The Effects of Speech Output Technology on Skill Acquisition in Children with Autism Spectrum Disorder: A Preliminary Investigation

A DISSERTATION

SUBMITTED TO THE FACULTY

OF

THE GRADUATE SCHOOL OF APPLIED AND PROFESSIONAL PSYCHOLOGY

OF

RUTGERS,

THE STATE UNIVERSITY OF NEW JERSEY

BY

MARIANA SIKIU TORRES-VISO

IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE

OF

DOCTOR OF PSYCHOLOGY

NEW BRUNSWICK, NEW JERSEY OCTOBER 2015

APPROVED: ______________________________

Kimberly N. Sloman, Ph.D.

____

Sandra L. Harris, Ph.D.

DEAN: ______________________________

Stanley B. Messer, Ph.D.

Abstract

Previous research on the use of voice output communication aids (VOCAs) has found a number of positive effects on the behavior of both the VOCA user and their communicative partners. Among these outcomes, studies have found that incorporating speech output into language learning tasks may result in faster and more efficient learning for adults with disabilities (e.g., Kohl & Schlosser, 2005; Schlosser et al, 1998). However, these studies have been conducted with adult participants, thus the effects of VOCA on the learning of graphic symbols in children are still unknown. Furthermore, the relationship between speech output and skill acquisition has not yet been evaluated for individuals with autism spectrum disorders (ASD). The present study aimed to assess whether previous findings on the effects of VOCA on learning could extend to school-aged children with ASD. This study employed a single-case, multielement design with multiple baseline probes to evaluate differences in teaching with speech output (SO condition) versus no speech output (NSO condition) across three sets of stimuli for Alan, a 12-year-old male with ASD. Results showed that higher rates of correct responding and lower rates of errors for targets were obtained in the SO condition. Furthermore, Alan met mastery criteria with SO targets in approximately half the number of sessions required for mastery of NSO targets across tiers, indicating higher efficiency in the SO sessions. Findings from this investigation thus provide strong preliminary evidence for the benefits of speech output in skill acquisition for children with ASD, both in terms of student accuracy and session efficiency.

Acknowledgements

The author would like to thank Dr. Sloman and Dr. Harris for their continued warmth, support and guidance in the development of this investigation, as well as throughout her graduate career. Furthermore, the author would like to thank the research and clinical staff at the Douglass Developmental Disability Center, especially Meredith Bamond and Katelyn Selver, for their invaluable contributions to this project. Finally, the author would like to give special thanks to Professor Emerita Edith Neimark for creating the Neimark endowment for graduate student research. Your support was instrumental to the success of this project!

Table of Contents

List of Tables

List of Figures

Appendices

Chapter I

Introduction

Language impairments are one of the central features of autism spectrum disorders (ASD). A number of studies have found that 25% to 61% of children with ASD remain essentially nonspeaking or with severe language impairments (Weitz, Dexter, & Moore, 1997), thus, a great deal of empirical and clinical attention has focused on identifying therapeutic interventions for this population. In an effort to address some of the negative effects associated with the language impairments inherent in ASD, augmentative and alternative communication (AAC) modalities were developed as means to facilitate communicative behaviors. A variety of AAC systems presently exist, with the primary goal of supplementing (i.e. augmenting) a learner's current language abilities, or acting as a primary (i.e., alternative) method of expressive communication (Mirenda, 2003). AAC modalities are typically categorized into unaided and aided categories, where unaided AAC does not require materials or visual prompts, and include modalities such as sign language, gesturing, and finger spelling. Aided AAC, on the other hand, make use of additional materials such as a pictographic symbols and electronic/non-electronic boards to facilitate communication. In recent years, researchers and clinicians have used computer technology advances to develop voice output communication aids (VOCAs) as an additional aided AAC modality.

Voice Output Communication Aids (VOCAs)

VOCAs, also known as speech-generating devices, are portable devices that emit speech output, or the vocal or auditory representation of words and images, by pressing graphic symbols on screens (Mirenda, 2003). These devices extend the capability of other forms of non-electronic aided AAC by providing speech output to both the user and the communicative partner (Drager, Reichle & Pinkoski, 2010). Learners are taught to navigate through categories within the device to locate visual symbols for items, activities, and individuals, as well as sentence components. Once an individual has pressed the desired icon, the device emits the vocal label for the symbol. With the development of voice output software for more common household electronic devices, such as the IPod® and IPad®, VOCAs are becoming increasingly common within the ASD population (Mirenda, 2003; van der Meer & Rispoli, 2010). While VOCAs may vary in physical or software features, the central characteristic of these devices is the speech output component, which is emitted in either digitized or synthesized speech. Synthetic speech has been defined as "any voice produced by a (rule-based) text-to-speech system that converts an input string of text characters to an output speech wave" (Allen, 1992). Synthesized speech often possesses a "robotic" quality, and is developed by the device itself. Digitized speech output, on the other hand, refers to the playback of a human voice that has been previously recorded and stored within the device (Schlosser & Blischak, 2001).

Parallel to the recent increase in the introduction of VOCAs in clinical settings there has also been a rapid increase in the number of studies evaluating the efficacy of VOCA, particularly in terms of its value as an AAC modality for learners with ASD (van der Meer & Rispoli, 2010). Research has found a number of potential benefits for VOCA, including positive behavioral changes for both the communicative partner (the "listener") and the user (the "speaker").

Effects of VOCA on Listener Behaviors

Various studies have described the benefits of using VOCA in terms of "listener" effects; that is, the effects of VOCA in the attitudes and behaviors of those who engage in communicative exchanges with the VOCA user. For example, Gorenflo and Gorenflo (1991) found that listener attitudes improved when an individual with disabilities utilized an AAC system with speech output rather than when he used an AAC system without speech output. Listeners reported stronger positive beliefs about the speaker, as well as a higher likelihood of interacting with the individual when he utilized a speech output device. These results are similar to those found by Lilienfield and Alant (2002), in which attitudes of communicative partners towards a 13-year old boy with cerebral palsy were found to be more favorable when the boy communicated with speech output than without it. The authors found that participants who observed a videotape of an individual communicating with a speech output device indicated in the Communication Aid/Device Attitudinal Questionnaire (CADAQ) that they perceived him as having higher communicative competence. Participants also stated that they were more likely to interact with him and think positively of his abilities.

The use of VOCA has also been linked to increases in the frequency and quality of speaker-listener interactions. Communicative partners may be more likely to understand and respond to VOCA outputs than to other aided AAC systems without speech output, such as non-electronic boards or graphic symbols. Schepis and Reid (1995) evaluated the effects of a treatment package that involved VOCA on the interactions of clinical staff and a young woman with multiple disabilities, and found significant increases in staff interactions with the learner across a number of settings, as compared to baseline and a control condition. Similarly, Schepis, Reid, Berhmann, and Sutton (1998) found that, after teaching four preschoolers with ASD to communicate with classroom staff and peers via VOCA, staff communicative behaviors towards children increased substantially, particularly following the first five seconds of a child's VOCA use. The authors suggested that these increases were due to the ease in understanding the children's VOCA-mediated communication and subsequent responses in the speech output activations, given the natural form of communication that is inherent in speech output.

These findings, in combination with data that indicate improvements in listener attitudes following VOCA use, have important implications on the social validity of AAC. The use of VOCA may allow for increased opportunities for the user to communicate across a wider range of listeners, settings, and distances. This increase in the amount of potential listeners, in turn, could lead to a higher number and variety of opportunities for communicative exchanges and interactions. Consequently, speech output may positively influence the development of social skills for learners who utilize AAC, given that VOCAs may expose these individuals to interactions and relationships that would otherwise not be available (Drager, Reichle, & Pinkoski, 2010). While further research is necessary to evaluate the effect of VOCA on social skills, many agree that increases in communicative interactions may promote greater community inclusion and participation (Rotholz & Berkowitz, 1989).

Effects of VOCA on Speaker Behaviors

Many studies have also described how the use of VOCA has resulted in various benefits for individuals utilizing VOCA as means of communication. Positive outcomes include successful acquisition and utilization of the device as an AAC, increased comprehension of speech, development of natural speech production, efficient learning of graphic symbols, and decreases in challenging behavior.

VOCAs have been demonstrated to be an efficient AAC system for children and adults with ASD. Children with ASD have been able to effectively utilize VOCA to engage in a number of communicative exchanges, such as requesting preferred objects or activities, identifying environmental stimuli, and providing social commentary (e.g., van der Meer & Rispoli, 2010; Schlosser et al., 1995; McGregor, Young, Gerak, Thomas, & Vogelsber, 1992). Moreover, other studies have demonstrated that skills acquired through VOCA in a classroom setting can generalize across individuals and community settings (Dyches, Davis, Lucido, & Young, 2002).

Researchers have also found positive results for the incorporation of VOCA in treatment packages for challenging behaviors. Durand (1993) examined the effects of using speech-generating devices for functional communication training (FCT) procedures, and found that participants were able to successfully acquire function-based alternative responses with the use of VOCAs, which consequently resulted in a reduction of problem behavior. Follow-up data indicated that the newly acquired alternative responses generalized to community environments and with novel communicative partners following classroom instruction (Durand, 1999).

Many authors have also discussed the influence of digitized speech on listeners' speech production abilities. Paul (1997) proposed that speech output might help children who are currently attempting to vocalize to build on these attempts and facilitate speech acquisition. Persons and La Sorte (1993) also found that children produced higher numbers of spontaneous natural speech during instruction with a speech output device than during a no speech output condition, suggesting that exposure to speech output may in some ways positively influence speech production for participants. Nonetheless, despite these promising results, the overall literature on the role of speech output in the development of natural speech continues to yield mixed results, and additional research is necessary in order to further understand the effects of VOCA on speech production.

Supporters of VOCAs also argue that this AAC modality may have a particular role in speech comprehension. Romski and Sevcik (1993) argued that speech output provides more immediate and consistent vocal models, which may be preferable to individuals with ASD, given the preference for sameness often observed in this population (Rimland, 1964). Specifically, the consistency of the synthetic speech output of VOCAs preserves dimensions of the auditory signal, which may potentially allow the speaker to segment and comprehend a stream of speech more effectively (Romski & Sevcik 1993). This consistency may also facilitate acquisition of VOCA use and subsequent skill generalization (Blischak, 2003; Romski & Sevcik, 1996). Furthermore, increased opportunities for the exposure of speech output may also facilitate speech comprehension (Romski & Sevcik, 1993). Koul and Clapsaddle (2006) found that repeated listening increased accuracy in the comprehension of synthetic speech for adults with mild to moderate intellectual disabilities during word and sentence recognition tasks. Moreover, participants were also able to generalize knowledge of acoustic-phonetic properties of the synthetic speech to novel auditory stimuli. Koul and Hester (2006) found similar results for individuals with severe intellectual disabilities, where performance in word recognition tasks significantly improved following repeated practice. Other studies have also linked repeated exposure to speech output with increases in natural speech comprehension, where VOCA training led to improvements in natural speech comprehension (Brady, 2000). These findings provide strong evidence for the success of VOCA, in comparison to other forms of AAC, in facilitating language comprehension for individuals with ASD and other disabilities.

While VOCAs may emit speech output prior to or following the speaker's communicative behavior, exposure to digitized or synthetic speech following a communicative response (i.e., consequent speech output) may result in additional benefits for the speaker. Research on instructional feedback conducted by Reichow and Wolery (2011) found that children with ASD and developmental delays who were provided with instructional feedback (IF; additional information provided following a correct response in a language task) acquired both the target response and the additional information without additional instructional time, thereby resulting in more efficient instruction. Delmolino, Hansford, Bamond, Fiske and LaRue (2013) also found similar results for one of their four participants, who acquired both the target verbal responses and the content of the IF, despite the fact that there was no additional training for the IF. Given that speech output emitted by the device often follows the speaker's communicative response, it is possible that the speech output emitted by VOCAs following the speaker's response may act as IF. Specifically, consequent speech output may act as IF and become paired with the visual image (i.e., the digital icon pressed during the communicative response) of the vocal label, which could potentially result in improvements in the association of the vocal and visual representations of a word, object, or concept. This is an empirical question that warrants further investigation.

Researchers have also evaluated how speech output may help individuals with intellectual disabilities comprehend graphic symbols, such as images and other pictorial representations of language. Given that many learners with ASD and other intellectual disabilities have demonstrated strengths in the visual modality (Mirenda & Iacono, 1988), clinicians have often developed treatments that incorporate graphic symbols into skill acquisition and behavior modification programs. Consequently, studies that evaluate how speech output influences the learning of graphic symbols and their referents are of great value to the developmental disability and ASD field.

Schlosser, Belfiore, Nigam, Blischajm and Hertzroni (1995) evaluated the effects of speech output on the learning of graphic symbols by isolating speech output as an independent variable in their study. Authors taught two adults with severe mental disabilities to identify a number of familiar and unfamiliar graphic symbols using a VOCA device with activated and deactivated synthesized speech output. In the VOCA condition, researchers provided the instruction "point to____," and pressed the target icon in the communication board, which in turn activated the voice output prior to the response (i.e., antecedent speech output). When participants provided a correct response by pressing down the corresponding target icon, the voice output was activated again, thereby acting as consequent speech output. The non-VOCA condition was identical to the VOCA, with the exception that this condition did not offer speech output upon pressing the target icon. Authors found that, while two of the three participants learned the names of graphic symbols in both conditions, participants made fewer errors and reached mastery criteria with fewer training sessions in the VOCA condition, thus making learning faster and more efficient. Moreover, the third participant acquired the labels for graphic symbols in the VOCA condition only. This study provided compelling evidence that speech output was successful in helping learners with intellectual disabilities associate graphic symbols with their word referents more efficiently. For the participant who exhibited more difficulties, VOCA appeared to have played an even more crucial role in skill acquisition.

Koul and Schlosser (2004) extended Schlosser's study to examine the effects of speech output in the learning of pictorial symbols of high versus low translucency, or graphic symbols that had a low degree of visual resemblance to their referent (i.e., low translucency, LT) versus a high degree of resemblance (i.e., high translucency, HT), for two adults with severe intellectual disabilities. For example, HT symbols included the word "yes" with an image of a face with arrows pointing up and down, while LT symbols included the word "little" with the image of a small dot. Authors hypothesized that speech would be particularly important in facilitating an association between the images and their referent for LT symbols, where the image and word referents were dissimilar. For one of the participants, she learned more LT items in the VOCA versus the non-VOCA condition, and more HT items in the non-VOCA condition. The second participant demonstrated superior learning for both LT and HT items in the VOCA condition. Authors concluded that auditory feedback in the form of speech output might be especially relevant for adults with intellectual disabilities, particularly when they are learning symbols that demonstrate a low degree of similarity with their concrete reference, and therefore cannot be coded solely through the visual modality.

While these studies are promising and clearly demonstrate differential responding in VOCA versus non-VOCA teaching strategies, several procedural issues may have played a role in participants' responding. For example, in the Schlosser et al. study (1995), all training sessions in this study included simultaneous prompting, where staff instructed participants to point to the target and then immediately modeled the correct response. Independent probes, where students were asked to identify the correct image without prompting strategies, were not conducted until maintenance. Thus, it is unclear whether the rates of correct responding during the acquisition phase represent increasing compliance to a model prompt or true acquisition of the graphic symbol. Moreover, the independent probes that took place during the maintenance phase resulted in more variable performance across targets between conditions for all participants. Additional studies that incorporate modeled and independent training sessions are necessary to further understand the role of speech output in learning for individuals with disabilities.

The populations targeted in these studies also limit how these results could be generalized to others. In both studies, authors evaluated speech output efficacy in learning for adults with developmental disabilities; therefore, it is unclear whether these results could be replicated with participants with ASD. Moreover, although previous research has focused on adult participants, it is necessary to extend this research to younger participants, given that communicative impairments are a defining component of ASD and are therefore present prior to adulthood.

Furthermore, since AAC interventions are typically carried out during childhood rather than adulthood years, it is then necessary to evaluate whether the empirical evidence found in the speech output literature extends to younger participants. Additional studies that attempt to replicate these findings in a school-aged ASD population are therefore necessary to provide further evidence for the benefits of speech output in learning.

The present study aims to contribute to the VOCA literature by further evaluating the effects of speech output technology in learning for a school-aged participant with ASD. Specifically, the purpose of this study is to evaluate whether results from previous investigations can be replicated by utilizing digitized speech technology to teach a school-aged child with ASD to identify graphic symbols.

Chapter II

Method

Participant and Setting

Alan is a 12 year, 8 month old student diagnosed with ASD who currently attends a center-based school program for individuals with ASD. While he exhibits some difficulty with pronunciation and clarity, such as producing the incorrect consonant sound in a word or skipping the first syllable, Alan mainly communicates independently and vocally with others at his school. Sessions were run in a large classroom designed to practice life skills, which contained a dining table, a kitchen, stove, and refrigerator, and a bed with a dresser. For the majority of sessions, only Alan, his assigned classroom instructor, and the principal investigator were present in the room. Alan sat at the end of a large table, and his instructor sat diagonal from him.

Materials

The targets consisted of laminated cards attached to a speech output device. The cards were 3x3-inch visual images of unfamiliar animals, foods, and objects randomly chosen to serve as potential targets for the study. In addition, Talk Point™ Mini-me devices were utilized as the speech output devices. These circular devices, measuring 1 ¾ inches in diameter, record up to 10 seconds of clear sound, which is produced upon pressing on the surface of the Mini-me. In the SO condition, the investigator recorded the vocal label of the target images and then attached the images to the Mini-me containing their corresponding label with Velcro®, so that when the participant touched the card, he also pressed down on the Mini-me, resulting in the emission of the recording (i.e., the image's vocal label). The Mini-mes attached to the visual images are referred to as the target stimuli.

Dependent Measures

Data were collected via paper and pencil data sheets for baseline, training, independent, and maintenance sessions. Trial-by-trial data were collected for correct responses, incorrect responses, or no response. Session data included percentage of correct responses, percentage of training errors, number of sessions at or above 80% mastery, and sessions to criterion. The following dependent measures were recorded.

Correct Response. Correct responses were defined as any instances in which Alan independently touched or pressed a target stimulus that corresponded to the discriminative stimulus (S^d) presented by the instructor. For example, when Alan touched the target stimulus for "toucan" following the S^d, "touch the toucan," the response was recorded as correct. Additionally, responses must have occurred no later than six seconds after the presentation of the S^d to be considered correct.

Incorrect Response. Incorrect responding was defined as any instance in which the participant pressed a target stimulus within six seconds that did not correspond to the S^d presented by the instructor.

No Response. No response was defined as any instance in which the participant did not touch any of the three target stimuli within six seconds of the presented S^d.

Percentage of Correct Responses. Percentage of correct responses was calculated by dividing the number of correct responses in each session by 15 (the total number of opportunities for responses), and multiplying by 100%.

Percentage of Training Errors. Percentage of training errors was calculated by dividing the number of incorrect responses in each session over 15 (the total number of opportunities for responses), then multiplying by 100%. No responses were not counted as errors.

Sessions to Criterion. Sessions to criterion was calculated by counting the total number of sessions until the participant reached mastery criteria, which was defined as reaching at least 80% accuracy over two consecutive sessions in the independent phase, and three consecutive sessions in the training phase. Mastery criteria was more stringent during the training session in order to ensure that Alan had received high amounts of exposure to correct responding prior to moving to the more challenging independent phase.

Sessions above Mastery. Sessions above mastery was calculated by counting the total number of sessions within a condition that had a percentage of correct responses at or above the 80% mastery line.

Vocal Approximation. Vocal approximation was defined as emitting, in the correct order, at least two of the consonant or vowel sounds composing the vocal expression of a target label. For example, "io-een," "ah-oh-leen," or "vah-leen" would all be coded as a vocal approximation for "violin."

Experimental Design

The current study utilized a single case, multielement design with multiple baselines to evaluate the effects of the independent variable across three separate and independent target sets. Each tier was composed of a specific set of randomly assigned targets (selected via the pre-assessment), and included the following phases: initial baseline (and baseline probes for Tier 2 and 3), training, independent probes, and maintenance. The multiple baseline design was selected in order to demonstrate experimental control of the teaching procedures by replicating the effects of the independent variable across three sets of novel targets. The multielement component of the design was included to evaluate differences in the acquisition of target stimuli in the Speech

Output (SO) and No Speech Output (NSO) conditions by alternating between conditions within each tier.

Following target selection via pre-assessment, initial baseline data were collected for both conditions in Tier 1, and probes were collected for Tiers 2 and 3. Prior to initiating training for Tier 1, baseline probes were again conducted for both conditions in Tiers 2 and 3. Once Alan completed training and reached mastery criteria in the independent phase for Tier 1, a short baseline for Tier 2 was collected, along with baseline probes for Tier 3. Training in Tier 2 was then initiated, and once Alan reached mastery in both training and independent probes, baseline for Tier 3 was initiated, and followed with the subsequent phases of training. Maintenance data were also collected for each tier at four, eight, and 12 weeks after mastery in order to evaluate differences in skill retention between the SO and NSO condition.

Procedure

The procedure for this study was informed by the protocol developed and carried out by Schlosser et al. (1995). While the general design of Schlosser and colleagues was maintained, additional components were added to the protocol in this investigation. These included preference assessments, systematic reinforcement according to phase, a modified number of targets per set, an independent probe phase, and three maintenance probes post-mastery.

Reinforcement. In order to maintain Alan's motivation and compliance throughout the various phases of the study, staff provided access to preferred items contingent on appropriate behaviors during baseline and maintenance sessions, and contingent on correct responding during training and independent phases. Since Alan typically utilized a choice board during regular academic sessions to request preferred activities or toys, the choice board was incorporated into the experimental sessions. Prior to starting a session, instructors waited for

Alan to verbally request a previously established reinforcer from his choice board, such as tickles, Ipad®, computer, and coloring. Following his request, instructors then placed a picture of the chosen item on Alan's token board, and stated that he would gain access to the item after earning all four tokens on his board. This procedure was repeated at the beginning of every research session and following the end of a reinforcement period to ensure that Alan received items or activities that were the most reinforcing to him at the time of the session.

Pre-Assessment. In order to control for learning history as a confounding variable, researchers ensured that the training materials used in this investigation were unfamiliar images that Alan could not correctly tact prior to the procedure. In this way, changes in performance could be more likely attributed to the independent variable rather than to previous exposure and experience with the tact. Thus, the purpose of the pre-assessment phase was to identify the set of graphic symbols to utilize as targets during teaching. In this phase, researchers presented graphic symbols, in the form of 3x3 laminated cards, of household objects, animals, foods, and transportation vehicles in a field of three. Probes were conducted within category to reduce the possibility of correct responding by discrimination; that is, the field of three always contained items from the same category. Once Alan independently identified a reinforcer to earn, the instructor placed the token board on his desk, and stated that he would gain access to his reinforcer once he earned four tokens. Then, the instructor placed three pictures on Alan's desk and provided the appropriate S^d for each target (e.g., "touch the bongo"). Instructors replaced all three images after two consecutive probes to reduce the possibility of correct responding by elimination, i.e., pressing the remaining target for the third S^d presented. Furthermore, instructors provided neutral response (e.g., "okay") or no feedback following Alan's responses, and instead awarded Alan tokens for attending behavior, such as keeping his hands down, staying in his seat, or making eye contact with his instructors, on an FR4 schedule. A target symbol was added to the unfamiliar target pool when Alan provided incorrect responses for that symbol in two out of three trials.

Alan's performance in the preassessment resulted in the identification of 20 potential targets. From the pool of unfamiliar targets, 18 symbols were randomly selected and assigned to the SO and NSO conditions, regardless of category. Table A depicts the final distribution of targets, arranged by tier and condition.

Table 1.
Randomization of Targets by Tier and Condition.

Condition	Tier 1	Tier 2	Tier 3
No Speech Output (NSO) Condition	Trombone Jeep Toucan	Beets Taxi Drums	Clarinet Oven Asparagus
Speech Output (SO) Condition	Bongo Helicopter Hammer	Aquarium Spinach Violin	Wrench Coat Rack Peas

The randomization of targets into groups to create non-overlapping sets for each tier and condition meant to address the threat of carryover effects that many studies face due to the use of the same symbols and words referents in both conditions (Schlosser, 2003). Nine targets were randomly assigned to each condition. These nine items were then randomly split into three sets for each tier, for a total of three stimuli per condition per tier. In this way, Tier 1 had a distinct set of SO and NSO targets, as did Tier 2 and Tier 3. Following target randomization, each target was attached to a Mini-Me, thereby creating a target stimulus.

Target stimuli were presented within their specified sets, so that the same targets were presented together throughout the phases of the study. For example, the targets "bongo," "helicopter," and "hammer" were assigned to the SO condition set for Tier 1; consequently, those three targets were consistently presented together during baseline, training, independent, and maintenance trials, where each stimulus acted act as both a distracter and a target stimulus within the set. In this way, it was ensured that Alan learned conditional discrimination, where his responses are under the stimulus control of the S^d (e.g., the Alan touches "bongo" when he hears, "touch bongo") rather than the contingency itself (e.g. Alan always touches the bongo image when the images of bongo, helicopter, and hammer are presented).

Baseline. A consecutive baseline was established for Tier 1, and baseline probes for tiers 2 and 3 were carried out at the beginning and end of Baseline 1. In this phase, the instructor presented Alan with his choice board and waited for him to identify a reinforcer for the session. The instructor then placed a picture of the reinforcer on Alan's token board, and stated that he could gain access to the reinforcer after he earned his four tokens. The instructor then waited for Alan to demonstrate attending skills (e.g., made eye contact with his instructor or placed his hands on his lap). Once Alan attended to his instructor, he was then presented with the three target stimuli assigned to the current condition, and the instructor delivered an S^d containing the vocal label of the target stimulus (e.g., "touch bongo"). Trials were recorded as correct, incorrect, or no response. If Alan did not respond to the instructor's S^d after six seconds, the instructor then moved on to the following S^d. The order of the target stimuli was altered after no more than two instructor S^ds. No corrective feedback or reinforcement was provided upon Alan’s response or nonresponse. Instead, only intermittent, nonspecific praise (e.g., "thanks for showing me") was provided, as well as tokens on an FR4 schedule for attending behaviors. Each session contained five interspersed presentations of each target, resulting in a total of 15 training trials per session. Furthermore, the speech output feature was turned off for both the SO and NSO target sets during baseline.

Training. Following baseline probes, training began. In this phase, Alan was trained to identify the target stimuli until he accurately identified targets in at least in 80% of trials across three consecutive training sessions in each of the conditions. Following the procedure published in Schlosser and colleagues (1995), instructors employed a simultaneous prompting strategy for training, which involved the pairing of the verbal S^d with the immediate model of the correct response (Schuster et al., 1992; as cited in Schlosser at al., 1995). Training procedures were identical in the SO and the NSO conditions, with the exception that the speech output devices were turned on during the SO condition, and off during the NSO condition. Consequently, following a correct response in the SO condition (i.e., touching the correct target stimulus corresponding to the vocal label), the speech output component of the Mini-me was activated, resulting in the production of the vocal label in the form of synthetic speech. During the NSO condition, Alan was also required to touch the correct target stimulus, but upon doing so, no speech output was emitted.

During each training session, the instructor first waited for Alan to identify his reinforcer for the session, and then placed a picture of the reinforcer on his token board. Once Alan exhibited appropriate attending skills, the instructor then presented Alan with the three target stimuli assigned to the current condition, then delivered the S^d containing the vocal label of the target stimulus (e.g., "touch helicopter") and immediately modeled the correct response by pressing on the corresponding target stimulus. In the SO condition, pressing a target stimulus resulted in consequent speech, while in the NSO condition, the speech output feature was turned off. Instructors also shadowed Alan's hands to indicate that he should wait until the end of the S^d, the model, and the vocal output emission following the model (for SO targets only), prior to responding. The instructor gave Alan high quality social praise for every instance that he followed the instructor's model and provided a correct response. Alan also received a token contingent on correct responding on a VR4 schedule. If Alan provided an incorrect response or did not respond six seconds after the S^d and model, the instructor implemented an error correction procedure. This consisted of the instructor stating, "*This* is ____," and immediately pressing the correct target stimulus. If Alan then provided a correct response, the instructor provided low quality social praise (e.g., "yes, that's it") and moved on to the next target. On the other hand, if Alan again provided an incorrect response or nonresponse, no consequence was given and the instructor moved to the next target. Each session contained five interspersed presentations of each target, resulting in a total of 15 training trials per session. Mastery was achieved when Alan followed the simultaneous prompt and provided correct responses in at least 80% of opportunities (i.e., 12 of the 15 trials) across three consecutive sessions in both conditions. While mastery criteria initially required only two consecutive sessions above 80%, Alan's dip in performance during Tier 1 training led the investigators to add an additional training session for each condition to ensure stable responding. This change to mastery criteria was applied across tiers and conditions to ensure consistency.

Independent Probes. Following mastery in the training condition, independent probes were conducted. The purpose of this phase was to evaluate Alan's rate of correct responding when asked to identify a target stimulus without additional prompting. Thus, instructors followed the protocol described in the training phase, but omitted the simultaneous prompting sequence. For these sessions, instructors waited for Alan to identify his reinforcer, placed the corresponding icon on the token board, and once Alan demonstrated attending skills, they presented the S^d, e.g., "touch the bongo." Instructors then waited six seconds for Alan to respond. Independent probes were identical in the NSO and SO condition, with the exception that the speech output component was turned on for the SO targets and off for NSO targets. Therefore, Alan was exposed to the speech output corresponding to the target stimulus he pressed following the S^d. In this way, the speech output acted as confirmatory feedback when his response was correct, and corrective feedback when his response was incorrect. In both conditions, if Alan provided a correct response, the instructor immediately reinforced the response with social praise and delivered a token on Alan's board on a FR1 reinforcement schedule. If Alan provided an incorrect response or did not respond six seconds after the S^d, the instructor implemented the error correction procedure described in the training session. In the NSO condition, Alan was exposed to the model of the correct response. In the SO condition, Alan was exposed to both the model and speech output of the correct response. Independent probes continued until Alan met mastery criteria of 80% or above in two consecutive sessions of both conditions.

Vocal Probes. Throughout the independent phase of Tier 1, investigators observed that Alan had begun attempting to echo some of the targets' vocal labels. Therefore, starting with Tier 2, vocal probes were systematically introduced to the protocol to evaluate whether training a receptive identification skill with and without speech output could also lead to changes in Alan's ability to expressively tact the target stimuli. Immediately following mastery of Tier 1 targets, instructors began Tier 2 training with vocal probes, followed by the baseline, training, independent, and maintenance phases. Procedures for reinforcer selection and establishing attending were identical to the other phases. The probe consisted of the instructor raising one image at a time and asking, "What is it?" Instructors then waited for a response for six seconds, and coded them as correct, approximation, incorrect, or no response. Similar to the baseline probe, no corrective feedback or reinforcement was provided upon responses, and instructors gave a neutral statement to each response, such as "Okay" or, "Thanks." Similar to baseline, instructors awarded tokens to Alan on an FR4 schedule contingent on attending behaviors. Fifteen trials were conducted for each session, where each target was presented individually five times, and one vocal probe session of each condition was conducted prior to baseline (ie., pre-test), immediately following mastery in the independent phase (i.e., post-test), and at each maintenance probe.

Maintenance Probes. Maintenance probes were conducted at four, eight, and 12 weeks after Alan met mastery for each of the three tiers in order to assess any differences in skill retention between conditions. Sessions were identical to baseline, in which instructors had Alan select a reinforcer for his token board and established attending behavior. Then, instructors asked Alan touch the tact stimulus corresponding to the specified S^d in a field of three. Alan received reinforcement on an FR4 schedule contingent only on appropriate behaviors. Reinforcement based on correct responding was not provided in an effort to avoid within-session learning following reinforcement of correct responses, which would then lead to a false inflation of correct responses rather than true maintenance. Furthermore, the speech output component of target stimuli was turned off for both conditions, given that the auditory feedback could also result in within-session learning.

Interobserver Agreement (IOA) and Treatment Integrity (TI)

IOA and TI data were collected by trained undergraduate and graduate research assistants, who viewed and coded videos of 34% of all sessions. Training for coders consisted of providing a written set of instructions, modeling correct coding with a sample video, watching trainers collect sample data, providing immediate feedback, and rewatching coding session to ensure feedback incorporation. Sessions were randomly selected for IOA and TI checks, and efforts were made to ensure similar percentages were taken from all phases and conditions.

Furthermore, coders were given a list of session numbers to code, and were asked to randomly select sessions for coding. Given that TI data sheets varied by phase and condition, coders were informed of conditions and phases for each session number in order to ensure correct data collection.

To calculate trial by trial IOA, independent recorders viewed a filmed session, and subsequently recorded Alan's responses as correct, incorrect, or no response. IOA for each session was calculated by coding agreement or no agreement between recorders' responses for each of the 15 trials composing a session, then dividing the number of trials with agreement by 15 and multiplying by 100%. All session agreement coefficients were then averaged to reveal a total IOA of 97.1% (range 95.3%-100%) across phases and conditions.

Independent recorders also collected TI data for 34% of all sessions. Sessions were evaluated according to procedural components specific to phase and condition. Components included conducting a preference assessment at the beginning of session, establishing attending, presenting a clear S^d, ensuring that speech output was turned on/off in the SO/NSO sessions, accordingly, providing appropriate reinforcement, etc. Recorders coded for the presence of each procedural component in the 15 trials that encompassed a session. If the instructor implemented all components of the procedure in a single trial, then that trial was scored as having 100% integrity. If an instructor failed to present one or more components on a given trial, then that trial was scored as having 0% integrity. Session TI was then calculated by dividing the number of trials with full integrity by 15, and then multiplying by 100%. All session coefficients were then averaged to reveal a total TI of 95.5% (range 87% - 100%).

Chapter III

Results

Results are presented according to dependent variable, which include percentage of correct responses, percentage of errors, sessions at or above mastery criteria, and sessions to criterion. Then, results from the expressive identification probes are discussed in terms correct responses and approximations in pre-post testing and maintenance.

Percentage of Correct Responses

Figure 1 depicts Alan's percentage of correct responses across tiers and between conditions. The top panel depicts Alan’s performance on Tier 1. During baseline, Alan's percentage of correct responses was equal between conditions. He scored an average of 21.5% accuracy for both the SO and NSO targets. Then, in the training phase, Alan exhibited an average of 93.3% of correct responses following simultaneous prompting in SO condition, and 95.5% in the NSO condition. Alan's accuracy dropped from 100% to 80% in the SO condition after two training sessions, therefore, an additional session was run for each condition, totaling three training sessions per condition. This extension of training was applied to subsequent tiers in order to ensure consistency across targets and tiers. Following training, independent probes resulted in overall higher performance for SO than NSO targets, where Alan correctly identified at least 80% of all targets in nine of the 12 sessions (*M*=84.42%) for SO targets, and in four of 12 the sessions (*M*=69.5%) for targets in the NSO condition. Following mastery, three maintenance probes were conducted for both conditions. Alan scored 100% accuracy for SO targets and 80% accuracy for NSO targets four weeks after mastery. At eight weeks, he scored 100% for both conditions, and at 12 weeks, he scored 100% for SO targets, and 73% accuracy for NSO targets.

In this tier, Alan's accuracy was both higher and more stable for targets that were taught with activated speech output.

The middle panel depicts the results for Alan's performance on Tier 2. Alan's accuracy during baseline was low for both conditions, with an average of 31.75% of correct responses for SO targets, and 30.25% for NSO targets. During training, Alan did not require more than the minimum 3 sessions for each condition to meet training mastery, and exhibited an average of 95.67% accuracy for SO targets, and 100% accuracy for NSO targets. In the independent phase, Alan's accuracy was significantly higher with SO targets than NSO targets, where 11 of the 15 independent probe sessions in the SO condition resulted in accuracy of 80% or above (*M*=89.73%), as compared to six of the 15 sessions (*M*=75.13%) in the NSO condition. Furthermore, Alan's performance during the four-week probe resulted in 73% and 60% accuracy for SO and NSO targets, respectively. He then scored 87% accuracy for SO targets and 53% accuracy for NSO targets eight weeks after mastery. Furthermore, 12 weeks post-mastery, Alan was able to correctly identify 93% of SO targets and 80% of NSO targets. Overall, Alan's performance in this tier revealed higher accuracy and more consistent responding for SO than NSO targets.

The bottom panel depicts responding in Tier 3. Alan's accuracy during baseline was comparable between conditions, with an average of 27.83% accuracy with SO targets and 28.83% accuracy with NSO targets. Similar to the previous tiers, Alan met mastery criteria with the minimum amount of sessions during the training phase, with an average of 97.67% and 100% accuracy for SO and NSO targets, respectively. Alan's rate of correct responding decreased for both conditions during the independent phase, although Alan continued to score higher rates in the SO condition (six of nine sessions above 80%; *M*=74.22%) versus the NSO condition (three of nine sessions above 80%; *M*=59.33%). Furthermore, while Alan met mastery with SO targets after the fourth session of the independent phase, his accuracy then dropped below 80% in the following session. This effect was likely due to the introduction of a novel instructor, given that Alan was observed having difficulty attending to materials and responding to the instructor. Consequently, original instructors were reintroduced for subsequent sessions, and accuracy returned to mastery levels. Finally, maintenance probes for Tier 3 yielded variable rates of accuracy for both conditions. Four weeks after mastery, Alan's performance resulted in 47% and 40% accuracy for SO and NSO targets, respectively. Alan then scored 100% accuracy and 47% accuracy for SO and NSO targets, respectively, at the 8-week probe. Twelve weeks after meeting mastery criteria, Alan scored 33% accuracy for both conditions. Alan's maintenance data were more unstable during this tier, resulting in no strong differentiation in learning between conditions.

Overall trends across tiers revealed that, while Alan’s performance was comparable between conditions during training, his overall percentage of correct responses during independent probes was significantly higher for SO targets than NSO targets. His performance resulted in more sessions at 80% or more accuracy in SO sessions (72.67% of sessions at or above criterion line) than in NSO sessions (35.56% of sessions at or above the criterion line). Results then indicate that Alan’s performance in receptive identification tasks was significantly higher and more accurate when images were accompanied with consequent speech output (i.e., when speech output followed his response) than when they were not. Furthermore, despite some variability, overall trends in maintenance data suggest that speech output may have contributed to more stable responding and higher rates of accuracy following mastery.

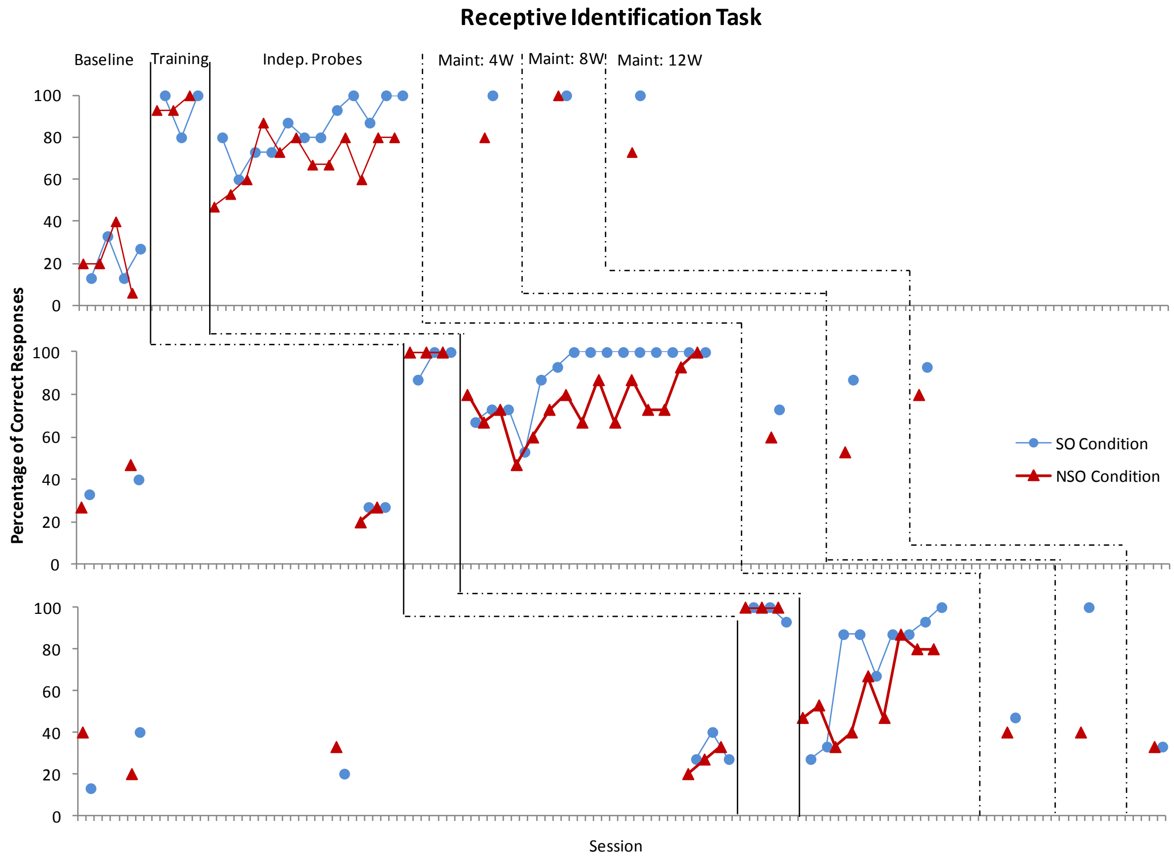


Figure 1. Alan's percentages of correct responses for the receptive identification task for each tear in the speech output (SO) and no speech output (NSO) conditions

Sessions to Criterion

Figure 2 depicts sessions to criterion during the independent phase for each tier, as well as the mean number of sessions per condition. The number of independent sessions Alan completed prior to meeting mastery criteria (i.e., two consecutive sessions at 80% accuracy or above) was lower for SO sessions than for NSO sessions across tiers. In Tier 1, Alan reached mastery criteria for SO targets after seven sessions, while he required 13 sessions to reach mastery in the NSO condition. Similarly, in Tier 2, Alan met mastery in six versus 15 trials for the SO and NSO targets, respectively. Finally, in Tier 3, Alan met mastery after four independent sessions for SO targets, while he met mastery for NSO targets after nine sessions. These results indicate that, on average, Alan required about half the number of sessions to learn novel targets when visual stimuli were presented with speech output (*M*=5.67) than when they were presented with no speech output (*M*=12.33), therefore making sessions more efficient in the SO condition than in the NSO condition.

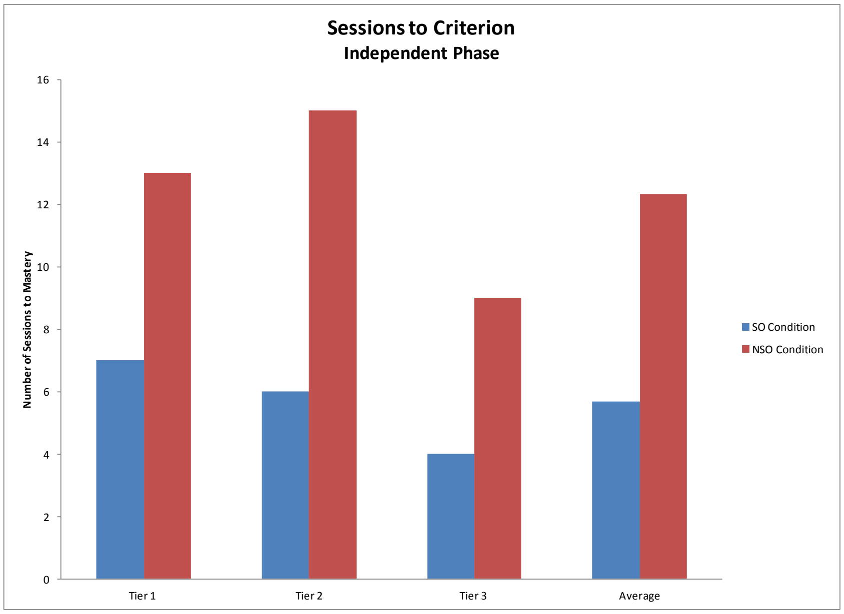


Figure 2. Number of independent sessions Alan completed prior to meeting mastery criteria for each tier and on average for the speech output (SO) and no speech output (NSO) conditions.

Percentage of Errors

Figure 3 illustrates Alan's percentage of errors during independent and maintenance phases. Alan made significantly fewer errors with SO targets across tiers, as compared to NSO targets. The top panel depicts percentage of errors during independent and maintenance probes for Tier 1. During independent probes, Alan's responses were, on average, incorrect on 17.75% of opportunities and 30.50% of opportunities for the SO and NSO conditions, respectively. Maintenance probes for this tier also revealed higher performance for SO targets versus NSO targets, where Alan made no errors with SO targets, but had an average of 16% inaccuracy with NSO targets. Differences in training errors were even more notable during Tier 2 probes (middle panel). In this Tier, Alan made errors during 8.67% of trials in the SO condition, and 20.81% of trials in the NSO condition. Maintenance probes for Tier 2 also revealed fewer errors in the SO condition, where Alan scored an average of 13.67% inaccuracy with SO targets and 31% inaccuracy with NSO targets. Finally, in Tier 3 (bottom panel), Alan continued to make fewer errors in the SO condition than the NSO condition, resulting in average error percentages of 24.44% and 40%, respectively. This trend also continued in maintenance probes, where his performance with SO targets resulted in an average of 26.67% errors, whereas he scored, on average, 57.67% inaccuracy with NSO targets. Alan's overall performance suggest that speech output may have aided Alan in making less errors during the receptive identification task.

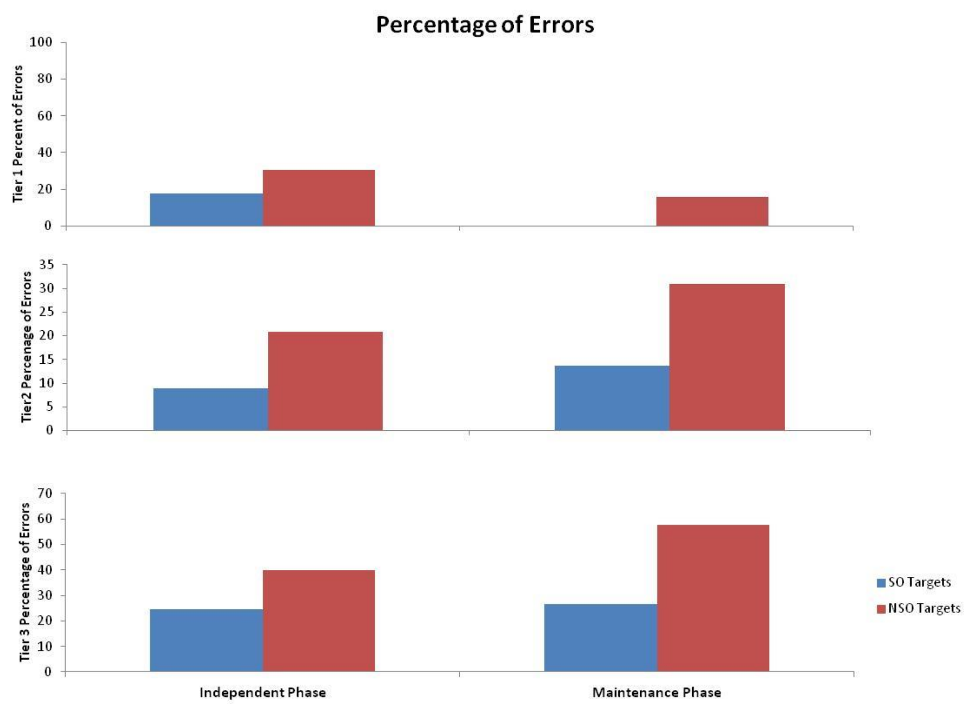


Figure 3. Average percentage of incorrect responses Alan exhibited during the independent and maintenance phases in each tier for targets in the speech output (SO) and no speech output (NSO) conditions.

Expressive Identification Probes

Figure 4 depicts percentages of correct responses (i.e., approximations or full words) for vocal probes during pre-training, post training, and maintenance conditions for Tier 2 (top panel) and 3 (bottom panel). Following observations that Alan attempted to echo vocal labels during Tier 1 training, formal data collection for expressive identification of targets began at the start of Tier 2. During pre-test for Tier 2, Alan did not provide answers for any of the SO targets, resulting in 0% correct responses. For the NSO targets, Alan identified “drums” three out of the five times this target was presented but no correct responses for other targets, resulting in 20% accuracy for pre-training. Immediately following mastery in both conditions, a post-test vocal probe was conducted in which Alan was again asked to expressively identify each stimulus, presented one at a time. During the post-test for the SO condition in Tier 2, Alan provided either an approximation or a full correct label in 73% of the trials, as compared to 40% of trials in the NSO condition. During maintenance probes for Tier 2, Alan's accuracy decreased across maintenance probes in the SO condition (Maintenance Probe 1=33%, Maintenance Probe 2=13%, and Maintenance Probe 3=7%). Alan’s responding was more stable for the NSO condition (Maintenance Probe 1=40%, Maintenance Probe 2=33%, and Maintenance Probe 3=27%). Differences between conditions during this phase mimicked those observed during pre-test, where Alan had higher accuracy scores with NSO targets. Therefore, data may be potentially skewed by Alan's accuracy score with one specific NSO target, "drums," which he correctly tacted in almost every presentation during all probes, thus contributing up to 20% of correct responses per session.

Vocal probes for Tier 3 (bottom panel) yielded similar trends. In Tier 3, Alan provided no responses for the SO and NSO conditions during pre-test, resulting in 0% accuracy for both conditions. During the post-test, Alan provided approximations or correct responses for 60% of SO trials and 40% accuracy in the NSO condition. Nonetheless, maintenance probes showed a significant decrease in responding. Alan only provided responses during Maintenance Probe 1, where he scored an average of 13% accuracy with SO targets, and 7% accuracy for NSO targets. Probes for SO conditions conducted eight and 12 weeks after mastery all resulted in 0% accuracy, where Alan most often provided no response, rather than incorrect responses, to the expressive identification task. Data thus suggest that, while the addition of a speech component during receptive identification training may result in the immediate, non-targeted learning of the vocal label for stimuli, it appears that speech output may not have as clear an impact in long-term retention of expressive identification skills.

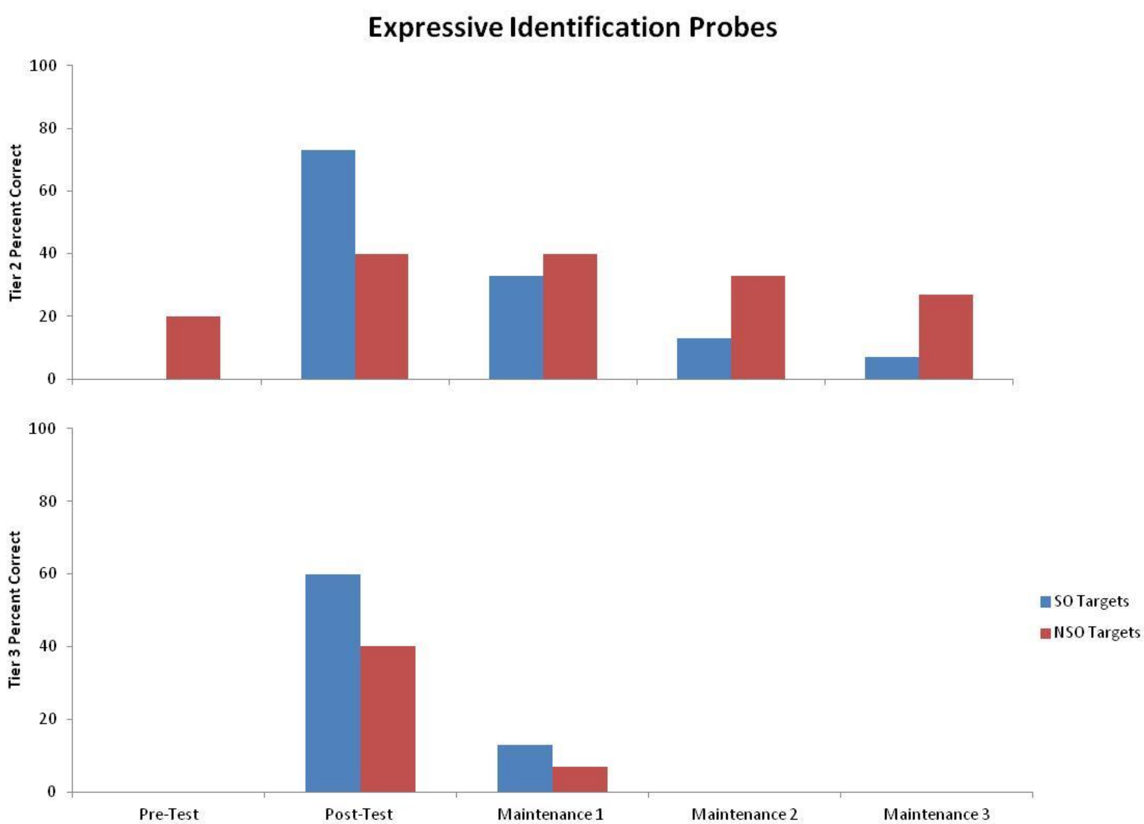


Figure 4. Percentage of approximations and correct responses Alan exhibited in the expressive identification task at pre-test, post-test, and maintenance probes for speech output (SO) and no speech output (NSO) targets in tiers 2 and 3.

Chapter IV

Discussion

The present study aimed to evaluate whether previous findings on the effects of speech output technology on the learning of novel images could be replicated with a school-aged participant diagnosed with ASD. Alan, a 12-year old male, was taught to identify novel visual stimuli with and without synthetic speech output across three separate sets of targets.

Results showed that Alan's percentages of correct responses were consistently higher for the SO condition than the NSO condition across all tiers, indicating that the incorporation of speech output in a teaching task facilitated learning for Alan. He also made significantly fewer errors in the SO condition than the NSO condition, which consequently resulted to a more limited exposure to error correction procedures in the SO condition. These general trends were also observed during maintenance probes, albeit with more variability in performance. Furthermore, while Alan eventually mastered all targets in both conditions, he met mastery criteria for SO targets in about half the number of sessions required for mastery of NSO targets across tiers, indicating higher efficiency in the SO sessions. Findings from this investigation thus provide strong preliminary evidence for the benefits of speech output in skill acquisition for children with ASD, both in terms of student accuracy and session efficiency. Moreover, lower percentages of errors in the speech output condition also suggest that speech output may be a particularly beneficial tool for individuals who perform best under errorless learning procedures (Touchette & Howard, 1984). While additional studies may be necessary to further corroborate data gathered in this investigation, these findings indicate that there may be a benefit to considering VOCA and other speech output technologies as potential supplements for communication and academic training for school-aged children with ASD.

While the main purpose of this study centered on the effects of speech output on a receptive identification task, secondary gains were also observed in Alan's ability to expressively tact the novel stimuli. Following instructors' observations that Alan attempted echo the vocal labels of Tier 1 targets, pre-post tests of expressive identification were formally assessed in the subsequent tiers. During these trials, reinforcement was only provided contingent on attending behaviors to control for within-session increases in accuracy due to the reinforcement of correct responses. These measures showed a moderate increase in untaught expressive identification accuracy for NSO targets, indicating that the receptive identification procedure itself may have primed Alan to vocally identify the targets. In contrast, higher percentages of correct responses for the untaught expressive labels were observed for the SO targets, suggesting that speech output may have played a role in boosting Alan's accuracy in the untaught vocal labels. While expressive identification maintenance data revealed an overall decline in responding for both conditions across time, this change in performance could have been influenced by changes in reinforcement, since none was provided for correct responding during vocal probes. Nonetheless, increases in accuracy during pre-post testing provide strong incentive for further evaluating the effects of speech output on vocal language learning.

Pre-post vocal probe findings are similar to results Schlosser and colleagues (2007) found for one of their participants, Michael, whose natural speech appeared to improve following requesting training with the use of VOCA. Authors reported that Michael was the only participant in their study to possess some vocal imitation skills prior to baseline; a skill that Alan also exhibited. Alan's performance is also consistent with findings by Blischak (1999), who found in her study that children who utilized synthetic speech output demonstrated marked increases in natural speech production while receiving individual phonological awareness instruction. Furthermore, Alan's unexpected increases in untaught vocal labels following the learning task support Paul's (1997) hypothesis that speech output may help children who are currently attempting to vocalize to build on these attempts and facilitate speech acquisition.

The current study provides promising data on the potential emergence of untaught skills following the use of speech output technology. However, further studies that isolate potential confounds should be conducted. More specifically, individual-specific variables, such as Alan's imitation skills and pre-existing phonetic repertoire, as well as procedural factors, including frequency of exposure to the vocal label during training, may have played a role in Alan's performance in the expressive identification task. Increasing participant numbers to include children with varying levels of imitation and expressive vocabulary skills may provide further information on the effects of speech output on expressive identification according to participants' skill set. Future research could also control for potential effects of exposure to labels by assigning a specified number of times each participant is exposed to a target vocal label as they undergo the learning procedure, given that increased exposure (and therefore practice with the label) could potentially lead to increased performance. Furthermore, given Alan's observed declines in performance in both conditions during maintenance probes, it is important to evaluate whether booster sessions, or increased exposure to targets, may aid in maintaining accuracy rates following formal training and mastery. Thus, additional studies focusing more explicitly on untaught skill emergence during speech output training, and the potential underlying mechanisms responsible for this phenomenon, could yield more conclusive evidence on this potential benefit of speech output technology.

In addition to further evaluate variables that influence untaught skill emergence, future research should also evaluate what components within our procedure contributed to the emergence of targeted skills, particularly in terms of antecedent versus consequent speech output. During the training phase in the SO condition, Alan was exposed to speech output prior to his response, as part of simultaneous prompting; and following his response, after he pressed a target stimulus. However, despite staff efforts to delay responding, Alan often did not wait until the antecedent speech output was fully emitted prior to responding, and instead followed the physical model of the correct response almost immediately after observing it. This observation, in addition to data showing comparable performance at or above 80% accuracy in training sessions in both the NSO and SO conditions, suggests that speech output (whether antecedent or consequent) did not play a significant role during training, and that increases in accuracy were more likely attributable to the physical model built into the simultaneous prompting strategy. Dips in performance when Alan moved from training to the independent phase, reaching close to baseline levels in Tier 3, further suggest that Alan's accuracy in both conditions was dependent on the model prompt.

During the independent phase, Alan was exposed to similar changes between the SO and NSO conditions, such as the simultaneous prompting sequence for error correction and reinforcement on an FR1 schedule, while consequent speech output was present only in the SO condition. This speech output was either confirmatory (when he pressed the correct target stimulus), or corrective (when he pressed the incorrect target stimulus and listened to the incorrect vocal label). In this phase, Alan was observed to attend to consequent output following his responses, and often waited for speech output to be completed prior to looking at his instructor for reinforcement or error correction. Therefore, data from the independent phase indicate that, while the error correction and reinforcement strategies this phase may have had positive effects on learning in both conditions, the consequent speech output may have been instrumental in boosting Alan's accuracy and efficiency in the SO condition. These findings provide valuable information on the potential benefits of antecedent versus consequent speech output as teaching strategies for individuals with disabilities since, for Alan, antecedent speech output appeared to have little effect on his learning of novel stimuli, while consequent speech output appeared to have significant implications on his performance on learning tasks.

Findings also indicate that consequent speech output may have acted as instructional feedback, given that Alan acquired the untaught response (i.e., the expressive label) during the receptive identification task. Data then support previous findings by Reichow and Wolery (2011), as well as Delmolino and colleagues (2013), on the acquisition of untaught responses when it follows correct responses of target tasks. Further study is needed to more systematically explore the benefits of antecedent and consequent speech output, as well as the potential role of speech output as instructional feedback for learners.

While the present study found important preliminary evidence for the effects of speech output on word and image learning in ASD populations, certain limitations require readers to interpret results with some caution. Given that this study included only one participant, it is unclear whether similar results would be produces with other individuals with ASD. Alan possessed some vocal imitation skills as well as history of communicating verbally with some success, consequently, whether results from this study would generalize with participants with less successful verbal abilities is an empirical question that requires additional inquiry.

Furthermore, while this study assessed for skill maintenance, it did not probe for generalization. This may have been a particularly important dimension of learning to further evaluate for Alan, given that, in the single instance where a novel instructor conducted an independent probe in the SO condition of Tier 3 (see arrow in Tier 3 of Figure 1), Alan's correct responding decreased despite the fact that he had previously reached mastery. Nonetheless, after familiar instructors were reintroduced for the following sessions, accuracy returned to mastery levels. Given that one of the main benefits of VOCA discussed in the literature include the notion that it provides the speaker with increased opportunities to communicate across settings and individuals (Rotholz & Berkowitz, 1989; Dyches, Davis, Lucido, & Young, 2002; Drager, Reichle, & Pinkoski, 2010), it is necessary to follow up results presented form this study with systematic assessments for generalization.

Finally, while untaught skill emergence in the vocal of expressive identification showed to be an unexpected occurrence during this protocol, the design of this study impeded authors from assessing which variables in the receptive identification task, if any, contributed to increases in expressive identification. This study demonstrated that, for Alan, the inclusion of speech output technology boosted his acquisition of the untaught vocal task, as evidenced in increases in expressive accuracy in the SO versus NSO condition. Nonetheless, this design did not isolate other variables that may have been accountable for the more modest increases in expressive identification present in the NSO condition. These variables may include increased exposure to the vocal label, the complexity of the vocal label itself, and unplanned changes in reinforcement potency (e.g., in instructor's quality of praise) following an approximation, among others. Additional studies should further investigate these variables to assess which teaching strategies in a receptive identification task may facilitate the learning of vocal labels.

The present study aimed to replicate findings from Schlosser et al (1995) and other researchers with a novel population and evaluate whether speech output could boost a child's performance in terms of accuracy and efficiency in a receptive identification task. Data collected on Alan's performance provide strong preliminary evidence that speech output devices that utilize synthetic speech may be effective in assisting a child with ASD in acquiring the vocal label for corresponding visual images.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.) Washington, DC: Author.

Allen, J. (1992). Overview of text-to-speech systems. In S. Furui & M. M. Sondhi (Eds.), *Advances in speech signal processing* (pp. 741-790). New York: Marcel Dekker.

Billingsley, F. F., White, O. R., & Munson, R. (1980). Procedural reliability: A rationale and an example. Behavioral Assessment, 2, 229–241.

Blischak, D.M. (1999). Increases in natural speech production following experience with synthetic speech. *Journal of Special Education Technology, 14*(2), 44-53.

Blischak, D. M. (2003). Use of speech-generating devices: In support of natural speech. Augmentative and Alternative Communication, 19, 29–35.

Brady, N. (2000). Improved comprehension of object names following voice output communication aid use: Two case studies. *Alternative and Augmentative Communication, 16*(3), 197-204.

National Research Council (2001). *Educating children with. autism*. Washington, DC: National Academy Press.

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

Drager, K., Reichle, J., & Pinkoski, C. (2010). Synthesized speech output and children: A scoping review. *American Journal of Speech-Language Pathology, 19*, 259-273.

Dyches, T., Davis, A., Lucido, B., & Young, J. (2002). Generalization of skills using pictographic and voice output communication devices. *Augmentative and Alternative Communication, 18*(2), 124-141.

Durand, V.M. (1993). Functional communication training using assistive devices: Effects on challenging behavior and affect. *Augmentative and Alternative Communication, 9*, 168-176.

Durand, V.M. (1999). Functional communication training using assistive devices: Recruiting natural communities of reinforcement. *Journal of Applied Behavior Analysis, 32*, 247-267.

Delmolino, L., Hansford, A., Bamond, M., Fiske, K., & LaRue, R. (2013). The use of instructive feedback for teaching language skills to children with autism. *Research in Autism Spectrum Disorders, 7*, 648-661.

Gorenflo, C.W. & Gorenflo, D.W. (1991). The effects of information and augmentative communication technique on attitudes towards nonspeaking individuals. *Journal of Speech and Hearing Research, 34,* 19-34.

Koul, R., & Clapsaddle, K.C. (2006). Effects of repeated listening experiences on the perception of synthetic speech by individuals with mild-to-moderate intellectual disabilities. *Augmentative and Alternative Communication, 22*(2), 112-122.

Koul, R., & Hester, K. (2006). Effects of repeated listening experiences on the perception of synthetic speech by individuals with severe intellectual disabilities. *Journal of Speech, Language, and Hearing Research, 49,* 47-57.

Koul, R. & Schlosser, R.W. (2004). Effects of synthetic speech output in the learning of graphic symbols with varied iconicity. *Disability and Rehabilitation, 26*(21-22), 1278-1285.

Lilienfeld, M.&Alant, E. (2002). Attitudes of children toward an unfamiliar peer using an AAC device with and without voice output. *Augmentative and Alternative Communication, 18,* 91-101.

Mirenda, P. (2003). Toward functional augmentative and alternative communication for students with autism. *Language, Speech, and Hearing Services in Schools, 34,* 203-216.

Mirenda, P., & Iacono, T. (1988). Strategies for promoting augmentative and alternative communication in natural contexts for children with autism. *Focus on Autistic Behavior, 3,* 1-15.

McGregor, G., Young, J., Gerak, J., Thomas, B., & Vogelsberg, R.T. (1992). Increasing functional use of an assistive communication device by a student with severe disabilities. *Augmentative and Alternative Communication, 8,* 243-250.

Paul, R. (1997). Facilitating transitions in language development for children using AAC. *Augmentative and Alternative Communication, 13,* 141-148.

Parsons, C.L., & LaSorte, D. (1993). The effect of computers with synthesized speech and no speech on the spontaneous communication of children with autism. *Australian Journal of Human Communication, 21,* 12-31.

Romski, M.A.m & Sevcik, R.A. (1993). Language comprehension: Considerations for augmentative and alternative communication. *Augmentative and Alternative Communication, 9,* 281-285.

Rimland, B. (1964). *Infantile autism.* New York: Appleton-Century-Crofts.

Romski, M.A., & Sevcik, R. (1996). Breaking the speech barrier: Language development through augmented means. Baltimore: Paul H. Brookes.

Reichow, B., & Wolery, M. (2011). Comparison of progressive prompt delay with and without instructive feedback. *Journal of Applied Behavioral Analysis, 44*(2), 327–340.

Rotholz, D. & Berkowitz, S. (1989). Functionality of two modes of communication in the community by students with developmental disabilities: A comparison of signing and communication books. *Journal of the Association for Persons with Severe Handicaps, 14*, 227–233.

Schepis, M. M., & Reid, D.H. (1995). Effects of a voice output communication aid on interactions between support personnel and an individual with multiple disabilities. *Journal of Applied Behavior Analysis*, *28*(1), 73-77.

Schepis, M.M., Reid, D.H., & Behram, M.M. (1996). Acquisition and functional use of voice output communication by persons with profound multiple disabilities. *Behavior Modification, 20*(4), 451-468.

Schepis, M., Reid, D., Behrmann, M., & Sutton, K. (1998). Increasing communicative interactions of young children with autism using a voice output communication aid and naturalistic teaching. *Journal of Applied Behavior Analysis, 31,* 561-578.

Schlosser, R.W. (2003). Roles of speech output in augmentative and alternative communication: Narrative review. *Augmentative and Alternative Communication, 19*(1), 5-27.

Schlosser, R.W., Belfiore, P.J., Nigam, R., Blischak, D., & Hetzroni, O. (1995). The effects of speech output technology in the learning of graphic symbols. *Journal of Applied Behavior Analysis, 28*(4), 537-549.

Schlosser, R.W., & Blischak, D. (2001). Is there a role for speech output in interventions for persons with autism? A review. *Focus on Autism and Other Developmental Disabilities, 16*(3), 170-178.

Schlosser, R.W., Sigafoos, J., Luiselli, J.K., Amgermeier, K., Harasymowyz, U., Schooley. K., & Belfiore, P.J. (2007) Effects of synthetic speech output on requesting and natural speech production in children with autism: a preliminary study. *Research in Autism Spectrum Disorders, 1*, 139-163.

Schuster, J. W., Griffen, A. K., & Wolery, M. (1992). Comparisons of simultaneous prompting and constant time delay procedures in teaching sight words to elementary students with moderate mental retardation. *Journal of Behavioral Education, 2,* 305-325.

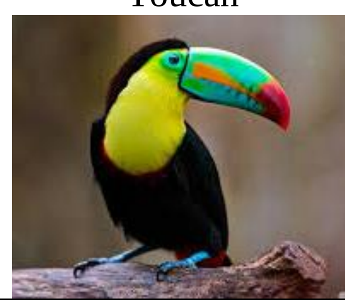
Touchette, P.E., & Howard, J.S. (1984) Errorless learning: Reinforcement contingencies and stimulus control transfer in delayed prompting. *Journal of Applied Behavior Analysis, 17*(2), 175-188.

Van der Meer, L.A. & Rispoli, M. (2010). Communication interventions involving speech-generating devices for children with autism: A review of the literature. *Developmental Neurorehabilitation*, *13*(4), 294–306.

Weitz, C., Dexter, M., & Moore, J. (1997). AAC and children with developmental disabilities. In S. Glennen & D. DeCoste (Eds.), *Handbook of augmentative and alternative communication* (pp. 395-431). San Diego: Singular.

Appendix A

Targets and Corresponding Images Divided by Tier

Tier 1	**Speech Output (SO) Condition**	**No Speech Output (NSO) Condition**
	Bongo 	Trombone 
	Helicopter 	Jeep 
	Hammer 	Toucan 

Tier 2	**SO Condition**	**NSO Condition**
	Aquarium 	Beets 
	Spinach 	Taxi 
	Violin 	Drums 

Tier 3	**SO Condition**	**NSO Condition**
	Wrench 	Clarinet 
	Coat Rack 	Oven 
	Peas 	Asparagus 

Appendix B

Data Collection and Interobserver Agreement Data Sheet.

Rater	Date	File Name	Cond		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	Extra Trials?
			SO NSO	Targets																
			SO NSO	Targets																
			SO NSO	Targets																
			SO NSO	Targets																
			SO NSO	Targets																
			SO NSO	Targets																

+	Correct Response
-	Incorrect Response
NR	No Response
r	social praise/ minor reinforcement
R	Token-earned Reinforcement
e	echoed label

Appendix C

Treatment Integrity Sheet - Baseline and Maintenance Phases

Rater:

Date of TI Check: Phase: File Name:

	Antecedent Strategies		Teaching Strategies				TI
Trial	Did teacher conduct a preference assessment, or did learner mand for a preferred object at the beginning of session/after earning reinforcement?	Did instructor establish attending behavior by calling learner's name/waiting for eye contact?	Did instructor present a clear/correct SD (e.g., "touch/find/where is the..." + target)?	Did the instructor provide a neutral response/praise for general behaviors after learner's response?	Did the instructor move the order of targets before delivering a new Sd?	Is Speech Output feature turned off in the pressed target?	
1	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
2	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
3	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
4	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
5	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
6	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
7	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
8	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
9	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
10	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
11	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
12	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
13	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
14	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
15	Y N N/A	Y N	Y N	NR Y N	Y N N/A	Y N	
						TOTAL TI	

Rate of Reinforcement:

Appendix D

Treatment Integrity Sheet - Training Phase

File Name: Date of TI Check: Rater:

Condition		Antecedent Strategies			Teaching Strategies				
SO NSO	Trial	Did teacher conduct a preference assessment, or did learner mand for a preferred object at the beginning of session/after earning reinforcement?	Did the instructor move the order of targets before delivering a new Sd?	Did instructor establish attending behavior by calling learner's name/waiting for eye contact?	Did instructor present a clear/correct SD (e.g., "touch/find/where is the..." + target)?	Did the instructor immediately follow the Sd with a model of the correct response?	When the instructor presses the mini-me, was a vocal model emitted?	Did the mini-me vocal model overlap with the instructor's Sd?	Did the instructor prevent responding until the model response/speech output was completed? (e.g., keeping materials away, hovering a hand over learner's hands)
	1	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	2	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	3	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	4	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	5	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	6	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	7	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	8	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	9	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	10	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	11	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	12	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	13	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	14	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N
	15	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N	Y N N/A	Y-worked Y- Didn't work N

	Consequences					
Trial	When the student pushed the mini-me as a response, was a vocal model emitted?	If student response was correct, did the instructor provide reinforcement?	If student response was incorrect, did the instructor provide feedback and model? ("this is ___")	How many vocal models are emmitted from the device in this trial?	Did the instructor move the order of targets before delivering a new Sd?	TI
1	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
2	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
3	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
4	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
5	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
6	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
7	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
8	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
9	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
10	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
11	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
12	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
13	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
14	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
15	Y N	Y - Verbal Y-Physical Y -	Y N	0 1 2 3	Y N N/A	
					TOTAL TI	

Appendix E

Treatment Integrity Sheet - Independent Phase

Phase: File Name: Date of TI Check: Rater:

Condition		Antecedent Strategies			Consequences					
SO NSO	Trial	Did teacher ask for, or did learner mand for a preferred object at the beginning of session/after earning reinforcement?	Did the instructor move the order of targets before delivering a new Sd?	Did instructor establish attending behavior by calling learner's name/waiting for eye contact?	Did instructor present a clear/correct SD (e.g., "touch/find/where is the..." + target)?	Did the instructor wait until the speech output was completed before delivering consequences (whether correct or incorrect)?	If response was correct: did the instructor provide a token as reinforcement?	If student response was incorrect, did the instructor provide feedback and model? ("no, this is ___")	If student did not respond after 7 seconds, did instructor provide instructive feedback and model?	TI
	1	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	2	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	3	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	4	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	5	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	6	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	7	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	8	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	9	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	10	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	11	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	12	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	13	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	14	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
	15	Y N N/A	Y N N/A	Y N	Y N	Y N	Y N N/A	Y N N/A	Y N N/A	
									TOTAL TI	