remedial training to learn the equivalence relations. Although prior research indicates that the one-to-many training structure is effective for adults of typical development (e.g., Arntzen, 2004), other researchers have demonstrated that the many-to-one training structure may also produce improved stimulus equivalence test performance (e.g., Hove, 2003). In addition, the complex-to-simple may demonstrate failures with emergent relations earlier in training.

Researchers should also evaluate the training package used in the present study with other skills. The results of the study demonstrate that a video model and task analysis training package is effective for training staff to create EBI. However, this type of training package may be applicable for training a variety of other skills. For example, researchers used video models to train the implementation of stimulus preference assessments (e.g., Lipshultz et al., 2015), but have not yet assessed incorporating a detailed task analysis in combination with a video model. Therefore, future researchers should evaluate the training procedure with other types of skills.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40732-021-00497-4.

Data Availability The data that support the findings of this study are openly available at https://github.com/katiemaranofrezza/Training-Staff-to-Create-Equivalence-Based-Instruction-Materials-in-Qualtrics

Declarations

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

References

Adams, B. J., Fields, L., & Verhave, T. (1993). The effects of test order on the establishment and expansion of equivalence classes. *The Psychological Record, 43*(4), 133–152. https://doi.org/10.1007/BF03395907

Albright, L., Schnell, L. K., Reeve, K. F., & Sidener, T. M. (2016). Using stimulus equivalence-based instruction to teach graduate students in applied behavior analysis to interpret operant functions of behavior. *Journal of Behavioral Education, 25*(3), 290–309. https://doi.org/10.1007/s10864-016-9249-0

Arntzen, E. (2004). Probability of equivalence formation: Familiar stimuli and training sequence. *The Psychological Record, 54*(2), 275–291. https://doi.org/10.1007/BF03395474

Arntzen, E., Halstadtro, L., Bjerke, E., & Halstadtro, M. (2010). Training and testing music skills in a boy with autism using a matching-to-sample format. *Behavioral Interventions, 25*(2), 129–143. https://doi.org/10.1002/bin.301

Behavior Analyst Certification Board. (2017). *BCBA/BCaBA task list* (5th ed.).

Berkman, S. J., Roscoe, E. M., & Bourret, J. C. (2019). Comparing self-directed methods for training staff to create graphs using GraphPad Prism. *Journal of Applied Behavior Analysis, 52*(1), 188–204. https://doi.org/10.1002/jaba.522

Blair, B. J., & Shawler, L. A. (2020). Developing and implementing emergent responding training systems with available and low-cost computer-based learning tools: Some best practices and a tutorial. *Behavior Analysis in Practice, 13*(2), 509–520. https://doi.org/10.1007/s40617-019-00405-x

Brodsky, J., & Fienup, D. M. (2018). Sidman goes to college: A meta-analysis of equivalence-based instruction in higher education. *Perspectives on Behavior Science, 41*(1), 95–119. https://doi.org/10.1007/s40614-018-0150-0

Cariveau, T., Hunt, K., & McCord, M. (2021). Recommendations for using PowerPoint 2016/2020 to create individualized matching to sample sessions on the iPad. *Behavior Analysis in Practice, 14*(1), 161–165. https://doi.org/10.1007/s40617-020-00484-1

Cooper, J. O., Heron, T. E., & Heward, W. L. (2020). *Applied behavior analysis* (3rd ed.). Pearson Education.

Cox, D. J., Plavnick, J. B., & Brodhead, M. T. (2020). A proposed process for risk mitigation during the COVID-19 pandemic. *Behavior Analysis in Practice, 13*(3), 299–305. https://doi.org/10.1007/s40617-020-00430-1

Critchfield, T. S., & Fienup, D. M. (2010). Using stimulus equivalence technology to teach statistical inference in a group setting. *Journal of Applied Behavior Analysis, 43*(4), 763–768. https://doi.org/10.1901/jaba.2010.43-763

Critchfield, T. S., Barnes-Holmes, D., & Dougher, M. J. (2018). Editorial: What Sidman did—Historical and contemporary significance of research on derived stimulus relations. *Perspectives on Behavior Science, 41*(1), 9–32. https://doi.org/10.1007/s40614-018-0154-9

Cummings, C., & Saunders, K. J. (2019). Using Powerpoint 2016 to create individualized matching to sample sessions. *Behavior Analysis in Practice, 12*(2), 483–490. https://doi.org/10.1007/s40617-018-0223-2

Fields, L., & Verhave, T. (1987). The structure of equivalence classes. *Journal of the Experimental Analysis of Behavior, 48*(2), 317–332. https://doi.org/10.1901/jeab.1987.48-317

Fields, L., Hobbie, S. A., Reevem, K. F., & Adams, B. J. (1999). Effects of training directionality and class size on equivalence class formation by adults. *The Psychological Record, 49*(4), 703–724. https://doi.org/10.1007/BF03395336

Fienup, D. M., Wright, N. A., & Fields, L. (2015). Optimizing equivalence-based instruction: Effects of training protocols on equivalence class formation. *Journal of Applied Behavior Analysis, 48*(3), 613–631. https://doi.org/10.1002/jaba.234

Gerencser, K. R., Akers, J. S., Becerra, L. A., Higbee, T. S., & Sellers, T. P. (2019). A review of asynchronous trainings for the implementation of behavior analytic assessments and interventions. *Journal of Behavioral Education, 29*(1), 122–152. https://doi.org/10.1007/s10864-019-09332-x

Graff, R. B., & Karsten, A. M. (2012). Evaluation of a self-instruction package for conducting stimulus preference assessments. *Journal of Applied Behavior Analysis, 45*(1), 69–82. https://doi.org/10.1901/jaba.2012.45-69

Green, G., & Saunders, R. R. (1998). Stimulus equivalence. In K. A. Lattal & M. Perone (Eds.), *Handbook of research methods in human operant behavior* (pp. 229–262). Plenum Press.

Gutierrez, J., Reeve, S. A., Vladescu, J. C., DeBar, R. M., & Giannakakos, A. R. (2019). Evaluation of manualized instruction to train staff to implement a token economy. *Behavior Analysis in Practice, 13*, 158–168. https://doi.org/10.1007/s40617-019-00386-x

Hove, O. (2003). Differential probability of equivalence class formation following a one-to-many versus a many-to-one training structure. *The Psychological Record, 53*(4), 617–634. https://doi.org/10.1007/BF03395456

Karsten, A. M., Axe, J. B., & Mann, C. C. (2015). Review and discussion of strategies to address low trainer to staff ratios. *Behavioral Interventions, 30*(4), 295–313. https://doi.org/10.1002/bin.1420

Lipshultz, J. L., Vladescu, J. C., Reeve, K. F., Reeve, S. A., & Dipsey, C. R. (2015). Using video modeling with voiceover instruction to train staff to conduct stimulus preference assessments. *Journal of Developmental & Physical Disabilities, 27*(4), 505–532. https://doi.org/10.1007/s10882-015-9434-4

Marano, K. E., Vladescu, J. C., Reeve, K. F., Sidener, T. M., & Cox, D. J. (2020). A review of the literature on staff training strategies that minimize trainer involvement. *Behavioral Interventions, 35*(4), 604–641. https://doi.org/10.1002/bin.1727.

Martens, B. K., Witt, J. C., Elliott, S. N., & Darveaux, D. X. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology: Research & Practice, 16*(2), 191–198. https://doi.org/10.1037/0735-7028.16.2.191

McPheters, C. J., Reeve, K. F., Fienup, D. M., Reeve, S. A., & DeBar, R. M. (2021). Effects of preliminary class membership on subsequent stimulus equivalence class formation. *Journal of the Experimental Analysis of Behavior*. Advance online publication. https://doi.org/10.1002/jeab.650

Mitteer, D. R., Greer, B. D., Fisher, W. W., & Cohrs, V. L. (2018). Teaching behavior technicians to create publication-quality, single-case design graphs in GraphPad Prism 7. *Journal of Applied Behavior Analysis, 51*(4), 998–1010. https://doi.org/10.1002/jaba.482

Mitteer, D. R., Greer, B. D., Randall, K. R., & Briggs, A. M. (2020). Further evaluation of teaching behavior technicians to input data and graph using GraphPad Prism. *Behavior Analysis: Research and Practice, 20*(2), 81–93. https://doi.org/10.1307/bar0000172

Pustejovsky, J. E. (2016). scdhlm: A web-based calculator for between-case standardized mean differences (Version 0.3.1) [Web application]. Retrieved from: https://jepusto.shinyapps.io/scdhlm. Accessed 1 Nov 2020.

Qualtrics. (2020). *Academic overview*. https://www.qualtrics.com/. Accessed 1 May 2020.

Rehfeldt, R. A. (2011). Toward a technology of derived stimulus relations: An analysis of articles published in the *Journal of Applied Behavior Analysis*, 1992–2009. *Journal of Applied Behavior Analysis, 44*(1), 109–119. https://doi.org/10.1901/jaba.2011.44-109

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech & Hearing Research, 14*(1), 5–13. https://doi.org/10.1044/jshr.1401.05

Stromer, R., Mackay, H. A., & Stoddard, L. T. (1992). Classroom applications of stimulus equivalence technology. *Journal of Behavioral Education, 2*(3), 225–256. https://doi.org/10.1007/BF00948817

Tyner, B. C., & Fienup, D. M. (2015). A comparison of video modeling, text-based instruction, and no instruction for creating multiple baseline graphs in Microsoft Excel. *Journal of Applied Behavior Analysis, 48*(3), 701–706. https://doi.org/10.1002/jaba.223

Tyner, B. C., & Fienup, D. M. (2016). The effects of describing antecedent stimuli and performance criteria in task analysis instruction for graphing. *Journal of Behavioral Education, 25*(3), 379–392. https://doi.org/10.1007/s10864-015-9242-z

Valentine, J. C., Tanner-Smith, E. E., Pustejovsky, J. E., & Lau, T. S. (2016). Between-case standardized mean difference effect sizes for single-case designs: A primer and tutorial using the scdhlm web application. *Campbell Systematic Reviews, 12*(1), 1–31. https://doi.org/10.4073/cmdp.2016.1

Vladescu, J. C., Marano, K. E., & Reeve, K. F. (2021). Equivalence-based instruction to improve portion-size estimation with different vessels. *The Psychological Record, 71*(1), 179–188. https://doi.org/10.1007/s40732-020-00388-0

Watson, P. J., & Workman, E. A. (1981). The non-concurrent multiple baseline across-individuals design: An extension of the traditional multiple baseline design. *Journal of Behavioral Therapy & Experimental Psychiatry, 12*(3), 257–259. https://doi.org/10.1016/0005-7916(81)90055-0