Effects of Video Modeling and Embedded Instructions to Teach Preference Assessment Procedures: A Comparison

by

Deborah Hatton

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

Individual Interdisciplinary Studies

Departments of Psychology and Disability Studies

University of Manitoba

Winnipeg

Acknowledgements

The author wishes to acknowledge her enthusiastic participants; and for graciously providing interobserver reliability, Keevan Brockie, Janet Wong, and Ryan Heckert; Johnathan Hatton for technical and emotional support; Timothy Hatton and Nathan Hatton for their acting, filming, and enduring patience; my family for considerable moral support; and especially my advisor, C.T. Yu, for sharing his vast wisdoms, insights, and guidance; and my committee members: Dr. Nancy Hansen, Dr. Toby Martin, and Dr. Shahin Shooshtari for their long standing patience and abiding encouragement.

Dedications

To Peter for his inspiration, love, patience, and support for time and all eternity.

To Timothy and Nathan for being there always and whenever.

To my committee members, Dr. Nancy Hansen, Dr. Toby Martin, and Dr. Shahin Shooshtari, for their inspiration, and for unreservedly giving me a strength of purpose through the thick and thin of it all.

Especially to my exceptional advisor, Dr. C.T. Yu, without whom I would not have made it, who graciously supported me through sickness and in health, for better and for worse, and even through retirement.

Table of Contents

Tables and Figures

Appendices

Abstract

Stimulus preference assessments are important procedures for identifying preferences with people diagnosed with intellectual disabilities and/or autism who are nonverbal. Research has shown that video modeling plus instruction was effective in teaching mediators to conduct preference assessment. The purpose of this study was to compare two ways to embed instructions into video modeling, namely, on-screen caption and voice-over instructions. Sixteen undergraduate university students participated. Participants were randomly assigned to view videos to teach paired-stimulus procedure with either on-screen or voice-over instructions and videos to teach multiple-stimulus-without-replacement procedure with either on-screen or voice-over instructions. All sessions were conducted online without in-person contact due to the COVID-19 pandemic. Participants' performance was measured during baseline, at post-training, and at a one-week follow-up. During each phase, participants were shown a video of the experimenter conducting a preference assessment with a confederate, with procedural errors deliberately built into the videos, and participants were asked to identify as many procedural errors as possible. Analysis of variance at post-training showed that the difference between the on-screen and voice-over instruction groups was not statistically significant. Analysis of covariance was also not statistically significant when baseline performance was controlled. However, there was a statistically significant gain in performance between baseline and post-training, and between baseline and follow-up in both voice-over groups and on-screen groups. In a post-study feedback survey, participants felt that the goal of the study was important and felt that the videos were easy to follow but overestimated the effectiveness of the videos. Factors influencing the effectiveness of the intervention in the present study are discussed.

Effects of Video Modeling and Embedded Instructions to Teach Preference Assessment Procedures: A Comparison

Stimulus preference assessment is a widely used procedure to allow people with severe intellectual disabilities including autism spectrum disorders (ASD) with no speech, to make choices and to identify their preferences (Autism Teaching Supports; Fisher et al., 1992; Piazza et al., 1996). The procedure has also been used successfully for other populations such as those with dementia and other memory impairments (Lucock et al., 2020; Virués-Ortega et al., 2012). The assessment is carried out by presenting choices to a person a number of times and asking the person to select their preferred item. The frequency with which an item is selected throughout an assessment reflects the degree of preference, with the most frequently selected item being the most preferred (Fisher et al.,1992; Roscoe et al., 1999). This procedure is important because preferred items work as reinforcers by strengthening desirable behaviour in teaching situations for people who may not be able to communicate their preferences in conventional ways. Moreover, learning to conduct preference assessments is included as one of the skills required for a behaviour analyst to become certified (Behavior Analyst Certification Board). Given the importance of the procedure, considerable research has taken place on evaluating effective and efficient methods to teach practitioners and care-givers to apply the procedure. One particular teaching strategy focuses on the use of video modeling (Higgins et al., 2017; Nottingham et al., 2017). In this approach, the video may be accompanied by instructions presented as written text on-screen (Higgins et al., 2017; Rosales et al., 2015; Weldy et al., 2014), or spoken vocally in a voice-over format (Delli Bovi et al., 2017; Marano et al., 2020). Although both types of instructions have been used in video modeling, the relative effectiveness of each is unknown. Examining the effects of the on-screen and voice-over components is important to streamline the

procedure (Ward-Horner & Sturmey, 2010). Therefore, the present study sought to compare the relative effectiveness of the on-screen (caption) video instruction and the voice-over (verbal) video instruction in the context of using video modeling to teach participants to apply one of the two different assessment procedures.

In the ensuing sections, stimulus preference assessment procedures and research on teaching using self-instructional manuals and video modeling are reviewed. This is followed by a statement of the problem, a description of the present study, its findings, and a discussion of the results.

Stimulus Preference Assessments

Daily, we make many choices that reflect our preferences for how we dress, what we eat, which activities we participate in, and which chair we will sit in, just to name a few. Our preferences enhance our lives by creating a more pleasant lifestyle and give us control and independence (Blair et al., 1999). However, there are many people who are not able to easily make their partialities known. Some people with moderate to severe intellectual disabilities may not be able to vocalize, may not be easily understood, and/or may not be accustomed to being allowed to share their preferences. Thus, many assumptions by the people working and living with them are constructed and the person in question is thus managed without any respective input (Goodley & Roets, 2008). Often people assume there are things everyone likes yet these assumptions may be incorrect. The problems with these faulty assumptions are that they could curtail options (why bother giving choices if they cannot make them) and result in using ineffective stimuli/items as reinforcers in teaching (Blair et al., 1999).

Some negative behaviours may even be unknowingly reinforced by social attention (Narozanick & Blair, 2019), or result in providing leisure activities that may be boring, or even worse, disliked (e.g., “most people like shopping, so let’s go to the mall!”). Preferred items are

routinely used by parents, schoolteachers, and other care-givers to teach individuals new skills. Children and adults alike may be presented with a favoured object or activity upon successful completion of a task. Thus, modifying classroom activities viewed as negative to a preferred chosen one, teaching more favourable replacement skills, and changing behavioural consequences to a preferred activity can increase a positive behaviour (Cho & Blair, 2017).

Life can be so difficult with some children with ASD that parents know no other way and any intervention can be seen as an intrusion to their already challenging situation; however, when home and school can work together significant changes can be observed (Lane et al., 2016). Blair et al. (2011) noted using increased novel activities, along with promoting choice for activities and materials and in giving frequent verbal praise, that the children's positive behaviours were increased both at home and at school.

Furthermore, many people assume individuals with severe intellectual disabilities who are nonverbal cannot make choices or that they do not have real partialities (Goodley & Runswick-Cole, 2012). However, preference assessment research has shown this not to be true (Clay et al., 2013; Fava & Strauss, 2010; Hagopian et al., 2004; Piazza et al., 1996; Sellers et al., 2013). Moreover, providing choices alone (e.g., choosing between two low preference items) appears to give a sense of control and thus is also associated with reduced problem behaviours (Romaniuk & Miltenberger, 2001). In some cases, regardless of the highly preferred preferences, choice can be the reinforcer in and of itself (Sellers et al., 2013). Similarly, in 2012, Keyl-Austin et al., found a choice of three medium rated alternating favourites would outlast one highly preferred choice as a reinforcer, perhaps as the point of satiation was not easily reached.

The solution lies in introducing preference assessments. Presenting a preference assessment will determine a person's preferences in the order in which they favour each item

with a rating from high to low. A preference assessment would not only benefit the person directly by improving their enjoyment and quality of life, but will also be of benefit to their parents, care-givers, practitioners and/or professionals who work with them on a regular basis (Blair et al., 1999; Fisher et al., 1992; Piazza et al., 1996). They would be better able to teach a new skill by presenting preferred items as reinforcers to promote learning.

Furthermore, people's preferences may change over a period of time so that one's favourite snack this month may be salty potato chips, while a couple of months later a sweet gummy bear may do the trick. This requires preference assessments to be done periodically to ensure the items being used are indeed having a reinforcing effect. Thus, it is essential for care-givers, parents, teachers, and practitioners to be able to administer these assessments. In the next section, research on the different types of preference assessments are reviewed in more detail.

Indirect Stimulus Preference Assessments

Of the two general types of preference assessments, indirect and direct, the indirect preference assessments consist of data gathered and reported by another person such as a parent, care-giver, or teacher. This may involve oral questions and answers, interviews either structured or casual, or by using written checklists. The assessor then assumes this reported information to be correct and proceeds with training using these described items as positive reinforcers. However, previous research has shown when indirect preferences are compared with direct preferences (discussed below) there is little correlation, thus indicating parents, care-givers and teachers are not necessarily aware of their charges' favourites.

Direct Stimulus Preference Assessments

Direct preference assessments require firsthand observation of preferences being shown in approach or engagement responses (Higbee et al., 2000). There are several approach-based

methods including: a single-stimulus procedure (Pace et al., 1985); the paired-stimulus procedure (Fisher et al., 1992); multiple-stimulus presentation procedure (e.g., Carr et al., 2000; Paramore & Higbee, 2005; Windsor et al., 1994); multiple-stimulus-without-replacement (DeLeon & Iwata, 1996); free-operant stimulus preference assessment (Roane et al., 1998); and response restriction stimulus preference assessment (Hanley et al., 2003).

Single-stimulus Preference Assessment. When stimuli are presented one at a time it is known as single-stimulus preference assessment. Individuals are timed as each item is presented. As the allotted time expires the item is removed. Preference is rated on the percentage of trials the person took, ate, or approached the presented item before it was taken away (Pace et al., 1985).

Paired-stimulus Preference Assessment. An assortment of stimuli presented two at a time is the main crux of the paired-stimulus (PS) preference assessment. Often from an initial selection of between six to eight items, two are presented on each trial and each item is paired with every other item during the assessment. One of the two items is chosen within a given time frame after which the other item is removed, and the stimuli are reset to present the next pair of choices. Preference is rated on the percentage of trials where each particular stimulus is chosen across the trials in which it is presented (Fisher et al., 1992; Hagopian et al, 2004). This assessment has strong reliability; however, it takes longer to administer than other preference assessments, as it requires more trials.

Multiple-stimulus Preference Assessment. During a multiple-stimulus preference assessment, all stimuli, usually six to eight items, are placed in a row in front of the individual. One item is chosen within a given time frame after which all items are removed, and the stimuli are reset with all the stimuli presented in a different order. This continues until all variations of

sequences have been presented. As with the PS, preference is rated on the percentage of trials where each particular stimulus is chosen across the trials in which it is presented (Hagopian et al., 2004; Windsor et al., 1994). Multiple-stimulus requires less time than PS; however, its test-retest reliability is lower. There is also a tendency for a person to select the same preferred item in the array on all trials, thus producing no preference information about all items.

Multiple-stimulus-without-replacement Preference Assessment. A multiple-stimulus-without-replacement (MSWO) preference assessment is similar to multiple-stimulus preference assessment except that after the initial choice has been made it is removed from the presented line-up of stimuli. Thus, the preliminary array has eight items, the chosen stimulus is consumed or otherwise removed, the remaining stimuli are reordered and re-presented. This next array holds seven stimuli, followed by an array of six stimuli, and so on until a choice has been made from a final presented pair of stimuli. Once again preference is rated on the percentage of trials where each stimulus is chosen across the trials in which it is presented (Hagopian et al., 2004). As with multiple-stimulus, MSWO requires less time than PS and its test-retest reliability is reported to be more reliable than multiple-stimulus and almost as reliable as PS (Hagopian et al., 2004).

Free-operant Stimulus Preference Assessments. During a free-operant stimulus preference assessment the full array of stimuli are displayed, usually eight items, such that the participant is free to engage with any item. Items are displayed for a specific time, typically five minutes, before being removed. The time the participant engages with each item is used to indicate the degree of preference (Roane et al., 1998). Free-operant has the lowest statistical correlation coefficients and Kendall rank coefficients (indicates the degree of correspondence across all replications) (Verriden & Roscoe, 2016) for stimulus preference assessments.

Response Restriction Stimulus Preference Assessments. This is a hybrid between MSWO and free-operant. The stimulus items are presented to the participant who is free to engage with any item as with free-operant, but with a 2-minute limit. Before a second presentation, the top ranked item is removed from the array, and so on during subsequent trials similar to that of MSWO (Hanley et al., 2003). The response restriction stimulus preference assessment proved to provide a better hierarchy of preferences than the free-operant, however it was time consuming (Hanley et al., 2003). Comparisons between free-operant, response restriction, single-stimulus, MSWO, and PS were studied by Verriden and Roscoe (2016) who found MSWO and PS to show the highest statistical correlations between each other's results for the highest preference, followed by single-stimulus and response restriction, with free-operant displaying the lowest statistical Kendall rank coefficients correlations.

Preference Assessment Factors

There are several issues that need to be considered while administering preference assessments to ensure accurate results.

Satiation and Deprivation. Aspects of the participants' environments such as satiation and deprivation must be considered. An individual will not respond well if tested with edibles just after being satiated with a filling meal, nor will they respond well if tested with toys after being satiated during a play period. Likewise, if a participant has not eaten for a few hours, they may react to every edible presented. If a person has not had an opportunity to play or to be outside for a time and is thus deprived in these manners, every toy and physical activity presented may be respectively desired.

Operational Definition of Behaviour

It is vital the choice response is well defined. Not all participants will be able to physically pick up an edible and put it in their mouth or be able to respond to a toy by actually playing with it. Depending on the individual, the behaviour to be accepted must be clear and distinct. Perhaps a head nod or a smile or a glance in the direction of the object may be delineated as a means of identifying preference for people with limited physical movement.

Applicability of Preference Assessments for Other Populations

Other than people with intellectual disabilities, preference assessments have been used with increasing frequency with people with memory issues. For example, studies have included people with diagnoses such as vascular dementia (Lucock et al., 2020; Virués-Ortega et al., 2012), Alzheimer's disease, (Lucock et al., 2020; Virués-Ortega et al., 2012), and mild cognitive syndrome (Virués-Ortega et al., 2012). Raetz et al. (2013) included participants with moderate and severe dementia. LeBlanc et al. (2006) included one participant having been diagnosed with severe aphasia and right-side hemiparesis due to a prior stroke, a second participant with probable Alzheimer's type dementia, multi-infarct dementia, and traumatic brain injury with severe cognitive impairment, a third participant with moderate to severe cognitive impairment with a diagnosis of probable Alzheimer's type dementia, and a fourth and final participant with dementia probable Alzheimer's type, along with depression, and moderate to severe cognitive impairment.

Responses for choices varied to match the participants' capabilities: verbal (LeBlanc et al., 2006; Lucock et al., 2020; Mesman et al., 2011; Virués-Ortega et al., 2012), nodding, pointing (Mesman et al., 2011), eye contact, touching (LeBlanc et al., 2006; Mesman et al., 2011; Virués-Ortega et al., 2012), taking the item (Raetz et al., 2013), and pictoral, and textual (LeBlanc et al., 2006). The stimuli used were either leisure activities (LeBlanc et al., 2006;

Mesman et al., 2011; Raetz et al., 2013), or leisure activities and edibles together (Lucock et al., 2020; Virués-Ortega et al., 2012).

Lucock et al. (2020) replicated the study by Virués-Ortega et al. (2012). After completing preference assessments for leisure items and then edibles separately from each other, they conducted preference assessments for both leisure and edible items together. Both studies found a clear preference for leisure activities over edible items, and both studies found the highly preferred items identified during the preference assessment functioned as reinforcers. Raetz et al., (2013) found the MSWO preference assessment to be well suited for people with dementia as they are not as lengthy as PS preference assessments. They found the high preference choice to be consistent across repeated measures, and that they functioned as effective reinforcers.

LeBlanc et al. (2006) compared four modes of stimulus responses during PS assessments: tangible (touching), vocal (spoken), pictoral (pictures), and textual (written). Two participants responded most predictively compared with their later engagements with items during the vocal mode of testing. One other participant responded most predictively with the tangible mode, and the other participant responded most predictively with the pictoral mode but only slightly above the textual and tangible modes. Mesman et al. (2011) compared the results of PS assessments of five people diagnosed with dementia with those of their caregivers and family members. Spearman's rank-order correlations between the participants' preference assessment results and the ratings of the caregivers ranged from -*.27* to *.52* and were not statistically significant. They also found no significant correlations between the participants' and three of the family members (ranged from -*.02* to *.04*), however, there was one significant negative correlation (-*.71*) and one significant positive correlation (*.61*). There were also no statistically significant correlations between the caregivers' and family members' ratings (ranged from -*.12* to *.56*).

Teaching by Modeling

Since video making became available for the general public to use in the form of filming for home videos in the early 1920s, people have been filming each other and themselves (Nicholson, 2003). Through the years non-professional filming became easier with safer, smaller, and more portable cameras, clearer resolution, and less expensive methods of filming. Furthermore, video has become an efficient way of training individuals by means of using playback to illustrate to a person what they look, and sound like to others, known as video self-modeling, while they are applying for a job, dancing, acting, participating in sports, or any other way in which one would like to improve. A person can view themselves and make changes that are deemed necessary for positive development. Albert Bandura (1997) stated that people, especially children, react well to correction when seeing themselves on video, or someone they identify as like them (physical characteristics, age, group affiliation, ethnicity, etc.).

Video modeling carries this concept a step further by providing teaching opportunities by using others as models. People are hired or supplied (as students) to correctly act out a situation. People watching the video are trained by means of observational learning as first described by Albert Bandura (1977) as part of his work on his social theory of psychology. The video can be repeated as many times is as necessary and can be referred to when desired for further clarification. It also will be a consummate example as only an ideal version would be videoed for modeling purposes so no errors would be portrayed as may happen in repetitious in vivo (live) presentations. To this end, when several people are being taught using the same video modeling there will be consistency between sessions as well as between participants.

In 1997 the James Stanfield Company, Inc. defined video modeling as a “Series of video tapes that teach appropriate social behaviour to special education students, by use of professional

actors and actresses who demonstrate appropriate behaviour (wrong way/right way)" (Trademarkia, n.d.). Although video modeling had been used for several decades in behavioural study and research, the James Stanfield Company, Inc., received an American trademark registration for Video Modeling in the category of *Computer & Software Products & Electrical & Scientific Products* using the above definition (Trademarkia, n.d.). Video modeling has been used in many areas including: competitive dancers (e.g., Giambrone & Miltenberger, 2020), increasing sales in retail (Loughrey et al., 2013), math skills (Kellems et al., 2016), laundry skills (Horn et al., 2008), gymnastic techniques (Boyer et al., 2009), handwriting (Candler et al., 2014), safe nursing practices (Nielsen et al., 2009), first aid skills (Ergenekon, 2012), and computer training for older adults (Czaja et al., 2009; Struve & Wandke, 2009), just to name a few. In fact, over two hundred uses of video modeling have appeared in print (Bandura, 1977; Hitchcock et al., 2003).

Moreover, video modeling has been used as a successful tool in teaching people diagnosed with intellectual disabilities and ASD (Ayres & Langone, 2005; Bellini & Akullian, 2007; Bellini et al., 2019; Charlop-Christy & Daneshvar, 2004; Charlop-Christy et al., 2000; McCoy & Hermansen, 2007; Mechling et al., 2014; Rayner et al., 2009; Reagon et al., 2006). The repetition of video-based intervention can be used to strengthen the positive behaviour being sought in the forms of socially acceptable behaviour, communication, functional skills, and in reducing problem behaviours (Rayner et al, 2009). As previously suggested, using self-like models, as in children viewing other children their age modeling on a video, aids in promoting attention to the video, interest in viewing how the model is doing what they are doing, and a desire to copy those behaviours; all traits necessary for viewing a video but sporadic in persons diagnosed with ASD. While these visages in a presentation are beneficial, using the video

modelling showing positive outcomes increases the amount the person with ASD is able to focus (Bandura, 1997; Bellini & Akullian, 2007).

Whether it is identified as video modeling, video prompting, video self-modeling, computer-based video instruction or video priming (Bellini & Akullian, 2007), they are all ways of teaching a person directly from the use of a video. There has been little research on using video modelling to teach a person to present a demonstration, or to administer a test.

Previous Research on Teaching Stimulus Preference Assessment Protocol with Manuals

Considerable research has also been done on teaching stimulus preference assessments using manuals. For example, in 2015 Miljkovic et al. trained 6 undergraduate university students to administer MSWO using a video showing modeling without procedural instructions during baseline (no instructions other than to conduct a preference assessment as best they could). After being shown the video there was great improvement from baseline although none of the participants achieved the 85% predetermined mastery. Participants were then given the manual by Ramon and Yu (2010) to read, after which all achieved mastery.

Arnal (2013) taught 6 university students and 4 staff members to administer MSWO through an online medium. The web-based Computer-Aided Personalized System of Instruction is primarily used to teach university classes online by breaking down units into smaller sections wherein each section must meet mastery criterion before moving on to the next section. For this study Computer-Aided Personalized System of Instruction was modified to include videos without instruction. Before baseline, participants received written directives for MSWO assessment in the form of excerpts from relevant peer reviewed published journal articles correctly outlining assessments' methods. The MSWO manual by Ramon and Yu (2010) was used as the format for the Computer-Aided Personalized System of Instruction intervention.

None of the participants achieved the 85% predetermined mastery after baseline. Although each participant showed great improvement over baseline post Computer-Aided Personalized System of Instruction , only 7 out of the 10 participants achieved mastery. Only 6 participants retained generalization at follow-up one week later.

Boris (2016) used a combination of a self-instructional manual and video modeling to teach 12 participant staff members to conduct the Assessment of Basic Learning Abilities-Revised ,DeWiele, Martin, Martin, Yu, & Thompson, 2014. The Assessment of Basic Learning Abilities-Revised is used to assess the learning abilities and levels of individuals with intellectual disabilities and who have been diagnosed with ASD. During baseline a brief script outlining the steps and their logistics were given to the participants. During treatment, the staff participants received the self-instructional manual which made references and guidance to the video modeling. None of the participants achieved mastery criterion (at least 80% on each level with an overall mean of at least 90% accuracy) on baseline. Of the 8 out of 12 participants who achieved mastery after the self-instructional manual and video modeling arrangement, 7 went on to a generalization session 2 months later where they again met the mastery criterion.

Ramon et al. (2015) evaluated a self-instructional training manual (Ramon & Yu, 2010) for conducting MSWO stimulus assessment with eighteen undergraduate students. All participants received a baseline without procedural instruction. In a cross over design, half of the participants initially were taught the MSWO assessment using the self-instructional manual and those in this first group who did not achieve mastery of 85% went on to receive training using a procedural description of MSWO. The other half of the participants began with the procedural description of MSWO, and those in this second group who did not achieve mastery of 85% went on to receive training using the self-instructional manual. After studying the manual only, 4 of

the 9 participants in the first group met the mastery criterion, although after the crossover one additional participant achieved mastery. However, after studying the procedural description only, none of the 9 participants achieved mastery, but 7 participants reached mastery after the crossover. There were 6 participants who did not attain mastery after receiving both training methods, although they met all met criterion after observing a live demonstration. Of the 16 participants who went on to a generalization session, 14 reached mastery after 5-64 days.

Similarly, Chand (2016) evaluated a self-instructional training manual (Chand & Yu, 2010) for conducting PS stimulus assessment with 4 university students and 6 staff members. As with the Ramon et al. (2015) study, Chand's baseline was without procedural instruction, and she used the same crossover design substituting the PS self-instruction manual for the MSWO self-instruction manual, and the PS procedural description for the MSWO procedural description. None of the 10 participants achieved the 80% mastery in baseline. All 5 participants who were initially taught the PS self-instruction manual achieved mastery, thus they were not exposed to the procedural description. On the other hand, none of those 5 participants in the second group who studied the procedural description met mastery, although all of them met mastery criterion once they were taught the self-instruction manual in the cross over design. Of the 9 participants who went on to a generalization session conducted 7-28 days later, 7 reached mastery.

Previous Research on Teaching Stimulus Preference Assessment Protocol with Video

Lavie and Sturmey (2002) trained three staff members (assistant teacher to children diagnosed with autism spectrum disorders) to conduct paired-stimulus stimulus preference assessments in a classroom in a specialist school for children with autism. Training included two 40-minute sessions each consisting of verbal instructions, a review of a checklist of instructions, watching a video of the procedure, and practice with feedback. The staff members rewatched the

video and repeated their practice with feedback until they were able to complete the assessment with 85% accuracy for two consecutive sessions. All three participants were able to learn the assessment with mastery (98-100%) within about 40 hours of training. No follow-up was done.

In 2012, Bishop and Kenzer taught 11 behavioural therapists working with children diagnosed with ASD, to conduct shortened modified paired-stimulus preference assessments as a group in a clinic classroom. This brief paired-stimulus preference assessment consisted of presenting only two previously known favoured items taking less than 30 seconds to complete. A third item was then introduced with the previously chosen item and repeated until the same item is chosen twice in row. These shortened assessments allow the behavioural therapists to administer them at the beginning of each therapy session permitting current preferences to be considered without taking much time out of the session. Seven of the eleven therapists were able to obtain mastery of the assessment after having been taught in a group classroom session including a lecture, video, modeling, role-playing and feedback. The other four therapists in addition to the group classroom session, required in vivo feedback before mastery was obtained. Follow-up 4 weeks later found 8 of the behaviourists still functioned with accurate performance.

Weldy et al. (2014) taught 9 behavioural staff members who work with children diagnosed with autism to conduct 2 brief preference assessments: MSWO, and free-operant. Seven of the participants held master's degrees in related fields and all were familiar with behavioural programming. They used 30-minute videos which also contained instructions and modeling. Group one consisting of 4 participants were first presented with the brief MSWO and the free-operant second, while group two consisting of 5 participants were first presented with free-operant and then the brief MSWO. Each video lasted 30 minutes for a total of 60 minutes. All participants mastery criterion for free-operant after viewing the video one time. Seven

participants met the mastery criterion of the MSWO after one viewing, while two participants required a second viewing.

In Deliperi et al. (2015), three female university students who were also staff members of a private autism centre at a suburban university, were taught to conduct a paired-stimulus preference assessment in a private autism center at a suburban university. Each participant completed a written pretest questionnaire followed by baseline where they were asked to conduct a paired-stimulus preference assessment live, with one of the experimenters. Then participants were shown an 18 min and 36 s video. The first half included voice-over, and the second half demonstrated a full assessment without voice-over. Participants watched the video until they were able to achieve 90% correct or better for two consecutive sessions, thus all participants met criterion. Two participants required three training sessions and the other required two. Follow-up was at 1 week, 2 weeks, and 2 months later when responding was at or above 90%. The dependent variable was the percentage of step opportunities implemented correctly by the staff trainees on the 18-step task analysis. They found that video modeling was effective with staff maintaining mastery criterion up to 2 months later.

Lipschultz et al. (2015) trained four staff members working with children diagnosed with autism spectrum disorders and other intellectual disabilities to conduct three separate assessments: single stimulus, paired-stimulus, and multiple-stimulus without replacement in a room in a university-based autism centre. For each teaching session participants watched the video with embedded verbal instruction displaying each of the three assessments and then had them administer each of the assessments. In subsequent sessions, they continued watching the complete video until they mastered all the assessments although afterward, they were only required to rehearse the assessments not previously mastered. One participant needed to see the

video twice (total viewing time: 38 min, 56 s), a second participant three times (total viewing time: 58 min, 24 s), and two other participants each required six times, with one requiring feedback (total viewing time: 116 min, 48 s), before mastery was achieved on all three assessments. On follow-up done one week later, three of the participants found them performing at 90% accuracy.

Also in 2015, Rosales et al. taught three teachers who work with children diagnosed with autism to conduct three separate assessments: paired-stimulus, multiple-stimulus without replacement, and free-operant preference assessments in an isolated classroom of a therapeutic centre for children with autism. The three teachers held bachelor's degrees with one also holding a master's degree all in unrelated areas. There were three separate videos with embedded instruction, one for each assessment. The paired-stimulus was 6.5 minutes long; the multiple-stimulus without replacement was 9.0 minutes long; and the free-operant preference assessment was 9.43 minutes long. After reviewing written instructions and watching the video for a particular assessment immediate practice of that assessment took place. A maximum of two sessions a day were done with only one participant requiring to watch the video more than once for two sessions. One teacher met mastery after two sessions, another after six sessions, and the third after four sessions plus feedback to correct a recurring issue. After training for the PS assessment responses were 87% correct, 93% for MSWO, and 99% for free-operant. Follow-up was done for two of the teachers at two weeks and at a month showing their performance to be accurate.

Delli Bovi et al. (2017) trained two public school staff to conduct a multiple stimulus without replacement preference assessment using video modeling with voice-over instruction. The training took place in a conference room in a suburban public school. During baseline the

participants were given written instructions, the necessary tools and asked to conduct a preference assessment. For treatment, a video (17 min, 13 s) showing how to conduct an MSWO assessment was shown. The first third of the video featured a third-person point-of-view voice-over for a description of each step, the second section displayed an entire multiple stimulus without replacement preference assessment without voice-over, and the final third of the video showed a first-person point-of-view with voice-over. Post-test was immediately following and continued until 90 % correct responding occurred for two consecutive sessions. The dependent variable was the percentage of opportunities implemented correctly by the staff trainees a 13-step task analysis. Follow-up sessions were conducted at 4 and 8 weeks and demonstrated mastery.

Higgins et al. (2017) taught 3 direct staff workers to implement MSWO to children diagnosed with autism. One of the workers had completed a bachelor's degree in psychology while the other two workers' bachelor's degrees also in psychology were in progress. Training was done in a conference room remotely through a broadband wireless connection, while the experimenter was in a private office on a different floor of the same building. Training consisted of written instructions for 15 minutes, followed by a 24-minute multimedia presentation (Microsoft PowerPoint™, textual display with narrative, and a video model), and participants were permitted to view at will for up to 45 minutes. This was followed by feedback from a baseline session and ended with a roleplaying session complete with feedback. Mastery was achieved when 11 out of 13 components were accomplished with a 90% accuracy. Total training times for the three participants were 4.1, 4.3 and 6 hours. Follow-up was done using two sessions: at one month for two of the participants (100% for both sessions; and 100% and 77.8%) and at two months for the other participant (100% for both sessions).

Nottingham et al. (2017), worked with three participants with varying degrees of experience providing care to people with ASD. All sessions took place in a room at a university-based center for individuals with ASD using a between-subjects comparison in a combination of treatments. After being given a case file to study, participants were to initially determine the correct stimulus assessment to use: free-operant, single stimulus, paired-stimulus, or multiple stimulus without replacement. Then the participant was given written instructions to set up the chosen stimulus assessment. In baseline, the participants were then to attempt the stimulus assessment. The 19 min 28 s video utilized both voice-over and on-screen open captions. Participants watched the video once but were permitted to take notes and to keep them to use. During the assessment which followed the treatment video, the experimenter scored each step as correct, incorrect, or not applicable. After each session, the experimenter provided in vivo performance feedback as was needed. The assessments were evaluated on a 12-point task list for the single stimulus, a 15-point task list for the paired-stimulus, or a 12-point task list for the multiple stimulus without replacement. The three participants respectively required six, four, and seven sessions to achieve the mastery criterion of 100% across two consecutive sessions. Follow-up for mastery criterion was 100% correctly completed steps for one session.

Hansard and Kazemi (2018), taught four female undergraduate students with no previous experience in stimulus assessment, to conduct a paired-stimulus assessment. All sessions took place in two rooms with a one-way observation window in a nonconcurrent multiple-baseline design across participants. In baseline, participants were given 30 min to read a written summary of the PS procedure then asked to conduct a paired-stimulus assessment. The training video incorporated vocal and written instruction, models, text prompts, and voiceover instruction. Participants had 30 minutes to review the video (although they could only view it once) and

could also use the baseline written instructions throughout the entire experiment. The grading was based on five target responses: stimulus presentation, stimulus placement, post selection response, response blocking, and trial termination. Participants achieved mastery criterion when they achieved 90% across two consecutive sessions.

Lim and Hu in 2020 taught a multiple stimulus without replacement assessment (as well as token economy, and error correction procedures) to three participants using video with voice-over. All participants had special education teaching licenses and were working with children with ASD. For one participant, the sessions took place at a summer camp for children with ASD at a university, and the other two participants had their sessions at a school for students with special needs. Baseline included verbal instruction of the stimulus assessment, including a question-and-answer section, a demonstration section, and a PowerPoint presentation. Then participants were asked to conduct the assessment with a child with ASD. In treatment, a 5 min and 16 s video with voice-over was shown wherein participants could pause, replay, and take notes. Intervention ended when a participant reached mastery criterion of 100% accuracy across two consecutive sessions on an 8-point task analysis. Participant 1 attained mastery criterion after two sessions, Participant 2 after four sessions, and Participant 3 after five sessions. The primary dependent variable was participants' implementation of the MSWO procedures. Follow-up sessions were conducted 2, 4, and 6 weeks after treatment and ranged from 71–100%.

In 2020, Marano et al. taught three college graduates with no prior experience to conduct a paired-stimulus assessment using a combination of techniques including video using voice-over in a multiple baseline across participants’ design. Sessions took place in a room at a university. Baseline included the participant having 10-minutes to review written protocol before attempting a PS. The treatment began with a tutorial module of playing and replaying

videos and advancing slides. The second module was two to four 30 s video clips showing each step of the assessment presented either correctly or incorrectly. Participants could only advance when they accurately identified the clip as either a correct or incorrect depiction. In the third module, participants again watched video clips, different from those in the second module, still identifying the clips as either correct or incorrect but without feedback as in the second module. All three participants responded to 100% of all questions correctly in the test module on their first attempts. All three participants retained 100% during their PS assessment with a consumer, and later at follow-up.

Statement of the Problem

Video modeling plus instructions have been shown to be effective in a number of studies. For example, Weldy et al. (2014) embedded instructions in the videos using voice-over, Rosales et al. (2015) displayed the instructions using on-screen text, and Lipschultz et al. (2015) used both. However, the relative effectiveness of these two formats of embedded instructions is unknown. With a multi-component intervention, it is important to understand the effects of individual components so that the intervention can be streamlined (Ward-Horner & Sturmey, 2010). Therefore, the purpose of this research was to compare the effectiveness of voice-over instructions versus on-screen instructions in the context of video modeling to teach participants to conduct two stimulus preference assessment procedures.

There are two methods commonly recommended for use in a component analysis in single subject experimental design: dropout and add-in analysis (Bell et al., 2013; Howard et al., 2020; Ward-Horner & Sturmey, 2010). The dropout approach requires the treatment package to be presented intact and is then repeated with one of the components removed from the treatment. The add-in analysis, as the nomenclature implies, requires the treatment package to be presented

intact and is then repeated with a additional component. In this way, the experimenter can see the contribution of the components being studied. These methods are appropriate for behaviours that are reversible (e.g., behaviour increases or decreases as a result of an intervention), but could not be used in a skill acquisition context. Therefore, the components in the present study were examined using a between-subject design where two groups of participants received the same intervention except for the types of embedded instructions.

In addition, video modeling studies using voice-over or on-screen instructions have mostly been conducted with staff with varied experience working with individuals with intellectual disabilities and/or behavioural staff with some behaviour analytic training, even though they have not been trained in conducting stimulus preference assessments. This may have resulted in elevated baselines, or contributed to generalization, as in being able to transfer their learned knowledge in teaching children to the process of stimulus assessments with children. This study evaluated the intervention with individuals with no or minimal training in intellectual disabilities or in behaviour analytic techniques in general. Only undergraduate university students who have not worked in the field of intellectual disabilities and who have not received any formal behavioural training were recruited to participate in this study. Undergraduate university students are prime candidates often hired as tutors/workers to support people with intellectual disabilities and it would be important to examine the effectiveness of the teaching procedures with this group.

Lastly, follow-up and social validity assessments were not conducted consistently in previous video modeling studies. Social validity is an important measure of treatment acceptability by consumers (Miltenberger et al., 1989). This study incorporated both assessments. Specifically, this research compared video modeling with voice-over (VMVO)

instructions and video modeling with on-screen (VMOS) instructions in a between-subjects design to teach PS and MSWO assessments to participants. In addition, the effectiveness of each intervention was evaluated in a within-subject, concurrent multiple-probe design across participants. This study received ethical approval from the University of Manitoba Research Ethics Board 1, and consent was obtained from all participants before the study commenced.

Method

Summary of Modifications Due to the COVID-19 Pandemic

My original research plan included face-to-face interactions taking place on campus in a research room. Participants would have been required to carry out a preference assessment with a "client" roleplayed by a research confederate, during baseline and post-training (after watching the video). Follow-up was to have been done with persons diagnosed with intellectual disabilities recruited from a local community agency. In March of 2020, the University of Manitoba suspended all in-person research due to COVID-19. Recruitment of persons with intellectual disabilities was also not possible as community agencies limited unnecessary in-person contacts. Not knowing when in-person research could resume at that time, two major less-than-ideal procedural modifications were made to the research plan. First, I had to eliminate all in-person contacts by conducting sessions by ZOOM™. Second, the primary dependent measure was changed: participants were asked to detect procedural errors from watching prepared videos instead of applying the procedure themselves. These changes are explained in more detail in the relevant sections below.

Modifications for the Experimenter

As a researcher with disabilities of my own, there were accommodations made. As I have mobility issues and use a wheelchair in public, when filming the training videos I had to be

sure I could be steady enough when the scene called for me to rise and welcome the client. I also have dyscalculia (a genetic learning disorder pertaining to the inability to understand number based information), which required me to spend many months working with my advisor to understand and generate the statistics necessary. I also have dyslexia (another genetic learning disorder, but is a reading disability. Both dyscalculia and dyslexia are often paired with Attention Deficit Hyperactivity Disorder with which I have also been diagnosed), which means I read slower and it is often required for me to reread printed materials several times to understand. The time required for this was incorporated into my research and my studies in general. Although the COVID-19 pandemic did cause many issues during this research, it also simplified other areas. I was not required to be at the university and could do my interviews and participant research from the comfort of my home. This removed travelling during the snowy winter months which in a wheelchair is a major challenge. I also experience aphasia anomia (issues with word retrieval and memory) due to several strokes. When necessitated, I explain this to people I am working with so they may consequently understand my situation.

Participants and Setting

Sixteen undergraduate students at the University of Manitoba, over the age of 18 were recruited. Their information is presented in Table 1. They all had little or no experience with people diagnosed with autism or intellectual disabilities and had received no training in behavioural techniques in general and particularly in stimulus preference assessments. Each participant was screened using the Demographics Questionnaire (Appendix A) and the Disability Experience Questionnaire (Appendix B). Although participants would have been excluded (a) if they had worked with a person with intellectual disabilities or autism spectrum disorders for more than 6 months, or (b) if they had received formal training (e.g., attended a workshop) in

conducting stimulus preference assessment, none met the exclusion criteria. Several psychology professors/instructors were contacted and asked if they would feel comfortable in presenting my recruitment poster to their classes. All of those professors/instructors approached were so inclined, thus interested students contacted me through email. Participants received an honorarium of $40 for participating. All sessions were conducted individually through on-line sessions on ZOOM™ due to the pandemic as noted earlier.

Materials

Participants were required to have access to a computer, and the internet with access to ZOOM™. They were emailed datasheets to use during the following sessions: baseline, post-training, and follow-up. During these sessions, participants were asked to watch test videos (described below) and to record any perceived errors on the datasheets, by noting the type of errors (e.g. interruption), and describing the error (e.g. loud music playing). All ZOOM™ sessions were video recorded by the experimenter.

Design

I used a between-groups design with repeated measures and random group assignment to compare the interventions. I compared two different ways of using embedded video instructions, on-screen and voice-over, to teach participants to identify procedural errors while viewing a video of someone conducting a preference assessment. Every two participants were assigned to the OS or VO group in the order they contacted me. That is, the first two participants who contacted me were assigned to OS group and the next two participants to the VO group. The first four participants (two in each group) viewed the PS training video and the next four participants (two in each group) viewed the MSWO training video. This group assignment process was

repeated with the next eight participants. Participant performance was measured at baseline and again at post-training and a third time at a one-week follow-up.

Overview of the Experimental Phases

The study consisted of three phases. During the baseline phase, each participant was given instructions to watch a test video and identify all procedural errors that they could detect in the video. During the training phase, each participant was given instructions to watch a training video first before being tested similar to baseline. During the follow-up phase, the procedure was similar to those in the baseline phase. The procedures in each phase are described below.

Procedure

Baseline Session Procedure

Once the ZOOM™ session began, the experimenter gave the participant instructions verbatim as shown in Appendix C. Participants were asked to watch a test video (described below) and to record any errors they observed in the video on the datasheet that they have received. The participant then emailed the completed form back to the experimenter. Half of the participants completed one baseline session, and the other half completed two baseline sessions on two different days and with two different videos such that an equal number of participants in each group had the same number of baseline sessions

Test Videos. As noted earlier, the original dependent variable, which would have required the participant to apply the preference assessment procedures, was replaced with requiring the participants to identify errors while watching videos of someone conducting the assessment. In order to measure this variable, eight test videos were made and used during baseline, post-training, and follow-up phases. These videos included four 5-min clips showing a preference assessment using the PS procedure and four 5-min clips using the MSWO assessment.

Procedural errors were deliberately programmed into these videos, ranging from 8 to 12 per video. The numbers and types of errors across videos were comparable, but the specifics varied, and they occurred at different times in the videos (e.g., a disturbance in one video may have been a phone ringing, and in another video may have been loud music playing).

A participant's response was scored as either (a) a "hit" (correct response) if an error in the video was correctly identified, (b) a "miss" if an error present in the video was not detected, or (c) a "false alarm" if an identified error was not an error in the stimulus assessment procedure. Percentage of hits was computed by dividing the number of hits by the number of errors in the video. Percentage of misses was computed by dividing the number of misses by the number of errors in the video. Therefore, percentages of hits and misses would sum to 100%. Percentage of false alarms was computed by dividing the number of false alarms by the sum of false alarms and hits (i.e., total errors reported by the participant). For example, if a video contained 12 errors and a participant correctly detected 8 errors (hits) and identified 3 errors that were not errors (false alarms), the percentage of hits would be 67% (8/12), percentage of misses would be 33% (4/12), and the percentage of false alarms would be 27% (3/(8+3)).

As mentioned earlier, identifying errors as a dependent measure was one of the modifications made due to the pandemic. This is not commonly used to measure learning although it has been the focus in some studies. For example, Cattaneo and Boldrini (2017) reported several experiments on error identification and procedural knowledge. In one experiment, they studied the effects of error identification on acquisition of procedural knowledge with a group of 38 students in a clothing design class in a vocational school. As part of the lesson, half of the students were shown a video on using a sewing machine with errors embedded. Students were asked to identify the errors, their causes, consequences, and their

solutions. The other half of the students (control) received the same lesson, but without the error embedded video component. In pre-post comparisons, students showed a statistically significant improvement independent of group assignment. Moreover, students in the experimental group performed significantly better than students in the control group at post-lesson. The results of this study suggested that error identification might be a useful training tool to promote knowledge acquisition. In another experiment reported by Cattaneo and Boldrini, error identification was used also as a dependent measure with 136 students in an Apprentice Office Clerk class. Half of the students, the experimental group, were shown two videos simulating an interview with customers with errors intentionally embedded. Students were asked to identify the errors, their causes and what should have been done differently. The other half of the students, the control group, saw videos only showing the correct procedure and were asked to identify the correct behaviours and why they thought the behaviours were correct. During a written post-test, the experimental group performed significantly better than the control group in identifying both correct and incorrect behaviors. Cattaneo and Boldrini reported two additional studies with students studying butchering and with perioperative nursing students, respectively, with positive qualitative feedback from students on the use of error identification as a teaching component.

In another study, Beal et al. (1990) evaluated a self-questioning and correction strategy on children's ability to find and correct errors in stories shown on a computer and read aloud by the teacher. In the first experiment with 24 third grade students and 18 fifth and sixth grade students combined, students were randomly assigned to either the self-questioning group or the control group. Children in the experimental group were read the "problem" stories with errors deliberately embedded (missing sentence and contradiction) whereas children in the control

group were read “clear” stories with no errors. The self-questioning group was taught to ask questions (given typed cards with who, what, why, where, and when questions) about the story after hearing it. The teacher modeled the behaviours by identifying a problem with the story and helped the child if necessary to respond. Each child practiced the procedure with several stories. The experimenter recorded each child’s response when asked how to fix and improve the stories. The teacher modeled the correct response if necessary. The control group was not taught to ask questions but was asked to fix and improve the stories. Although the results showed that the experimental group identified more missing sentence and contradiction errors than the control group, children in self-questioning group had practice with the procedure. Beal et al. eliminated the practice component in Experiment 2. In the second experiment, 114 grade three students were randomly assigned to five conditions: (1) test-only controls; (2) no-questioning, clear stories; (3) no-questioning, problem stories; (4) self-questioning, clear stories; and (5) self-questioning, problem stories. Results showed that there was no significant difference between conditions and between story types, but there was a statistically significant difference between self-questioning and no-questioning groups on error identification and revision scores. That is, students who engaged in self-questioning identify more errors and made more correct revisions than students who did not engage in self-questioning.

Although the studies by Cattaneo and Boldrini (2017), and Beal et al. (1990) used error identification as a performance measure, they also used it as a teaching tool. It should be noted that identifying errors is not an ideal replacement for applying a procedure since it cannot provide information on whether a participant is able to carry out the procedure correctly.

Training Session Procedure

Video Modelling with Voice-over Instructions Training. Once the ZOOM™ session had launched, the experimenter provided the instructions in Appendix D to each participant. When the participant indicated that they were ready to begin, the experimenter started playing the training video with voice-over instructions (described below) for the PS or MSWO procedure depending on the participant's group assignment. The video was played in its entirety and the participants were able to watch the video, or parts of the video, as many times as they wanted.

Video Modelling with On-screen Instructions Training. The procedure for this condition was similar to the "voice-over" condition except that the training videos with on-screen instructions were shown. The videos with on-screen instructions were identical to those with voice-over instructions in all respects except that the instructions during the video were not spoken, but were presented as written text (open captions) at the bottom of the screen. Participants were given the same instructions at the beginning of the session (Appendix D) as in the video modeling with voice-over instruction condition.

Post-Training Procedure

Immediately after the participants had finished watching the training video, they were given instructions to watch the test video and to record any errors that they identified as in the baseline session.

Training Video

The training video showed the assessment with either oral instructions (voice-over) or written instructions (on-screen) for each step. Steps included how to greet the person being tested, how to set up the initial staging for the assessments, how to present items on each trial, how to provide consequences for different types of choice responses, and how to record and compute the results of the assessment. I prepared the videos using the program Camtasia

Studio™, Version 2019, by Tech Smith. In the videos, I acted as the behaviour therapist, and two of my sons, ironically both have been diagnosed with autism, alternated acting as the people with intellectual disabilities being assessed, and with being the camera operator.

Two training videos for the PS procedure were prepared and used. The two videos were identical except for how the instructions were presented: one video used on-screen (OS) instructions and the other voice-over (VO) instructions. Both videos were 15.2 minutes in length ensuring an equivalency between conditions. The contents of the instructions were identical in the two presentations. In addition, two training videos for the MSWO procedure were prepared and used. Again, the two videos were identical except for how the instructions were presented: one video used on-screen instructions and the other used voice-over instructions. Both videos were 16.5 minutes in length. During the intervention, each participant was asked to view one of the videos depending on their group assignment (On-screen-Paired Stimulus, Voice-over-Paired Stimulus, On-screen-Multiple-stimulus Without Replacement, On-screen-Multiple-stimulus Without Replacement,).

Follow-up Assessment Procedures

This transpired approximately one week following the training and the assessment procedure was the same as in baseline phase. One week for follow-up was selected to accommodate the undergraduate participant's schedules and to avoid conflicts such as exams, and Spring/Winter breaks, and it kept study duration within one semester for each participant.

Interobserver Agreement on Hits and False Alarms.

All sessions were recorded for interobserver agreement checks. All participants and all sessions (baseline, post-training, and follow-up) were assessed. Prior to conducting reliability checks, several observers were trained by the experimenter. The observer scored each

participant assessment independently using the same list of errors used by the experimenter. For hits, an agreement was defined as both the experimenter and observer recording the same programmed error being detected by the participant. If the experimenter and the observer's recordings differed, it was a disagreement. Percent agreement for each session was computed by dividing the number of agreements by the sum of agreements and disagreements and converting it to a percentage. Mean percentage agreement for hits across sessions and participants was 91.48% (range 55-100%).

For false alarms, if the experimenter and observer recorded the same error being reported by the participant or if both the experimenter and observer recorded that no false alarm was reported by the participant, it was scored as an agreement. If the experimenter and the observer's recordings differed, it was a disagreement. For example, if a participant reported three false alarms and both the experimenter and observer agreed on two and disagreed on one, that would be scored as two agreements and one disagreement. If a participant reported no false alarm and both the experimenter and observer also agreed that no false alarm was reported, that would be scored as one agreement. Percent agreement for each session was computed by diving the number of agreements by the sum of agreements and disagreements and converting it to a percentage. Mean percent agreement for false alarms across sessions and participants was 83.14% (range 0-100%).

Procedural Integrity

A trained observer also conducted procedural integrity checks by reviewing the recorded sessions, noting whether the experimenter had given the correct instructions at the beginning of each session (see Appendices C and D). All participants and all sessions were assessed and the results showed that the correct instructions were given for all but three sessions. During those

three sessions, the correct instructions were given before the recording started and therefore were not captured by the recording. Procedural integrity was considered 100% if the instructions were correctly given for each session, and 0% if they were not, or were not caught on video. Mean percentage across sessions and participants was 96%.

Social Validity Questionnaire

At the closing of the follow-up session, participants were asked to fill out a feedback questionnaire. Participants were asked to rate each of seven statements on a 5-point scale, where 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, and 5 = strongly agree. The seven statements can be seen in Table 2.

Statistical Procedures

To compare the instructional methods, I used a Repeated Measures of Analysis of Variance (RM ANOVA) and a Repeated Measures Analysis of Covariance (RM ANCOVA). The RM ANOVA evaluated the difference between OS and VO groups across three time points (baseline, post-training, and follow-up). The RM ANCOVA was used to control for baseline differences between groups.

In this study the between-subject factor was the OS vs VO instructions. The within-subject factor was performance (defined as the percentage of hits) at baseline, post-training, and follow-up for each participant. Baseline performance was the covariate in the ANCOVA. All statistical tests were performed using SPSS Version 28™ and statistical significance was evaluated with alpha set at *0.05*.

Both ANOVA and ANCOVA tests require similar assumptions about the data except that ANCOVA also requires homogeneity of regression. First, the dependent variable and covariate variable(s) should be measured on a continuous scale (Howell, 2010). In this study the dependent variable was the percentage of hits (errors correctly detected by the participant). A

second assumption is that the independent variable should consist of two or more categorical, independent groups (Howell, 2010). The independent variable of this study consisted of two instructional methods (OS and VO instructions embedded in the videos). Additionally, there should be an independence of observations between groups (Howell, 2010). In this study eight participants were in one group (VO) and another 8 were in a second group (OS) without overlap.

Another assumption is that there should be no significant outliers in the data. Box plots for each group (Howell, 2010) were examined (Figure 1). These plots illustrate a tube-like box where the centre represents the median (of the percentage of hits in this study) with the bottom of the box representing the lower percentile (25th percentile) and the top representing the upper percentile (75th percentile). On each end, the top and bottom, is a whisker (line) extending to show the percentiles between seventy-fifth to one hundred, and twenty-fifth to zero, respectively. To identify outliers the interquartile-range rule (IQR) is used; the difference between the upper and lower quartiles is calculated and multiplied by 1.5, and minor outliers being identified as dots outside of the whisker lines. The difference is again multiplied by 3 to identify extreme outliers, which are identified as asterisks outside of the whisker lines. In this study, in baseline, there were two minor outliers in the OS group (*10.00, 61.25*), but no extreme outliers, and there were no outliers in the VO group. In both post-training and follow-up, there were no outliers in both the OS and VO groups.

Another assumption is that residuals should be approximately normally distributed for each category of the independent variable. This was evaluated using the Kolmogorov-Smirnov test (Otsu & Taniguchi, 2020). The test results were not statistically significant for the OS group at baseline (D = .230, p = .200), post-training (D = .174, p = .200), and follow-up (D = .164, p = .200). Likewise, the test results were not statistically significant for the VO group at baseline (D

= .212, p = .200), post-training (D = .162, p = .200), and follow-up (D = .142, p = .200). Therefore, we could not reject the null hypothesis, meaning that the data did not deviate significantly from a normal distribution.

The next assumption is the need for homogeneity (equality) of variances (Howell, 2010; Huang et al., 2022). This was examined using the Levene's test for homogeneity of variances (Gastwirth et al., 2009). The Levene's test was not statistically significant for baseline (F = .048, p = .830), post-training (F = .094, p = .764), and follow-up (F = 2.87, p = .112), respectively. Therefore, we could not reject the null hypothesis that the variances between OS and VO groups were not significantly different.

Furthermore, the covariate should be linearly related to the dependent variable at each level of the independent variable and can be shown using scatterplots (Aron & Aron, 1994). As such, scatterplots of the covariate (baseline) and dependent variable (percentage of hits) for each instructional group (OS and VO) were visually examined and the relations appeared to be linear.

A final assumption is the need for the homogeneity of regression slopes, meaning there is no significant interaction between the covariate and independent variable (Frigon & Laurencelle, 1993). In this study there was no significant interaction between baseline performance (covariate) and instructional methods (independent variable) (F = .024, p = .879); therefore, we cannot reject the null hypothesis that there is no interaction between the independent variable and covariate. Thus, except for the minor outliers in the OS group, the data in the present study met the assumptions for ANOVA and ANCOVA.

Results

The mean percentage of hits for the OS and VO groups during baseline, post-training, and follow-up are shown in Figure 2. For the VO group, mean percentage of hits was 32.16 in

baseline, 45.61 in post-training, and 55.74 at follow-up. For the OS group, mean percentage of hits was 34.74 in baseline, 52.31 in post-training, and 49.58 at follow-up.

RM ANOVA showed that the difference between the on-screen and voice-over groups was not statistically significant ($F = 0.008$, $p = 0.931$). However, there was a significant difference in within-subjects effects across the three time points ($F = 8.054$, $p = 0.002$). Pairwise-comparisons between means at each time point showed that for the OS group, the mean difference between baseline and post-training was statistically significant ($p = 0.022$) and the mean difference between baseline and follow-up was also significant ($p = 0.005$). However, mean difference between post-training and follow-up was not significant ($p = 0.316$). In the VO group, the mean difference between baseline and post-training was significant ($p = 0.008$) but was not significant between baseline and follow-up ($p = 0.052$), and not significant between post-training and follow-up ($p = 0.768$). There was no significant interaction between instructional methods across repeated measures ($F = 0.711$, $p = 0.500$).

ANCOVA was used to determine if there was a significant difference between OS and VO groups, after controlling for baseline performance differences. The results showed that there was no significant difference between groups ($F = 0.141$, $p = 0.714$), and there was no significant difference between post-training and follow-up ($F = 1.509$, $p = 0.241$). Furthermore, the interaction effects were also not significant ($F = 0.708$, $p = 0.415$).

Errors Missed and False Alarms

The types and frequency of all misses made by participants are summarized in Table 3. During baseline, most of the errors in the video were undetected. This was not surprising since the participants had not yet received any training. At post-training, there was a reduction in misses in almost all errors. The most common error missed was not securing the examinee's

attention (missed a total of 18 times), not giving an instruction for the examinee to select an item and not recording an examinee's response (missed a total of 14 times each), not praising the examinee after a selection (missed a total of 12 times), and not presenting a choice to the examinee (missed a total of 11 times).

False alarms, behaviours as mistaken errors, in the OS group were highest in baseline (average across participants = 1), and it decreased in post-training (average across participants = 0.38) and at follow-up (average across participants = 0.25). In the VO group there was little change from baseline (average across participants = 1.25) to post-training (average across participants = 1.13) and at follow-up (average across participants = 1.38). Overall, there were slightly fewer false alarms in the OS group than in the VO group.

Social Validity Questionnaire Results

The findings of the questionnaire are summarized in Table 4. Fifteen of the 16 participants completed the questionnaire. Overall, the mean ratings were relatively high across all statements. All participants agreed that it was important for staff and care-givers to learn to conduct preference assessments. The training video was easy to follow and understand, and it provided all the necessary information to do the assessment. Participants felt that they had successfully learned to conduct the assessment, even though their perception was not corroborated by their ability to identify errors in the procedure.

Discussion

The purpose of this study was to compare two different ways to embed instructions, on-screen vs. voice-over, in videos to teach stimulus preference assessment to university students in a between-groups comparison. The results showed that the difference between the two groups was not statistically significant at post-training. That is, after viewing the training video,

participants in the voice-over and on-screen groups performed similarly. Furthermore, the difference between the two groups at post-training was not statistically significant when baseline performance was statistically controlled. However, there was a significant difference in within-subjects effects between baseline and post-training, and between baseline and follow-up, in both VO and OS groups. That is, participants in each group improved their performance significantly from baseline after watching the training video and performance at follow-up was maintained at levels similar to post-training.

Previous studies have shown that videos with voice-over instructions (e.g., Bishop & Kenzer, 2012; Deliperi et al., 2015; Delli Bovi et al., 2017; Hansard & Kazemi, 2018; Lavie & Sturmey, 2002; Lim & Hu, 2020; Lipschultz et al., 2015; Marano et al., 2020) and on-screen instructions (Rosales et al., 2015; Weldy et al., 2014) were effective in teaching stimulus preference assessment with percent accuracy ranging from 85% to 100% following training. Studies that have examined both on-screen and voice-over instructions concurrently have found them to be effective (e.g., Higgins et al., 2017, Nottingham et al., 2017). Several factors may have contributed to the differences between the results of the present study and previous research.

First, participants' experience may have contributed to the results. The majority of participants in previous research included staff who worked in the fields of teaching children, health care and/or with people diagnosed with autism or other cognitive disabilities. For example, Lim and Hu (2020) taught MSWO Assessments to participants who were working with children with ASD and Rosales et al. (2015) and Lipschultz et al. (2015) worked with teachers. However, two studies taught undergraduate students (Hansard & Kazemi, 2018) and college graduates (Marano et al., 2020) with no previous experience in the above fields.

Second, the present study differed from previous research in the setting in which the training was delivered. In all previous studies, training was held in-person in offices, classrooms, or laboratories. However, in-person delivery was not possible in this study due to the COVID-19 pandemic and the study was modified to deliver training remotely by video over the internet. This entailed the experimenter being in her home using her computer, while connecting with the participants who were in their own homes on their own devices. Participants accessed the videos using computers with different screen sizes, including smart phones. It was also not possible to control for interruptions or distractions in the participant's home (e.g., conversations nearby and roaming pets. An important advantage of training by video modeling is that it could be accessed by learners remotely and at their convenience. However, the robustness of the training procedure in different settings to overcome varying distractions require further study.

Third, a difference in the dependent measure may have also contributed to the results of the present study. In the present study, participants were asked to view videos of someone carrying out the procedure and to report errors that were made in the video. The primary dependent measure in previous research was the performance of the participants in carrying out the procedure following training. The two represent different repertoires as one entailed reproducing the behaviours while this study necessitated identifying incorrect behaviour, but participants were not required to carry out the procedure. Future research is needed to examine how these different repertoires are related and if they are transferable, for example, by adding a requirement for the participants to apply the procedures. This was not possible due to the suspension of in-person research because of COVID-19 restrictions.

Fourth, all previous studies have used a training criterion, a predetermined level required to perform the assessment at 80-100%, depending on the study. This was done by reviewing

training videos often with feedback, and frequently with another form of teaching until participants have achieved the criterion (e.g., Lipschultz et al., 2015; Weldy et al., 2014). In contrast, the present study did not use a training criterion, although participants were permitted to view the training video as many times as they would like, no participant viewed the entire video more than once, although a few did ask for a particular section to be replayed. With most participants in the present study, viewing the video only once produced only small to moderate gains in their performance in identifying errors.

In addition, to provide a comparison of embedded on-screen and voice-over instructions, the present study relied solely on the video as the training procedure. In previous studies involving video modeling training often included additional components such as: verbal instructions (Hansard & Kazemi, 2018; Lavie & Sturmey, 2002; Lim & Hu, 2020), a review of a checklist of instructions (Lavie & Sturmey, 2002), practice with feedback (Bishop & Kenzer , 2012; Lavie & Sturmey, 2002), in-vivo performance feedback (Nottingham et al., 2017), lectures, live role-playing and live modeling (Bishop & Kenzer, 2012; Hansard & Kazemi, 2018), written instructions or written protocol (Delli Bovi et al., 2017; Hansard & Kazemi, 2018; Higgins et al., 2017; Marano et al., 2020; Nottingham et al., 2017), Microsoft PowerPoint (Higgins et al., 2017; Lim & Hu, 2020), self notes (Lim & Hu, 2020; Nottingham et al., 2017), text prompts (Hansard & Kazemi, 2018), and video stills/slides (Hansard & Kazemi, 2018; Marano et al., 2020). In the present study, it appears that video modeling only, with either on-screen or voice-over instructions, was insufficient to yield a high level of accuracy in identifying errors. While adding other instructional components to the training procedure may improve performance accuracy, it is not clear whether the difference, or the lack of it, between the two types of instructions will remain the same.

The effects of the COVID-19 pandemic on the well-being and motivation of the participants during the study are unknown. Originally, all sessions were to be presented on campus in face-to-face sessions where participants would have been presented with tools to complete an actual assessment with a confederate and then with a person diagnosed with an intellectual disability. However, throughout the present study, the university campus was shutdown, all in-person research was suspended, and students had to complete all classes remotely. The impact of these changes and changes in students' everyday lives (work and isolation) should not be underestimated (Agnello & Giubellini, 2021; Schade, 2020) and may have a direct impact on the motivation of the participants in the present study.

The results during follow-up were mixed. For the OS group, five of the eight participants showed clear increases in percentage of hits at follow-up over those at post-training one showed little change, and two showed a decrease. For the VO group, one of the eight participants showed clear increase in percentage of hits at follow-up over those at post-video, three showed little change, and four showed a decrease. The improvements may be due to "wait time" (McOwen et al., 2008; Rowe, 1986; Straney & et al., 2018; Whiteoak, 2018) and "repeated practice" (Anderson, 1982; Brucks & Huang, 2020; Johnstone et al., 2002). Wait time may allow students to reflect on the information and to reconsider their responses (Whiteoak, 2018).

The results of the present study must also be considered in light of other limitations. The difference between OS and VO at post-video training was small and not statistically significant. Although this finding contributes to the limited literature showing that either method may be an acceptable teaching component, it must be interpreted cautiously as the sample size is relatively small for the between-group comparisons. Although it is possible that a larger sample may yield a statistically significant difference between groups, the small effect may not be clinically

important. Nevertheless, the findings of the present study seem to suggest that either OS or VO instruction could be used in training videos. This contributes to our understanding of the separate effects of the two components in the context of video modeling. This could represent a time saving for the future preparation of training videos.

In addition to the above, future research is needed to examine the effects of participant experience as a potential variable. It seems reasonable to expect individuals with a background working with vulnerable populations or with some training in behavioural principles to learn to perform preference assessments more easily than those with no experience and behavioural training. However, it is important to recognize that effective training procedures are needed for the latter group as was done in some previous research on teaching preference assessments (Deliperi et al., 2015; Hansard & Kazemi, 2018; Marano et al., 2020). Undergraduate university students are prime candidates who are recruited to work as tutors and support staff for individuals with intellectual and related disabilities and other impairments.

As a result of delivering the training remotely to participants in their homes due to the COVID-19 pandemic, future research is needed to examine setting characteristics more closely. The results of most research study are obtained under well-controlled settings (e.g., office, work/clinical setting, university laboratory). There is a need to evaluate the effectiveness of online training procedures (video modeling or remote learning) in less well-controlled settings such as home environments.

Based on the error analysis in the present study, there is clearly room for improving the videos in future research. Revising the videos to address the top errors noted such as emphasizing the instructions to the client and recording the client's responses should hopefully improve the effectiveness of the training videos. Furthermore, future research should ask

participants to also record the time each identified error occurred in the videos, for example, by displaying the time on the video or allowing the participant to pause the video.

Although the results of the present study showed a lack of difference between on-screen and voice-over instructions, it would be premature to conclude that the two types of instructions are equal given the limitations of the study and that participants did not make large performance gains with either procedure. Further studies in these comparisons will be required to see if there truly is no difference, and/or if variations in technique are required to improve efficacy.

References

Agnello, I., & Giubellini, C. (2021). Clinical experiences during the COVID-19 pandemic. *Journal of Analytical Psychology, 66*(3), 379–398. DOI: 10.1111/1468-5922.12688

Agnello, & Giubellini, C. (2021). Clinical experiences during the COVID-19 pandemic. *Journal of Analytical Psychology*, *66*(3), 379–398. https://doi.org/10.1111/1468-5922.12688

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, *89*(4), 369–406. https://doi.org/10.1037/0033-295X.89.4.369

Arnal, L. (2013). *Teaching individuals to conduct a preference assessment procedure using computer-aided personalized system of instruction*. (Doctoral dissertation). Retrieved from University of Manitoba, Department of Psychology, (Thesis A7427 Te)

Aron, A., & Aron, E. (1994). Statistics for psychology. *Prentice Hall.*

Autism Teaching Supports. (n.d.). *Preference Assessments/Reinforcer Inventories.* https://autismteachingsupports.weebly.com/preference-assessments.html

Ayres, K. M., & Langone, J. (2005, June). Intervention and instruction with video for students with autism: a review of the literature. *Education and Training in Developmental Disabilities: Division on Autism and Developmental Disabilities, 40*(2), 183-196. https://www.jstor.org/stable/23880090

Bandura, A. (1977). Social learning theory. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: W.H. Freeman.

Beal, C. R., Garrod, A. C., & Bonitatibus, G. J. (1990). Fostering children's revision skills through training in comprehension monitoring. *Journal of Educational Psychology, 82*(2), 275–280. https://doi-org.uml.idm.oclc.org/10.1037/0022-0663.82.2.275

Behavior Analyst Certification Board. (n.d.). *Credentials*. https://www.bacb.com/

Bell, E. C., Marcus, D. K., & Goodlad, J. K. (2013). Are the parts as good as the whole? A meta-analysis of component treatment studies. *Journal of Consulting and Clinical Psychology, 81*(4), 722-736. DOI: 10.1037/a0033004

Bellini, S., & Akullian, J. (2007). A meta-analysis of video modeling and video self-modeling interventions for children and adolescents with autism spectrum disorders. Exceptional *Children*, 73(3), 264.

Bellini, S., Akulian, J., & Hopf, A. (2019). Increasing social engagement in young children with autism spectrum disorders using video self-modeling. *School Psychology Review, 36(1),* 80-90. DOI: 10.1080/02796015.2007.12087953

Bishop, M.R., & Kenzer, A.L. (2012, January-March). Teaching behavioural therapists to conduct brief preference assessments during therapy sessions. *Research in Autism Spectrum Disorders, 6,* 450–457. doi: 10.1016/.rasd.2011.07.005

Blair, K.-S. C., Lee, I.-S., Cho, S.-J., & Dunlap, G. (2011). Positive behavior support through family–school collaboration for young children with autism. *Topics in Early Childhood Special Education*, *31*(1), 22–36. https://doi.org/10.1177/0271121410377510

Blair, K.-S. C., Umbreit, J., & Bos, C. S. (1999). Using functional assessment and children's preferences to improve the behavior of young children with behavioral disorders. *Behavioral Disorders*, *24*(2), 151–166. https://doi.org/10.1177/019874299902400202

Bloom, S. E., Lambert, J. M., Dayton, E., & Samaha, A. L. (2013). Teacher-conducted trial-based functional analyses as the basis for intervention. *Journal of Applied Behavior Analysis*, *46*(1), 208–218. https://doi.org/10.1002/jaba.21

Boris, A. (2016). *Training ABA service providers to conduct the Assessment of Basic Learning Abilities-Revised using a self-instructional manual and video modeling*. (Doctoral dissertation). Retrieved from University of Manitoba, Department of Psychology, (Thesis B64488)

Boyer, E., Miltenberger, R. G., Batsche, C., Fogel, V., & Leblanc, L. (2009). Video modeling by experts with video feedback to enhance gymnastics skills. *Journal of Applied Behaviour Analysis, 42*(4), 855-860. doi: 10.1901/jaba.2009.42-855

Brucks, M. S., Huang, S. (2020). Does practice make perfect? The contrasting effects of repeated practice on creativity. *Journal of the Association for Consumer Research, 5*(3), 291–301. https://doi.org/10.1086/709174

Candler, C., Mulder, A., & Nall, K. (2014, 03 April). Embedding video-based modeling handwriting instruction in a Montessori preschool phonics program. *Journal Of Occupational Therapy, Schools, & Early Intervention, 7*(2), 151-160. doi.org.uml.idm.oclc.org/10.1080/19411243.2014.930618

Carr, J. E., Nicolson, A. C., & Higbee, T. S. (2000). Evaluation of a brief multiple stimulus preference assessment in a naturalistic context. *Journal of Applied Behavior Analysis, 33*, 353-357. DOI: 10.1901/jaba.2000.33-353

Cattaneo, A. A. P., & Boldrini, E. (2017). Learning from errors in dual vocational education: Video-enhanced instructional strategies. *Journal of Workplace Learning. 29*(5), 353-369. DOI 10.1108/JWL-01-2017-0006

Chand, C. (2016). *Evaluation of a self-instructional manual for conducting paired-stimulus preference assessment with individuals with developmental disabilities*. (Doctoral dissertation). Retrieved from University of Manitoba, Department of Psychology, (Thesis 22112)

Chand, C., & Yu, C.T. (2010). How to conduct direct preference assessments for persons with developmental disabilities using a paired-stimulus procedure: A self-instructional manual. *University of Manitoba and St.Amant Research Centre.*

Charlop-Christy, M. H., & Daneshvar, S. (2004). Using video modeling to teach perspective taking to children with autism. *Journal of Positive Behaviour Interventions, 5*(1), 12-21. doi: 10.1177/10983007030050010101

Charlop-Christy, M. H., Le, L., & Freeman, K. A. (2000). A comparison of video modeling with in vivo modeling for teaching children with autism. *Journal of Autism and Developmental Disorders, 30*(6), 2000.

Cho, S.-J., & Blair, K.-S. C. (2017). Using a multicomponent function-based intervention to support students with attention deficit hyperactivity disorder. *The Journal of Special Education*, *50*(4), 227–238. https://doi.org/10.1177/0022466916655186

Clay, C.J., Andrew, L.S., Bloom, S.E., Bogoev, B.K., & Boyle, M.A. (2013). Assessing preference for social interactions. *Research in Developmental Disabilities: A Multidisciplinary Journal, 34*(1), p.362-371.

Czaja, S. J., Gregor, P., & Hanson, V. L. (2009). Introduction to the special issue on aging and information technology. *ACM Transactions on Accessible Computing (TACCESS), 2*(1), 1-4. doi.acm.org/10.1145/1525840.1525841

DeLeon, I.G., & Iwata, B.A. (1996). Evaluation of a multiple-stimulus presentation format for assessing reinforcer preferences. *Journal of Applied Behaviour Analysis,29*(4), 519-533.

Deliperi, P., Vladescu, J. C., Reeve, K. F., Reeve, S. A., & DeBar, R. M. (2015). Training staff to implement a paired-stimulus preference assessment using video modeling with

voiceover instruction. *Behavioral Interventions, 30*(4), 314-332. https://doi.org/10.1002/bin.1421

Delli Bovi, G.M., Vladescu, J.C., DeBar, R.M., Carroll, R.A., & Sarokoff, R.A. (2017). Using video modeling with voice-over instruction to train public school staff to implement a preference assessment. *Behaviour Analysis Practice, 10*, 72-76. doi 10.1007/s40617-016-0135-y

DeWiele, L., Martin, G. L., Martin, T. L., Yu, C. T., & Thomson, K. (2014). The Kerr-Meyerson Assessment of Basic Learning Abilities Revised: A self-instructional manual (2nd ed.). St.Amant Research Centre, Winnipeg, MB, Canada. Retrieved from https://stamant.ca/wp-content/uploads/2023/01/ABLA-R-self-instructional-manual-20140630-1.pdf

Ergenekon, Y. (2012, Autumn). Teaching basic first-aid skills against home accidents to children with autism through video modeling. *Educational Sciences: Theory & Practice, 12*(4), 2759-2766. EISSN: 2148-7561

Fava, L., & Strauss, K. (2010). Multi-sensory rooms: Comparing effects of the Snoezelan and the stimulus preference environment on the behaviour of adults with profound mental retardation. *Research in Developmental Disabilities: A Multidisciplinary Journal, 31*(1),160-171. DOI: 10.1016/j.ridd.2009.08.006

Fisher, W., Piazza, C.C., Bowman, L.G., Hagopian, L.P., Owens, J.C., Slevin, I. (1992). A comparison of two approaches for identifying reinforcers for persons with severe and profound disabilities. *Journal of Applied Behavior Analysis, 25*, 491–498. 10.1901/jaba.1992.25-491

Frigon, J., and Laurencelle, L. (1993). Analysis of Covariance: A proposed algorithm. *Educational and Psychological Measurement, 53*(1), 1-18.

Gastwirth, J. L., Gel, Y. R., Miao, W. (2009). The impact of Levene's test of equality of variances on statistical theory and practice. *Statistical Science, 24*(3), 343-360. DOI: 10.1214/09-STS301

Giambrone, & Miltenberger, R. G. (2020). Using video self-evaluation to enhance performance in competitive dancers. *Behavior Analysis in Practice, 13*(2), 445–453. https://doi.org/10.1007/s40617-019-00395-w

Goodley, D., & Roets, G. (2008). The (be)comings and goings of 'developmental disabilities': The cultural politics of 'impairment'. *Discourse: Studies in the Cultural Politics of Education, 29*(2), 239-255, doi: 10.1080/01596300801966971

Goodley, D., & Runswick-Cole, K. (2012). Reading Rosie: The postmodern disabled child, *British Psychological Society: Educational & Child Psychology, 29*(2), 53-66.

Hagopian, L.P., Long, E.S., & Rush K.S. (2004). Preference assessment procedures for individuals with developmental disabilities. *Behaviour Modification, 28*(5), 668-677. doi:10.1177/0145445503259836

Hansard, C., & Kazemi, E. (2018). Evaluation of video self-instruction for implementing paired stimulus paired-stimulus preference assessments. *Journal of Applied Behavior Analysis, 51*(3), 675-680. https://doi.org/10.1002/jaba.476

Hanley, G.P., Iwata, B.A., Lindberg, J.S., & Conners, J. (2003). Response-restriction analysis: I. assessment of activity preferences. *Journal of Applied Behaviour Analysis, 36(1)*, 47–58. doi:10.1901/jaba.2003.36-47

Higbee, T. S., Carr, J. E., & Harrison, C. D. (2000). Further evaluation of the multiple stimulus preference assessment. *Research in Developmental Disabilities, 21*, 61-73.

DOI: 10.1016/S0891-4222(99)00030-X

Higgins, W. J., Luczynski, K. C., Carroll, R.A., Fisher, W.W., & Mudford, O.C., (2017). Evaluation of a telehealth training package to remotely train staff to conduct a preference assessment. *Journal of Applied Behaviour Analysis, 50*(1), 1-14.

Hitchcock, C. H., Dowrick, P., & Prater, M. A., (2003). Video self-modeling in school based settings. *Remedial and Special Education, 56,* 36-45.

Horn, J. A., Miltenberger, R. G., Weil, T., Mowery, J., Conn, M., & Sams, L. (2008). Teaching laundry skills to individuals with developmental disabilities using video prompting. *International Journal of Behavioural Consultation and Therapy, 4*(3), 279-286.

Howard, A. J., Morrison, J. Q., & Collins, T. (2020). Evaluating self-management interventions: Analysis of Component Combinations. *School Psychology Review, 49*(2), p.130-143. DOI: 10.1080/2372966X.2020.1717367

Huang, F. L., Wiedermann, W., & Zhang, B. (2022). Accounting for heteroskedasticity resulting from between-group differences in multilevel models. Multivariate Behavioral Research, Vol.ahead-of-print (ahead-of-print), 1-21. https://doi-org.uml.idm.oclc.org/10.1080/00273171.2022.2077290

Johnstone, K. M., Ashbaugh, H., & Warfield, T. D. (2002). Effects of repeated practice and contextual-writing experiences on college students' writing skills. *Journal of Educational Psychology, 94*(2), 305–315. https://doi.org/10.1037//0022-0663.94.2.305

Kellems, R. O., Frandsen, K., Hansen, B., Gabrielsen, T., Clarke, B., Simons, K., & Clements, K. (2016). Teaching multi-step math skills to adults with disabilities via video prompting. *Research in Developmental Disabilities, 58,* 31–44. doi.org/10.1016/j.ridd.2016.08.013

Keyl-Austin, A. A., Samaha, A. L., Bloom, S. E., & Boyle, M. A. (2012). Effects of preference and reinforcer variation on within-session patterns of responding. *Journal of Applied*

Behavior Analysis, 45(3), 637–641. https://doi.org/10.1901/jaba.2012.45-637

Lane, J. D., Ledford, J. R., Shepley, C., Mataras, T. K., Ayres, K. M., & Davis, A. B. (2016). A brief coaching intervention for teaching naturalistic strategies to parents. *Journal of Early Intervention, 38*(3), 135–150. https://doi.org/10.1177/1053815116663178

Lavie, T., and Sturmey, P. (2002). Training staff to conduct a paired-stimulus preference assessment. *Journal of Applied Behaviour Analysis, 35*(2), 209–211. DOI: 10.1901/jaba.2002.35-209

LeBlanc, L. A., Cherup, S. M., Feliciano, L., & Sidener, T. M. (2006). Using choice-making opportunities to increase activity engagement in individuals with dementia. American *Journal of Alzheimer's Disease and Other Dementias, 21*(5), 318-325. DOI: 10.1177/1533317506292183

Lim, N., & Hu, X. (2020). The effects of video modeling on staff implementation of behavioral procedures in China. *Behavioral Interventions*, *35*(4), 704–716. https://doi.org/10.1002/bin.1735

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

Loughrey, T. O., Marshall, G. K., Bellizzi, A., & Wilder, D. A. (2013). The use of video modeling, prompting, and feedback to increase credit card promotion in a retail setting. *Journal of Organizational Behaviour Management, 33*(3), 200-208. DOI: 10.1080/01608061.2013.815097

Lucock, Z. R., Sharp, R. A., Jones, R. S. (2020). Preference for leisure items over edible items in individuals with dementia: A replication**.** *Journal of Applied Behavior Analysis, 53*(3),

1780-1788. DOI: 10.1002/jaba.679

Marano, K. E., Vladescu, J. C., Reeve, K. F., & Reed, F. D. (2020). Effect of conducting behavioral observations and ratings on staff implementation of a paired-stimulus preference assessment. *Journal of Applied Behavior Analysis, 53*(1), 296-304. https://doi.org/10.1002/jaba.584

McCoy, K., & Hermansen, E. (2007). Video modeling for individuals with autism: A review of model types and effects. *Education and Treatment of Children, 30*(4), 183-213. doi:10.1353/etc.2007.0029

McOwen, K. S., Kogan, J. R., & Shea, J. A. (2008). Elapsed time between teaching and evaluation: does it matter? *Academic Medicine*, *83*(10), S29–32. https://doi.org/10.1097/acm.0b013e318183e37c

Mechling, L. C., Ayres, K. M., Bryant, K. J., & Foster, A. L. (2014, December). Comparison of the effects of continuous video modeling, video prompting, and video modeling on task completion by young adults with moderate intellectual disability. *Education and Training in Autism and Developmental Disabilities: Division on Autism and Developmental Disabilities, 49*(4), 491-504. https://www.jstor.org/stable/24582346

Mesman, G. R., Buchanan, J. A., Husfeldt, J. D., Berg, T. M. (2011). Identifying preferences in persons with dementia: Systematic preference testing vs. caregiver and family member report. *Clinical Gerontologist, 34*(2), 154-159. DOI: 10.1080/07317115.2011.539516

Miljkovic, M., Kaminski, L., Yu, CT., & Wishnowski, L. (2015). Evaluation of video modelling and self-instructional manual to teach students to conduct a preference assessment. *Journal on Developmental Disabilities, 21*(1), 3-10. EISSN: 1917-7909

Miltenberger, R., Parrish, J. M., Rickert, V., & Kohr, M. (1989). Assessing treatment acceptability with consumers of outpatient child behavior management services. *Child & Family Behavior Therapy*, *11*(1), 35–44. https://doi.org/10.1300/J019v11n01_03

Narozanick, T., & Blair, K.-S. C. (2019). Evaluation of the class pass intervention: An application to improve classroom behavior in children with disabilities. *Journal of Positive Behavior Interventions*, *21*(3), 159–170. https://doi.org/10.1177/1098300718806650

Nicholson, H.N. (2003). British holiday films of the Mediterranean: At home and abroad with home movies: Small-gauge and amateur film, ca.1925-1936. *Film History, 15*(2), 152-165. DOI: 10.2979/FIL.2003.15.2.152

Nielsen, D., Sigurdsson, S. O., & Austin, J. (2009). Preventing back injuries in hospital settings: The effects of video modeling on safe patient lifting by nurses. *Journal of Applied Behaviour Analysis, 42*(3), 551-561.

Nottingham, C. L., Vladescu, J. C., Giannakakos, A. R., Schnell, L. K., & Lipschultz, J. L. (2017). Using video modeling with voiceover instruction plus feedback to train implementation of stimulus preference assessments. *Learning and Motivation, 58*, 37-47. https://doi.org/10.1016/j.lmot.2017.01.00

Otsu, T., & Taniguchi, G. (2020, October). Kolmogorov–Smirnov type test for generated variables. *Economics Letters, 195*. https://doi.org/10.1016/j.econlet.2020.109401

Pace, G. M., Ivancic, M. T., Edwards, G. L., Iwata, B. A., & Page, T. J. (1985). Assessment of stimulus preference and reinforcer value with profoundly retarded individuals. *Journal of Applied Behaviour Analysis, 18*, 249–255. doi:10.1901/jaba.1985.18-249

Paramore, N.W., & Higbee, T. S. (2005). An evaluation of a brief multiple-stimulus preference

assessment with adolescents with emotional-behavioral disorders in an educational setting. *Journal of Applied Behavior Analysis, 38(3),* 399-403.

Piazza, C.C., Fisher, W.W., Hagopian, L.P., Bowman, L.G., & Toole, L. (1996). Using a choice assessment to predict reinforcer effectiveness. *Journal of Applied Behaviour Analysis, 29*(1), 1–9. doi:10.1901/jaba.1996.29-1 https://doi-org.uml.idm.oclc.org/10.1177/109830070100300303

Ramon, D., & Yu, C. T. (2010). How to conduct direct preference assessments for persons with developmental disabilities using a multiple stimulus without replacement procedure: A self-instructional manual. *University of Manitoba and St.Amant Research Centre.*

Ramon, D., Yu, C.T., Martin, G.L., & Martin, T. (2015). Evaluation of a self-instructional manual to teach multiple-stimulus without replacement preference assessment procedure. *Journal of Behavioural Education*, *24*(3), 289-303. doi 10.1007/s10864-015-9222-3

Raetz, P. B., LeBlanc, L. A., Baker, J. C., Hilton, L. C. (2013). Utility of the multiple-stimulus without replacement procedure and stability of preferences of older adults with dementia. *Journal of Applied Behavior Analysis*, *46*(4), 765-780. DOI: 10.1002/jaba.88

Rayner, C., Denholm, C., & Sigafoos, J. (2009). Video-based intervention for individuals with autism: Key questions that remain unanswered. *Research in Autism Spectrum Disorders, 3*(2), 291–303. doi.org.uml.idm.oclc.org/10.1016/j.rasd.2008.09.001

Reagon, K.A., Higbee, T.S., & Endicott, K. (2006). Teaching pretend play skills to a student with autism using video modeling with a sibling as model and play partner. *Education and Treatment of Children*, 29(3), 517-528.

Roane, H. S., Vollmer, T. R., Ringdahl, J. E., & Marcus, B. A. (1998). Evaluation of a brief

stimulus preference assessment. *Journal of Applied Behaviour Analysis, 31*, 605–620. doi: 10.1901/jaba.1998.31-605

Romaniuk, C., & Miltenberger, R. (2001). The influence of preference and choice of activity on problem behavior. *Journal of Positive Behavioral Interventions, 3*(3), 152-159.

Rosales, R., Gongola, L., & Homlitas, C. (2015). An evaluation of video modeling with embedded instructions to teach implementation of stimulus preference assessments. *Journal of Applied Behaviour Analysis, 48*(1), 209–214. doi: 10.1002/jaba.174

Roscoe, E. M., Iwata, B. A., & Kahng, S. W. (1999, Winter). Relative versus absolute reinforcement effects: Implications for preference assessments. *Journal Of Applied Behavior Analysis, 32*(4), 479–493. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1284210/pdf/10641302.pdf

Rowe, M. B. (1986). Wait time: Slowing down may be a way of speeding up! *Journal of Teacher Education, 37*(1):43-50. doi:10.1177/002248718603700110

Schade, L. D. (2020, April 03). 6 tips for avoiding Zoom fatigue in the age of COVID-19. *EcoPreacher*. Retrieved February 017, 2022, from https://www.patheos.com/blogs/ecopreacher/2020/04/tips-avoiding-zoom-fatiguecovid-19/

Sellers, T. P., Bloom, S. E., Samaha, A. L., Dayton, E., Lambert, J. M., & Keyl-Austin, A. A. (2013). Evaluation of some components of choice making. *Journal of Applied Behavior Analysis, 46*(2), 455–464. https://doi.org/10.1002/jaba.46

Straney, G. R., Ranger, C., & Blosser, E. (2018). Time out for teaching: Implementing a post-procedure time out to improve surgical evaluations. *Obstetrics and Gynecology (New York. 1953), 132*(1), 58–58S. https://doi.org/10.1097/01.AOG.0000546677.68958.78

Struve, D., & Wandke, H. (2009). Video modeling for training older adults to use new technologies. *ACM Transactions on Accessible Computing (TACCESS), 2*(1), 1-24. doi:10.1145/1525840.1525844

Trademarkia. (n.d.) *Video Modeling Trademark Information*. Retrieved from http://www.trademarkia.com/video-modeling-75372737.html

Verriden, A. L., & Roscoe, E. M. (2016). A comparison of preference-assessment methods. *Journal of Applied Behaviour Analysis, 49*(2), 265–285. doi:10.1002/jaba.302

Virués-Ortega, J. V., Iwata, B. A., Nogales-González, C., Frades, B. (2012). Assessment of preference for edible and leisure items in individuals with dementia. *Journal of Applied Behavior Analysis, 45*(4), 839-844. DOI: 10.1901/jaba.2012.45-839

Ward-Horner, J., & Sturmey, P. (2010). Component analyses using single-subject experimental designs: A review. *Journal of Applied Behavior Analysis, 43*(4), 685–704. doi:10.1901/jaba.2010.43-685

Weldy, C.R., Rapp, J.T., & Capocasa, K. (2014). Training staff to implement brief stimulus preference assessments. *Journal of Applied Behaviour Analysis, 47*(1), 214–218.

Whiteoak, H. (*2018).* The importance of "*wait-time*" in the technology-based classroom. *Schoology*. http://www.schoolology.com.

Windsor, J., Piché, L. M., & Lock, P. A. (1994). Preference tests: A comparison of two presentation methods. *Research in Developmental Disabilities, 15*, 439–45. DOI: 10.1016/0891-4222(94)90028-0

Table 1

Participant Characteristics and Intervention Received

Participant	Gender	Age	Year in Program	Major	Intervention
P1	Female	19	Second	Science	OS-PS
P2	Female	22	Second	Psychology	OS-PS
P3	Female	21	Second	Psychology	OS-PS
P4	Female	21	Third	Native Studies	OS-PS
P5	Female	21	Third	Psychology	VO-PS
P6	Male	58	Second	Integrated Studies	VO-PS
P7	Female	23	Second	Psychology	VO-PS
P8	Female	19	Second	Psychology	VO-PS
P9	Female	21	Second	Psychology	OS-MSWO
P10	Female	35	Third	Psychology	OS-MSWO
P11	Male	29	Third	Environmental Design	OS-MSWO
P12	Female	21	Second	Psychology	OS-MSWO
P13	Male	23	Fourth	Economics	VO-MSWO
P14	Female	20	Third	Psychology	VO-MSWO
P15	Male	26	Third	Psychology	VO-MSWO
P16	Non-binary	22	Third	Psychology/Geography	VO-MSWO

Table 2

Participant Feedback Questionnaire (Social Validity Questionnaire)

Participant Code: _____
Date: _____

Please indicate how strongly you agree or disagree with each statement by circling the number after each statement.

1. It is important for staff working with individuals with developmental disabilities to learn to conduct preference assessments.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree/Disagree	Agree	Strongly Agree

2. It is important for people living with individuals with developmental disabilities to learn to conduct preference assessments.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree/Disagree	Agree	Strongly Agree

3. The video was easy to follow and understand.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree/Disagree	Agree	Strongly Agree

4. The video provided all the necessary information for me to do the assessment.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree/Disagree	Agree	Strongly Agree

5. I believe I have successfully learned how to conduct the preference assessment procedure shown in the video.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree/Disagree	Agree	Strongly Agree

6. I feel confident and ready to conduct the preference assessment with clients after watching the video.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree/Disagree	Agree	Strongly Agree

7. I would recommend the videos to others who wish to learn this procedure.

1	2	3	4	5
Strongly Disagree	Disagree	Neither Agree/Disagree	Agree	Strongly Agree

Other Comments:

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

Table 3

Types and Total Number of Errors Missed Across All Participants

Types of Errors Missed	Baseline	Post-Training	Follow-Up
Disruption or interruption	5	1	1
No choice presented for testee	23	11	14
Testee selected more than one item	19	5	4
Items were not rotated	18	8	6
Tester did not record	19	14	10
Items in incorrect position	11	6	5
Tester did not praise selection	26	12	16
Tester did not secure testee's attention	25	18	11
Tester did not instruct testee to select an item	22	14	11
Tester gave testee inappropriate attention for not selecting	5	5	3

Table 4

Social Validity Questionnaire's Mean Ratings Across Participants (n = 15)*

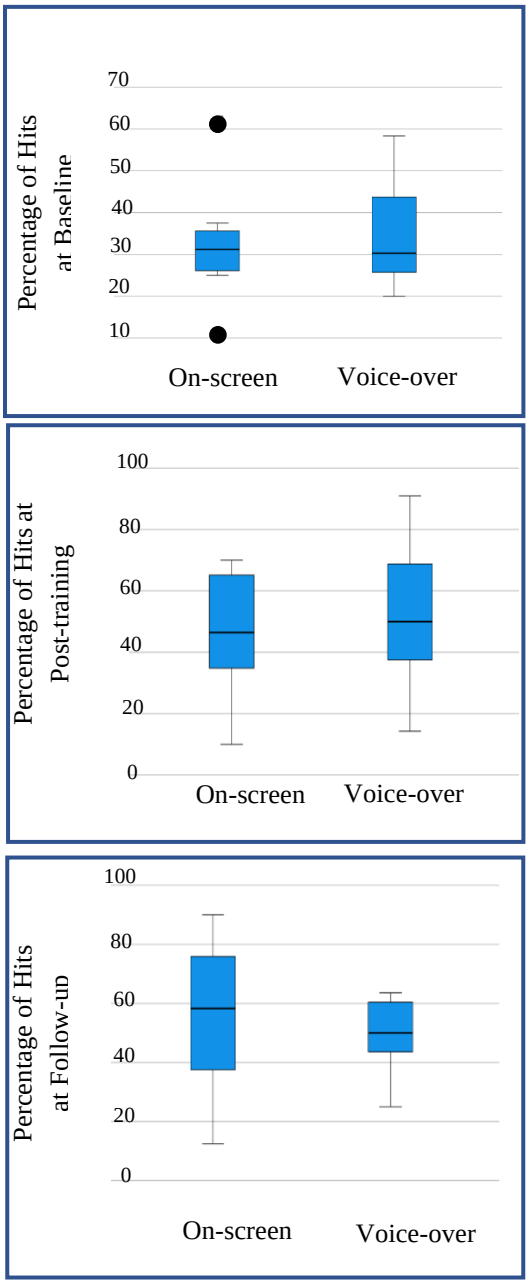
Statements	Mean Rating (Range)
It is important for staff working with individuals with developmental disabilities to learn to conduct preference assessments.	4.9 (4-5)
It is important for people living with individuals with developmental disabilities to learn to conduct preference assessments.	4.7 (4-5)
The video was easy to follow and understand.	4.7 (3-5)
The video provided all the necessary information for me to do the assessment.	4.3 (2-5)
I believe I have successfully learned how to conduct the preference assessment procedure shown in the video	4.1 (2-5)
I feel confident and ready to conduct the preference assessment with clients after watching the video.	3.9 (2-5)
I would recommend the videos to others who wish to learn this procedure.	4.5 (3-5)

* One participant did not complete the questionnaire.

Note. A 5-point Likert scale was used: 1 = strongly disagree, 2 = disagree, 3 = neither agree nor disagree, 4 = agree, and 5 = strongly agree.

Figure 1

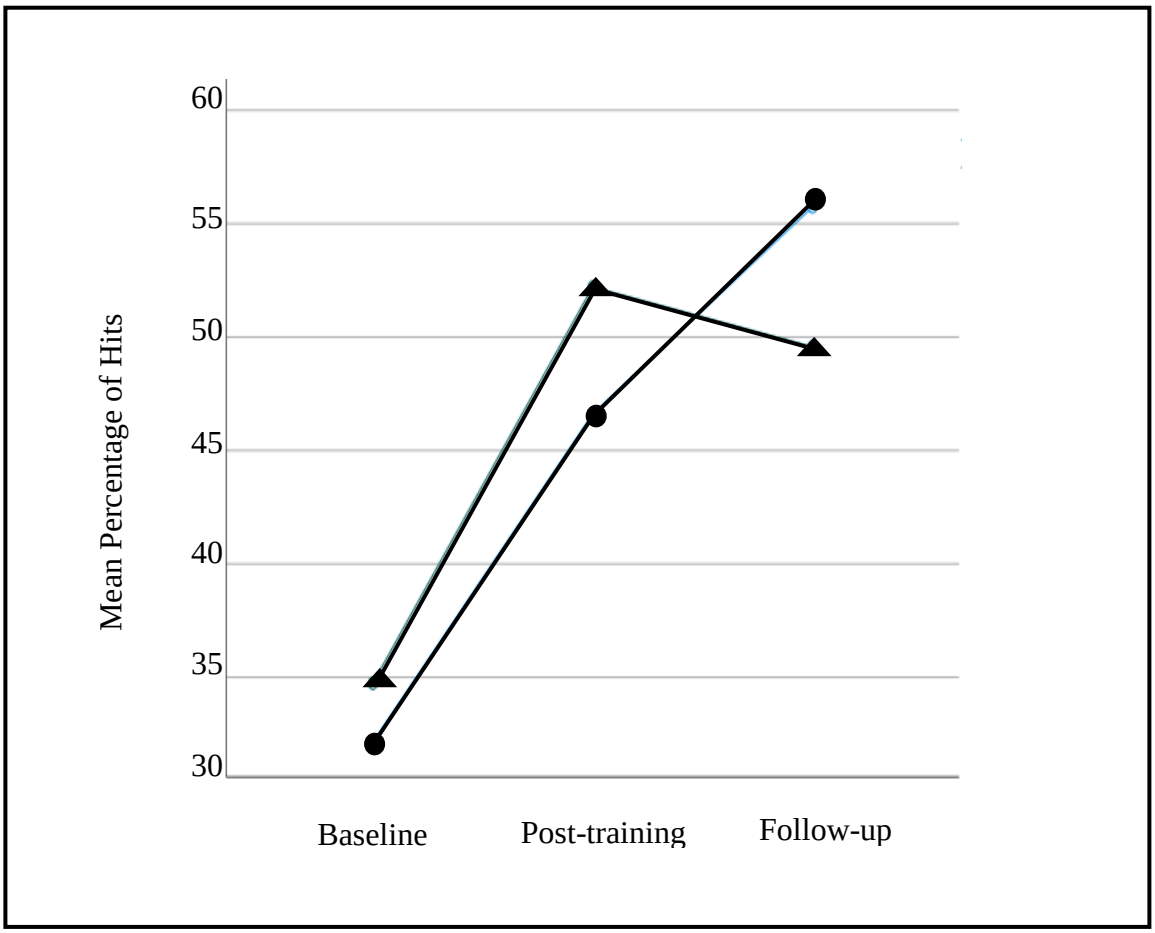
Box Plots for On-screen and Voice-over Groups at Baseline, Post-training, and Follow-up.



Note. The rectangle box's centre represents the median. The bottom of the box represents the lower percentile (25th percentile) and the top represents the upper percentile (75th percentile). On each end a whisker (line) extends to show the percentiles between seventy-fifth to one hundred, and twenty-fifth to zero, respectively, demonstrating the extreme points of the data within acceptable boundaries. Minor outliers are identified as dots outside of the whisker lines and extreme outliers are identified as asterisks. Graphs have different scales on the vertical axis.

Figure 2

Mean Percentage of Hits for the OS (triangles in graph) and VO (circles in graph) Groups.



Appendix A

Demographics Questionnaire

Gender: _____

Age: _____

Year at University: _____

Major (if known): _____

Minor (if known): _____

Have you worked, or lived, with anyone with a developmental disability or an autism spectrum disorder?

_____

If yes, for how long? _____

In what capacity? _____

Please Describe. _____

_____

_____

If yes, have you received any training to work in this area? _____

If yes, what type of training? _____

Please Describe. _____

_____

_____

Appendix B

Disability Experience Questionnaire

Please complete this questionnaire about yourself without including any identifying information about yourself or others. You may decline to answer any question if you wish.

1. Do you know people with physical disability(ies)?

YES ☐ NO ☐

2. Do you know people with developmental/emotional/psychiatric disability(ies)?

YES ☐ NO ☐

3. Tell me about your experiences, if applicable, with disability and disabled people…

_____

_____

_____

_____

_____

_____

_____

Appendix C

Oral Instructions to Participants to Watch Test Videos During Baseline, Post-Training, and Follow-Up Sessions

"Thank you for helping me with this study. Today you are going to watch a 5minute video of me, on the left, conducting a stimulus preference assessment with <name of confederate>, on the right, who is playing the role of a client, a person with no speech. I will be using a procedure called paired-stimulus using two items [or multiple-stimulus-without-replacement using five items]. You have received a scoring sheet for this procedure and your task is to score the number and types of mistakes I have made when carrying out the procedure. Record every mistake that you are able to identify in the scoring sheet.

I cannot answer any questions or provide any additional information about the preference assessment procedure at this time. I can start the video whenever you are ready. If the video freezes during play, let me know I will rewind it to a couple of seconds before the freeze. You will have up to 10 minutes to score the preference assessment using the scoring sheet I have sent you, or you can let me know when you would like to stop. Tell me if you need me to rewind, rewatch a certain part, or stop. Please let me know when you are ready to begin. Just do the best you can."

Appendix D

Oral Instructions for Participants Before Viewing the Training Videos with Voice-over or On-screen Instructions

"Thank you for helping me with this study. Today, you will watch a 15-minute video. The video will show you how to conduct a stimulus preference assessment using the paired-stimulus [or multiple-stimulus-without-replacement] procedure. You ***must*** initially watch the video in its entirety first. If the video freezes during play, let me know I will rewind it to a couple of seconds before the freeze. After you have watched the complete video, you may watch the entire video again as many times as you wish, or you may review specific sections of the video as many times as you would like.