

Comparison of Flexible Prompt Fading to Error Correction for Children with Autism Spectrum Disorder

Justin B. Leaf · Ronald Leaf · Mitchell Taubman ·
John McEachin · Lara Delmolino

Published online: 12 September 2013
© Springer Science+Business Media New York 2013

Abstract This study compared flexible prompt fading to an error correction procedure involving feedback and remedial trials for teaching four children with Autism Spectrum Disorder. Using a parallel treatment design nested into a multiple probe design, researchers taught each participant how to expressively label six pictures of Muppet characters with the flexible prompt fading procedure and six pictures of Muppet characters with the error correction procedure. The researchers evaluated the effectiveness, maintenance, efficiency, and acquisition during teaching for each participant across the two teaching conditions. Results indicated that both teaching procedures were effective, resulted in high rates of maintenance, and that participants responded correctly during the majority of teaching trials. However, flexible prompt fading was more efficient in terms of total number of trials and sessions, as well as total amount of time for participants to learn all targeted skills.

Keywords Autism · Discrete trial teaching · Error correction · Flexible prompt fading · Prompting

Discrete trial teaching (DTT) is commonly implemented to help teach students diagnosed with an autism spectrum disorder (ASD) (Lovaas 1987; Smith 2001). The three main components of DTT are: (a) a discriminative stimulus (S^D) from the

Author Note We wish to thank Jeremy A. Leaf, Christine Miline, Amy Lentel, Marlene Brown, and Amanda Kwok for their help running sessions throughout the study. We wish to thank Shelli Imfeld, Julie Stiglich, and Cliff Anderson for their help throughout the project. Finally, we wish to thank Misty L. Oppenheim-Leaf for her insight on previous versions of this manuscript.

J. B. Leaf · R. Leaf · M. Taubman · J. McEachin
Autism Partnership Foundation, Seal Beach, CA, USA

L. Delmolino
Rutgers University, New Brunswick, NJ, USA

J. B. Leaf (✉)
200 Marina Drive, Seal Beach, CA 90740, USA
e-mail: Jblautpar@aol.com

teacher; (b) a response by the student; and (c) a consequence provided by the teacher. Since students with ASD often need assistance from the teacher in order to display the correct response, an optional fourth step of DTT is prompting. Prompts can take many forms (known as prompt types). Prompt types can include: pointing to the correct response (e.g., Leaf et al. 2010), verbally stating the correct response (e.g., Leaf et al. 2011a, b), modeling the correct behavior (e.g., Bozkurt and Gursel 2005), reducing the number of choices (e.g., Soluaga et al. 2008), within-stimulus prompts (e.g., Schreibman 1975), or physically guiding the learner to the correct response or to engage in the correct behavior (e.g., Leaf et al. 2010).

When prompts are utilized, it is intended for the control of the response to be systematically transferred from the prompt to the intended S^D . Researchers have created prompting systems to help ensure that teachers provide prompts correctly, fade prompts appropriately, prevent unintended prompts, and avoid the student becoming dependent on teacher prompts. Today, there are several different prompting systems that have been evaluated in the literature and are implemented clinically to teach students diagnosed with ASD. These prompting systems include: time delay (e.g., Charlop and Trashowech 1991; Morse and Schuster 2000), least-to-most prompting (e.g., Tarbox et al. 2007), and most-to-least prompting (e.g., Bloh 2008).

One prompting system that has been clinically implemented with numerous students with ASD, but has limited empirical evidence as to its effectiveness, is flexible prompt fading (FPF). Flexible prompt fading was first described by Lovaas and colleagues during investigations at the UCLA Young Autism Project (Lovaas 1987) and has more recently been described by Leaf and McEachin (1999). Flexible prompt fading is a prompting technique that relies on the teacher using his or her clinical judgment (that is, making in-the-moment interventional decisions based on defined parameters) to decide whether or not to prompt a student and what type of prompt to implement. Thus, FPF is similar to graduated guidance (e.g., MacDuff et al. 1993; Wolery and Gast 1984), as it allows clinicians the freedom to prompt based on general guidelines rather than specific rules.

Although clinicians prompt student responses based on in-the-moment decisions, there are several important guidelines that the clinician follows when implementing FPF. In general, the clinician aims for the student to maintain a high level of success (e.g., 80 % correct with or without a prompt). Second, the clinician should provide a prompt if the student has had a recent history of errors on the task. If the student has had a long history with the task or has had a recent history of responding correctly, the clinician may elect to reduce the level of assistance or not provide a prompt at all. Ultimately, the clinician must determine whether the student is likely to make a correct response on the next trial, based on the criteria above, and prompt accordingly. If the student is likely to respond correctly, the teacher should provide a less intrusive prompt or not prompt; if the student is likely to respond incorrectly, the teacher should provide a prompt. Further guidelines of FPF have been described by Leaf and McEachin (1999).

The first study to evaluate flexible prompt fading was conducted by Soluaga et al. (2008). This study compared FPF to time delay for teaching various academic tasks to five children diagnosed with ASD. The FPF procedure consisted of the teacher implementing five different prompt types (i.e., physical, pointing, modeling, positioning prompts, and field reduction prompts) in a one-to-one instructional format.

During the implementation of the time delay procedure, however, only controlling prompts (i.e., the least intrusive prompt type that guarantees a correct response by the learner) were implemented. A modified parallel treatments design was utilized to compare the effectiveness of the two prompting procedures. Results of the study indicated that both prompting procedures were effective and there was mixed results in terms of efficiency.

Although prompting has been demonstrated to be an effective component of DTT, some clinicians may elect not to implement DTT with the provision of antecedent prompts. When a teacher does not provide an antecedent prompt (e.g., prompting prior to the student's behavior) they are relying on either pure trial-and-error learning or on providing some type of explicit error correction procedure (Rodgers and Iwata 1991). Since trial-and-error procedures can have undesirable side effects, error correction procedures (EC) are more widely implemented. In EC, teachers provide reinforcement for correct responses and corrective feedback (e.g., “Nope, that is not it.”) followed by modeling the correct response (e.g., “This is an apple.”) for incorrect responses. Providing corrective feedback only, without modeling the correct response, for incorrect responses is an example of consequence-based EC that does not directly assist the student in identifying the correct response (i.e., trial-and-error). Providing the correct model in addition to the corrective feedback may increase the rate of learning (Smith et al. 2006). Finally, the teacher may implement another immediate unprompted opportunity for the student to display the appropriate behavior (i.e., remedial trial) after the model has been provided.

Researchers have found that error correction procedures can be effective in teaching a wide variety of skills including: verb usage (Schumaker and Sherman 1970), matching to sample tasks (Rodgers and Iwata 1991), and expressive labeling of sight words (Worsdell et al. 2005). In 2005, Worsdell and colleagues evaluated the effects of error correction procedures in teaching 11 adults with development disabilities to improve their ability to recognize sight words. Worsdell et al. demonstrated that error correction procedures utilizing a remedial trial after every incorrect response were effective in increasing sight word recognition.

Smith et al. (2006) compared three different teaching conditions for teaching matching words to pictures for six participants diagnosed with ASD. In the first condition, error statement, the teacher said “no” anytime the participants made an incorrect response (similar to corrective feedback stated above). In the second condition, modeling, the teacher stated the correct response anytime the participants made an incorrect response. In the third condition, the teacher provided no feedback (pure extinction) any time the participants made an incorrect response. Results of the study were idiosyncratic across participants in regards to acquisition rate. For four of the six subjects, EC was superior to no feedback. The other two participants were fast learners and performed as well in the no-feedback condition as they did in the EC condition. Of the four who performed better with EC, two did equally well with both EC methods, while one made fewer errors with the error statement and one made fewer errors with modeling of the correct response.

While the research to date has shown that a number of error correction procedures are effective in teaching new skills, many professionals still warn against teaching procedures that allow students to make errors (e.g., Gast 2011). Research has shown that under some circumstances errors can lead to more errors, students may display

aberrant behaviors after making an error, and that error correction procedures may not be as effective as other prompting procedures (e.g., Ferster and DeMeyer 1962). Therefore, more direct comparison studies are warranted to provide further evidence about which procedures are most effective and efficient for teaching acquisition of new skills to children with ASD. Additionally, clinicians working with individuals diagnosed with ASD should implement the most effective and efficient procedures. Thus, the purpose of this study was to compare a consequence based procedure (i.e., an error correction procedure), which did not attempt to minimize participant errors, to an antecedent prompt procedure (i.e., a flexible prompt fading procedure), which attempted to minimize errors through the use of antecedent prompts. In doing so, we compared the effectiveness, maintenance, and efficiency of the two procedures in teaching expressive labeling to four high-functioning children diagnosed with ASD.

Method

Participants

Participants all had a formal diagnosis of autistic disorder from an outside agency, ranged in age from 4- to 6-years-old, and had an IQ score ranging from 86 to 128. Three of the four participants had a history of educational intervention that used a flexible prompt fading procedure. None of the participants had a history of error correction.

Rob was a 5-year-old boy independently diagnosed with autistic disorder. Rob had a Wechsler Preschool and Primary Scale of Intelligence-Third Edition (WPPSI-III) FSIQ score of 128, a Vineland-II Adaptive Behavior Scales Survey Interview Form (VABS-II) adaptive behavior score of 94, a Gilliam Autism Rating Scale (GARS-II) autism quotient of 98 (probability of autism very likely), and a PPVT-4 standard score of 123. Rob had received a mean of 20 h of behavioral treatment per week over the prior 18 months and was placed in a special education preschool classroom with supports.

Jimmy was a 4-year-old boy independently diagnosed with autistic disorder. Jimmy had a WPPSI-III FSIQ score of 86, a VABS-II adaptive behavior score of 81, a GARS-II autism quotient of 98 (probability of autism very likely), and a PPVT-4 standard score of 100. Jimmy had received a mean of 39.5 h of behavioral treatment per week over the prior 24 months and was placed in a special education classroom without supports.

Billy was a 5-year-old boy independently diagnosed with autistic disorder. Billy had a WPPSI-III FSIQ score of 99, a VABS-II adaptive behavior score of 88, a GARS-II autism quotient of 89 (probability of autism very likely), and a PPVT-4 standard score of 117. Billy had received a mean of 13 h of behavioral treatment per week over the prior 21 months and was placed in a general education preschool classroom without supports.

Kenny was a 6-year-old boy independently diagnosed with autistic disorder. Kenny had a Stanford Binet-Fifth Edition FSIQ score of 88, an ADOS (Module 3) score meeting the Autism cut-off (communication and social interaction combined score of 17), and a PLS-4 standard score of 88 (percentile of 21 and age equivalent of 5 years 11 months). Kenny had received a mean of 25 h of behavioral treatment per

week for a period of 24 months and was placed in an integrated preschool classroom with supports. He is the only participant without prior exposure to FPF.

Setting and Researchers

This study took place in two different settings. The setting for three of the participants (i.e., Rob, Jimmy, and Billy) was a small research room in a private behavior intervention agency's Southern California office. The research room measured approximately 2.7 m by 2.7 m and contained a table, cabinets, chairs, couch, closets for research materials, and a desk. At this research site, there were three researchers who conducted the study on a daily basis. Two of the researchers had a Bachelor's degree in psychology and one had a Master's degree in education. Each researcher had received an initial intensive training lasting at least 2 months. The training consisted of both didactic and hands on training on various topics, such as: applied behavior analysis, autism, reinforcement, prompting, discrete trial teaching, error correction, and teaching interactions. After this initial training, each researcher had over 1-year direct experience working with individuals diagnosed with autism and implementing the procedures utilized in this study.

The second setting, for Kenny only, was a small research room in a New Jersey university that provides behavioral intervention for children and adults diagnosed with ASD. The research room contained a table, chairs, and file cabinets. On a few occasions, due to scheduling conflicts, sessions took place in an office or an unused classroom, both of which were familiar to the student. At this research site, there were two primary researchers who conducted the study on a daily basis. One of the researchers held a doctorate degree and one had a master's degree pending at the time of the study. Both researchers had over 20 years of experience in the field and had received intensive training in applied behavior analysis, autism, reinforcement, prompting, discrete trial teaching, and error correction procedures. Both of the researchers also had extensive experience (over 10 years) utilizing the procedures in this study.

Skills Taught

Each participant was taught to expressively label the names of 12 pictures of Muppet® characters. The selection of these skills were made based upon each participant's supervisors recommendation as they were teaching each participant pop culture knowledge, a skill that all participants needed. Additionally, these skills were not being targeted in each participant's current clinical intervention, so skills would not be inadvertently taught. Character names were taught in pairs, and the stimulus pairs were randomly assigned to one of the two conditions prior to baseline. Table 1 shows the item pairs that were taught to each participant with each procedure.

General Procedure

The researchers conducted research sessions 3 to 5 days per week; only one research session occurred per day. The length of the sessions ranged from 5 to 25 min dependent upon the type of session (e.g., probe session only or probe session plus

Table 1 Targeted skills

Participant	First stimulus pair FPF	First stimulus pair EC	Second stimulus pair FPF	Second stimulus pair EC	Third stimulus pair FPF	Third stimulus pair EC
Jimmy	Scooter & Honeydew	Beaker & Janice	Sweetums & Camilla	Rizzo & Sam	Floyd & Lew	Dr. Teeth & Animal
Rob	Beaker & Janice	Scooter & Honeydew	Lew & Sweetums	Rizzo & Sam	Dr. Teeth & Zoot	Camilla & Floyd
Billy	Beaker & Janice	Scooter & Honeydew	Rizzo & Pepe	Sweetums & Camailla	Dr. Teeth & Zoot	Floyd & Lew
Kenny	Fozzie & Woldorf	Sweetums & Camilla	Zoot & Lew	Rowlf & Floyd	Dr. Teeth & Statler	Rizzo & Sam

teaching sessions) and participant responding (e.g., more reinforcement breaks for correct responding). During some sessions, the participant only received probe trials (full probe sessions) to assess baseline levels for skills not yet taught and to assess maintenance levels for skills previously taught (see below). During the majority of research sessions, the researchers implemented daily probe trials to test for acquisition, a short break (approximately 2 min), one of the teaching conditions (i.e., FPF or EC), another short break (approximately 2 min), and then the second teaching condition (e.g., the procedure that was not implemented first). The order of FPF and EC were randomly determined prior to each research session.

Each trial (probe trials and teaching trials) began by the researcher holding up one of the cards displaying a Muppet character in view of the participant. Next, the researcher gave an instruction to the participant to provide the name of the Muppet character (e.g., “What is his or her name?”), and allowed approximately 5 s for the participant to respond. During probe trials, no prompts, reinforcement, or feedback was provided to the participant; the researcher provided neutral feedback (e.g., “Thanks” or “Thank You”) regardless of the participant’s response (i.e., correct or incorrect response). During teaching trials, however, the researcher provided prompts, reinforcement, and feedback dependent upon the teaching condition being implemented (see below).

Prior to beginning intervention, potential tangible reinforcers (e.g., toys or edibles) were selected, which were used during full probe sessions, daily probe sessions, FPF teaching sessions, and EC teaching sessions. The researchers selected tangible reinforcers by observing the participant, asking the participant what he wanted to work for, or interviewing the participant’s teachers and/or parents. The researchers selected approximately 5 different tangible reinforcers for each participant. The reinforcers were held constant across both teaching conditions throughout the study.

Full Probe Sessions

The researchers conducted full probe sessions prior to the teaching of any new stimulus items to determine current baseline performance. Additionally, after the participant met mastery criterion (i.e., 100 % correct on all daily probe trials for 3

consecutive daily probes) on at least one stimulus pair, researchers administered a full probe session on all stimulus pairs to evaluate whether correct responding on previously taught pairs was maintained. The researchers evaluated all stimulus items four times each during full probe sessions and randomly determined the order for presentation during these sessions; thus, each full probe session consisted of 48 full probe trials. No reinforcement was provided to the participants contingent upon correct responding during full probe sessions. The researchers did provide reinforcement to participants on a fixed ratio schedule (FR3 or FR4) contingent upon the participant displaying appropriate behaviors (e.g., sitting in his or her chair and not engaging in any aberrant behaviors); the reinforcer provided was randomly selected.

Daily Probe Sessions

The researchers conducted daily probes prior to each teaching session to evaluate whether participants were learning to correctly label the Muppet characters that were currently being taught to them. Daily probe trials were conducted in the same manner as full probe trials. The daily probe sessions consisted of 16 randomized probe trials; four probe trials were conducted for each target skill currently being taught (2 skills with EC and 2 skills with FPF). Mastery criterion was set at 100 % correct responding on all probe trials for a stimulus pair (i.e., 8 probe trials) across three consecutive daily probes. The researchers provided reinforcement to the participants for displaying appropriate behavior (e.g., sitting correctly) on an FR-4 schedule (similar to the reinforcement provided during full probe sessions). Once a participant met mastery criterion for a stimulus pair, teaching on that stimulus pair stopped; daily probes, however, were continued until at least three more daily probe sessions were completed or the second stimulus pair reached mastery criterion. Following daily probes, the researchers provided the participant with a brief 1 to 2-min break prior to beginning the first teaching session. During the first session in which new stimulus pairs were being taught no daily probe was implemented.

Teaching Session

Flexible Prompt Fading (FPF) A total of 20 teaching trials per session were implemented in this condition. The FPF condition started with the researcher placing a color mat (e.g., yellow mat) in front of the participant; the color mat indicated that the FPF condition was going to be implemented. In the FPF condition, a trial started with the researcher holding up a picture of the Muppet character in sight of the participant. Next, the researcher provided a discriminative stimulus and gave the participant 5 s to respond to the instruction. If the participant labeled the character correctly (prompted or unprompted) the researcher provided the participant with praise and brief access (approximately 5 s) to the reinforcer (described above). The researcher randomly selected one of the five reinforcers and provided it to the participant. Thus, the inter-trial interval was approximately 5 s following correct responses. If the participant labeled the character incorrectly, the researcher said “Nope, that’s not it” and moved to the next trial; the following trial could be a remedial trial or the researcher had the autonomy to move to the next predetermined trial. Thus, the inter-trial interval was approximately 3 s following incorrect responses.

During the FPF condition, the researchers had the flexibility to provide antecedent prompts to help ensure that the participant maintained a high level of correct responding. Although the use of prompts during FPF is based primarily upon researcher judgment, those decisions were governed by several guidelines that the researchers were instructed to follow.

Most importantly, the researchers aimed to have the participant respond correctly (prompted or unprompted) on at least 80 % of trials. The researchers were instructed to assess prior to each trial whether or not the participant was likely to respond correctly. If the researcher determined that the participant was likely to respond correctly without a prompt, then the researcher did not provide a prompt to the participant. If the researcher determined that the participant was likely to respond incorrectly, then the researcher provided a prompt to the participant. In order to make this assessment, the researcher first looked at the previous responses of the participant. If the participant was responding correctly without prompts on previous trials, or if the researcher had prompted the participant on several previous trials, then the researcher could either reduce the level of assistance or not prompt at all. If the participant was responding incorrectly with a less assistive prompt then the researcher could provide a more intrusive prompt. Additionally, if the participant had many previous sessions with the target, the researcher may elect not to provide a prompt. Finally, the researcher assessed the participant's current behaviors. If the participant's tolerance for frustration was low, the researcher was more likely to provide a prompt.

Second, the researcher had the flexibility to implement multiple prompt types (e.g., verbal prompt, partial verbal prompt, model prompt) at his or her discretion. The researchers were guided to implement any prompt type that he or she thought would result in the participant responding correctly on any given trial. Thus, unlike other prompting systems (e.g., most-to-least prompting) where the researcher has to provide a given prompt at a given point the researcher had the discretion to provide any prompt type at any point. Furthermore, the researchers were instructed to fade prompts as quickly as possible to transfer stimulus control from the prompt to the instruction alone.

A third guideline was that the researcher had to provide all prompts directly after the instruction and prior to the participant engaging in a correct or incorrect response. This was different than the error correction procedure where the researcher provided instructional feedback following a participant's incorrect response (see below).

Error Correction (EC) A total of 20 teaching trials per research session were implemented in this condition. The EC condition started with the researcher placing a color mat (e.g., red mat) in front of the participant; the color mat indicated that the EC condition was going to be implemented. In the EC condition a trial started with the researcher holding up a picture of the Muppet character in sight of the participant. Next, the researcher provided a discriminative stimulus (e.g., "What is his or her name?") and gave the participant 5 s to respond to the instruction. If the participant labeled the character correctly, the researcher provided the participant with praise and provided one of the same reinforcers used in the FPF condition for approximately 5 s. After 5 s the researcher asked the participant to hand him or her back the toy and implemented the next planned teaching trial. Thus, the inter trial interval was approximately 5 s following correct responses.

If the participant incorrectly labeled the character or did not respond to the instruction within the 5 s, the researcher said, “No, that’s not it” followed by stating the correct name (e.g., “This is [character’s name].”) of the character. The participant was not required to imitate the modeled response. Instead, the researcher provided one remedial trial, providing the participant with the opportunity to demonstrate the correct response. The remedial trial started with the researcher holding up a picture of the Muppet character in sight of the participant. Next, the researcher provided a discriminative stimulus and gave the participant 5 s to respond to the instruction. If the participant labeled the character correctly, the researcher provided the participant with praise, but no tangible reinforcement. If the participant did not respond correctly or did not respond to the instruction within the 5 s, the researcher said, “No, that is not it” followed by stating the correct name of the character. Regardless of the outcome of the remedial trial, the researcher moved on to the next planned trial.

Dependent Variable and Data Collection

The primary measure was participants’ skill acquisition as measured by daily probe trials. The researchers measured how many stimulus pairs the participants mastered across the two teaching conditions. Mastery criterion was set as the participant responding 100 % correct for targets of a stimulus pair for three consecutive daily probe sessions. During all probe trials, the researcher recorded the response of the participant. A correct response was recorded if the participant correctly named the picture of the Muppet character within 5 s of the researcher’s instruction. An incorrect response was recorded if the participant incorrectly named the picture of the Muppet character within 5 s of the researcher’s instruction. A no-response was recorded if the participant did not give any response within 5 s of the researcher’s instruction.

The second measure was how well the participants maintained skills taught to them, which was assessed on full probe trials. During each full probe session the researchers recorded participant responding during each probe trial. As described above, the participants could respond correctly, incorrectly, or have no response.

The third measure was the relative efficiency of the two interventions. We measured the total number of teaching sessions, total number of trials, and total amount of teaching time required for each participant to master all of his targets across the two teaching conditions. Each research session consisted of 20 total teaching trials for FPF and 20 total teaching trials for EC. A teaching trial was defined as anytime the researcher presented an instruction for the participant to respond, regardless of whether the trial was prompted or not or was preceded by an error. Thus, each remedial trial in the EC condition was counted as a separate trial (i.e., 1 out of the 20 trials). A remedial trial in the EC condition was defined as anytime the participant made an incorrect response and the teacher re-presented the same targeted behavior on the next trial. Therefore, if a participant was incorrect on the first opportunity and was incorrect on a remedial trial, this was scored as two incorrect responses. Remedial trials in the FPF condition also counted as separate trials. A remedial trial in the FPF condition could occur at three times: (1) if the participant responded incorrectly and the teacher provided a follow-up trial of the same targeted response (similar to the EC condition); (2) if the participant responded correctly independently and the teacher

decided to provide another opportunity to the learner to respond to the same target; and (3) if the participant responded correctly with the provision of a prompt and the teacher elected for the student to have an opportunity to respond correctly but without a prompt being provided.

The final measure captured the percentage of participant responses during teaching trials across the two conditions. A total of five response types were evaluated, which included: (1) overall correct trials without prompts (first opportunity and remedial trials); (2) correct trials without prompts on the first opportunity; (3) incorrect/ no response trials without prompts (first opportunity and remedial trials); (4) prompted correct trials; and (5) prompted incorrect/ no response trials.

Correct and incorrect/no response trials had the same operational definition as responses during probe trials. Prompted correct trials were scored if the researcher provided an antecedent prompt (e.g., verbally stating the correct response) and the participant correctly labeled the picture. Prompted incorrect trials were scored if the researcher provided an antecedent prompt (e.g., verbally stating the correct response) and the participant incorrectly labeled the picture. Prompted responses were never counted as correct or incorrect. Remedial trials were scored based upon the learners response.

Experimental Design

A parallel treatment design (Gast and Wolery 1988) nested in a multiple probe design across skill sets and replicated across participants was used to evaluate the effectiveness of the two prompting procedures. It is critical that when implementing a parallel treatment design, that the order of the two procedures are randomly determined ahead of time, which was done throughout this study. With a parallel treatment design, experimental control is established when one of the prompting procedures results in more rapid skill acquisition than the other prompting procedure. Since experimental control may be undermined if both procedures result in equal rates of acquisition, the additional use of the multiple probe design helps to ensure experimental control. With the multiple probe design the researcher implements the independent variable (i.e., the prompting systems) on one of the dependent variables (i.e., one of the stimulus sets) and does not intervene on the other independent variables until an increasing trend is shown. Thus, experimental control is established if learning occurs when, and only when, the intervention is implemented.

Interobserver Agreement

The researcher scored the participants responses during every session. A second observer (i.e., research assistant) simultaneously and independently recorded participant responses during 52.1 % (range, 41.6 % to 75 % across participants) of the full probe sessions, 48 % (range, 30 % to 71.4 % across participants) of the daily probe sessions, 58.1 % (range, 33 % to 100 % across participants) of the FPF sessions, and 62.5 % (range, 36 % to 100 % across participants) of the EC sessions; interobserver reliability was scored both in-vivo and by watching videotapes of the research sessions. Interobserver agreement was calculated by totaling the number of agreements (i.e., trials in which both observers scored the same response) divided by the

number of agreements plus disagreements (i.e., trials in which the two observers scored a different participant response) and converting this ratio to a percentage. Percentage agreement across all participant responses was 99.5 % (range, 95.8 % to 100 % per session) for full probe trials, 98.9 % (range, 87.5 % to 100 % per session) for daily probe trials, 99.6 % (range, 95 % to 100 % per session) for FPF teaching trials, and 98.5 % (range, 90 % to 100 % per session) for EC teaching trials, summed across all four participants.

Treatment Fidelity

The researchers measured correct instructor behaviors during full probe sessions, daily probe sessions, flexible prompt fading trials, and error correction trials (contact author for treatment fidelity checklists). During full and daily probe trials, correct instructor behaviors included: (a) holding up the picture in the participant's view; (b) delivering an instruction for the participant to name the Muppet character; (c) the researcher allowing approximately 5 s (e.g., plus or minus 1 s) for the participant to respond; and (d) providing the participant with neutral praise (e.g., "Thank You" or "Thanks") regardless of the participant's response.

During flexible prompt fading trials, correct instructor behaviors included: (a) holding up the picture in the participant's view; (b) delivering an instruction for the participant to name the Muppet character; (c) the researcher allowing approximately 5 s (e.g., plus or minus 1 s) for the participant to respond; (d) the researcher providing reinforcement (i.e., social praise and a toy) only if the participant responded correctly; and (e) the researcher providing corrective feedback (i.e., "That's not it") only if the participant responded incorrectly. In addition to these correct teacher behaviors, we also analyzed whether or not the researcher(s) maintained a participant's correct level of responding (correct or correct after the provision of a prompt) at 80 % or above (e.g., the most important guideline of FPF).

Correct instructor behaviors measured for error correction were: (a) holding up the picture in the participant's view; (b) delivering an instruction for the participant to name the Muppet character; (c) the researcher allowing approximately 5 s (e.g., plus or minus 1 s) for the participant to respond; (d) the researcher providing reinforcement (i.e., social praise and a toy) only if the participant responded correctly; (e) the researcher providing corrective feedback (i.e., "That's not it") only if the participant responded incorrectly; (f) the teacher providing informative feedback only after an incorrect response; and (g) the researcher providing a remedial trial if the participant responded incorrectly on the first opportunity to respond independently.

To assess treatment fidelity, an independent observer (e.g., research assistant) recorded the researchers behaviors during 37.5 % (range, 33.3 % to 41.6 % across participants) of full probe sessions, 34 % (range, 30 % to 35.7 % across participants) of daily probe sessions, 53.4 % (range, 33 % to 100 % across participants) of the flexible prompt fading sessions, and 55.1 % (range, 35.7 % to 100 % across participants) of error correction sessions. The observer reported that the researcher engaged in correct instructor behaviors on 99.2 % (range, 94 % to 100 %, across sessions) of full probe trials; 99.6 % (range, 94 % to 100 % across sessions) of daily probe trials; 98 % (range, 80 % to 100 % across sessions) of flexible prompt fading trials; and 97.4 % (range, 80 % to 100 % across sessions) of error

correction trials. Additionally, the researchers maintained participant responding above 80 % during 100 % of flexible prompt fading sessions in which treatment fidelity was taken.

Results

Skill Acquisition, Mastery Criterion, and Maintenance

The researchers taught Jimmy three stimulus pairs using FPF and three stimulus pairs using EC (see Fig. 1). Jimmy reached mastery criterion for all of the stimulus pairs taught using FPF and all of the stimulus pairs taught using EC. The assessment of maintenance was conducted during full probe sessions after Jimmy reached mastery criterion. The first assessment of maintenance for skills taught with FPF was 7 days (set 1), 3 days (set 2), and 5 days (set 3) after mastery criterion was met. The first assessment of maintenance for skills taught with EC was 5 days (set 1), 4 days (set 2), and 3 days (set 3) after mastery criterion was met. The final assessment of maintenance for skills taught with FPF was 58 days (set 1), 38 days (set 2), and 8 days (set 3) after mastery criterion was met. The final assessment of maintenance for skills taught with EC was 56 days (set 1), 37 days (set 2), and 6 days (set 3) after mastery criterion was met. During the assessment of maintenance, Jimmy's mean correct responding on the stimulus pairs taught with FPF and EC was 90.2 % (range, 75–100 %) and 92.1 % (range, 87.5–100 %), respectively.

The researchers taught Rob three stimulus pairs using FPF and three stimulus pairs using EC (see Fig. 2). Rob reached mastery criterion for all of the stimulus pairs taught using FPF and all of the stimulus pairs taught using EC. The assessment of maintenance was conducted during full probe sessions after Rob reached mastery criterion. The first assessment of maintenance for skills taught with FPF and EC was 1 day (set 1), 1 day (set 2), and 4 days (set 3) after Rob reached mastery criterion. The final assessment of maintenance for skills taught with FPF and EC was 53 days (set 1), 31 days (set 2), and 7 days (set 3) after Rob reached mastery criterion. During the assessment of maintenance, Rob's mean correct responding for all stimulus pairs taught with FPF and EC was 100 %.

The investigators taught Billy three stimulus pairs using FPF and three stimulus pairs using EC (see Fig. 3). Billy reached mastery criterion for all of the stimulus pairs taught using FPF and all of the stimulus pairs taught using EC. The assessment of maintenance was conducted during full probe sessions after Billy reached mastery criterion. The first assessment of maintenance for skills taught with FPF was 1 day (set 1), 7 days (set 2), and 4 days (set 3) after Billy reached mastery criterion. The first assessment of maintenance for skills taught with EC was 1 day (set 1), 6 days (set 2), and 4 days (set 3) after Billy reached mastery criterion. The final assessment of maintenance for skills taught with FPF was 51 days (set 1), 29 days (set 2), and 13 days (set 3) after Billy reached mastery criterion. The final assessment of maintenance for skills taught with EC was 51 days (set 1), 28 days (set 2), and 13 days (set 3) after Billy reached mastery criterion. During the assessment of maintenance, Billy's mean correct responding on stimulus pairs taught with both FPF and EC was 98.6 % (range, 87.5 % to 100 %).

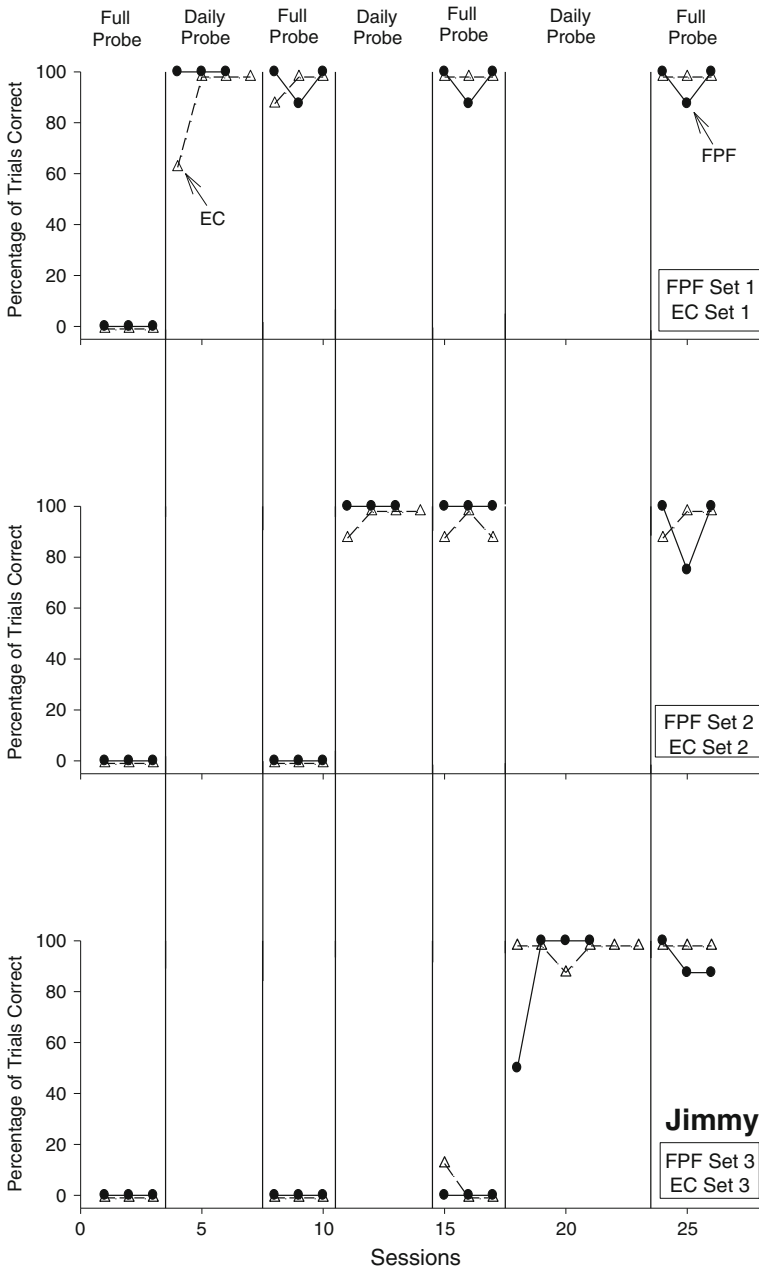


Fig. 1 Jimmy probe data

The investigators taught Kenny three stimulus pairs using FPF and three stimulus pairs using EC (see Fig. 4). Kenny reached mastery criterion for all of the stimulus pairs taught using FPF and all of the stimulus pairs taught using EC. The assessment of maintenance was conducted during full probe sessions after Kenny reached mastery criterion. The first assessment of maintenance for skills taught with FPF was 11 days (set 1) and 1 day (set 2

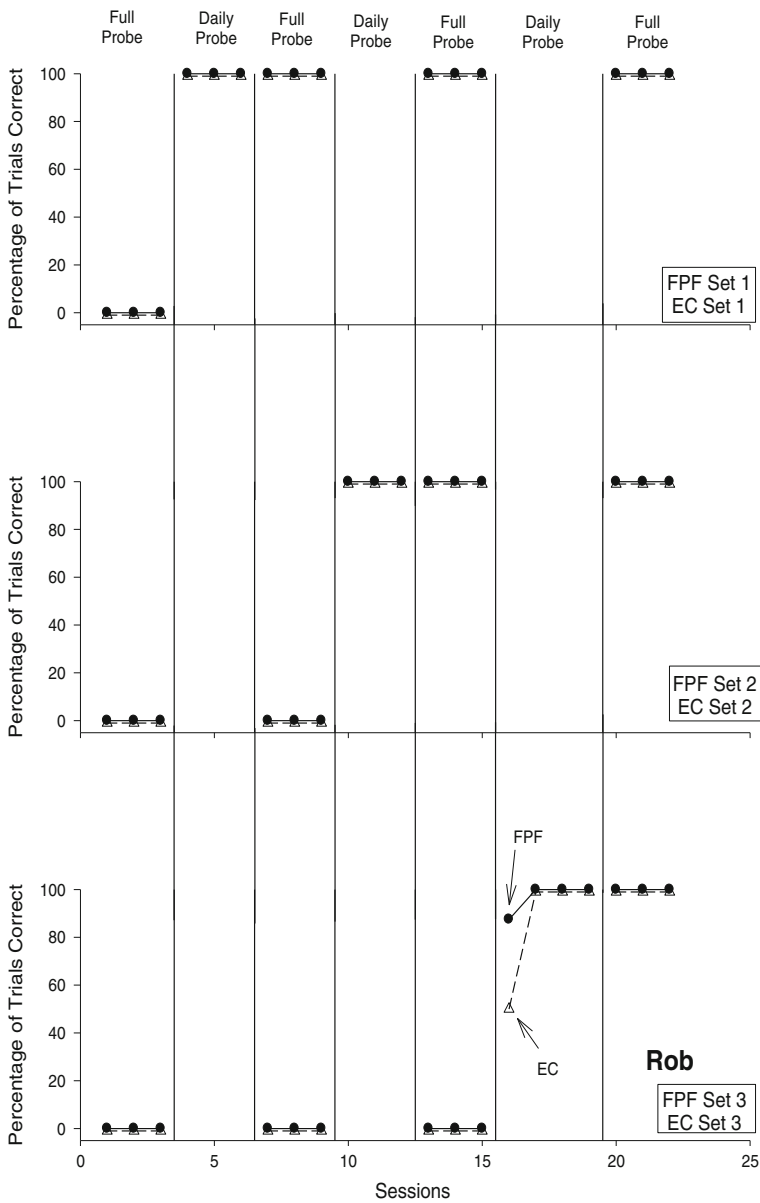


Fig. 2 Rob probe data

and set 3) after Kenny reached mastery criterion. The first assessment of maintenance for skills taught with EC was 1 day (set 1, set 2, and set 3) after Kenny reached mastery criterion. The final assessment of maintenance for skills taught with FPF was 39 days (set 1), 14 days (set 2), and 5 days (set 3) after Kenny reached mastery criterion. The final assessment of maintenance for skills taught with EC was 31 days (set 1), 14 days (set 2), and 5 days (set 3) after Kenny reached mastery criterion. During the assessment of maintenance, Kenny's

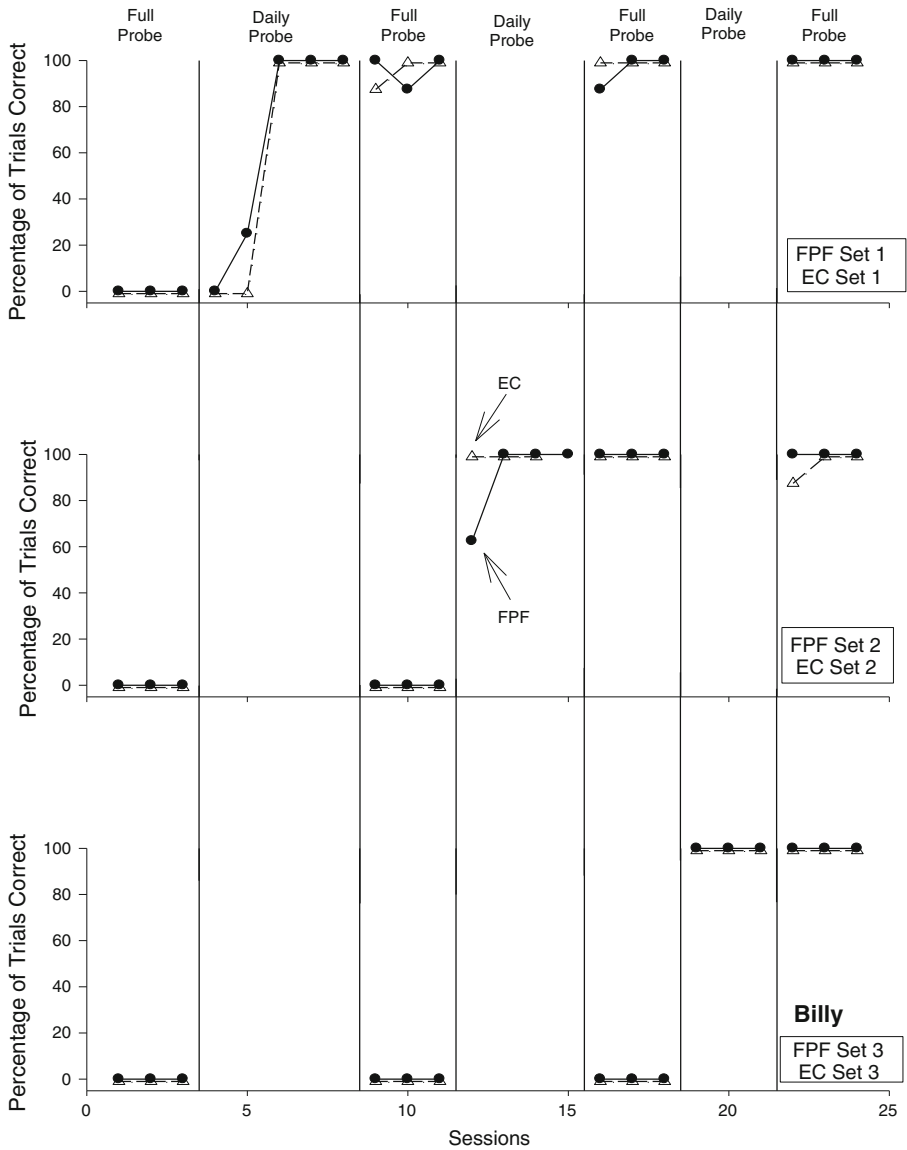


Fig. 3 Billy probe data

mean correct responding on stimulus pairs taught with FPF and EC was 92.8 % (range, 50 % to 100 %) and 97.9 % (range, 87.5 % to 100 %), respectively.

Efficiency

The researchers measured the total amount of sessions, total amount of teaching trials, and total amount of time it took participants to reach mastery criterion across the two teaching methodologies (see Table 2). Data summarized across all participants indicated that targets taught with FPF required fewer sessions, trials, and total amount

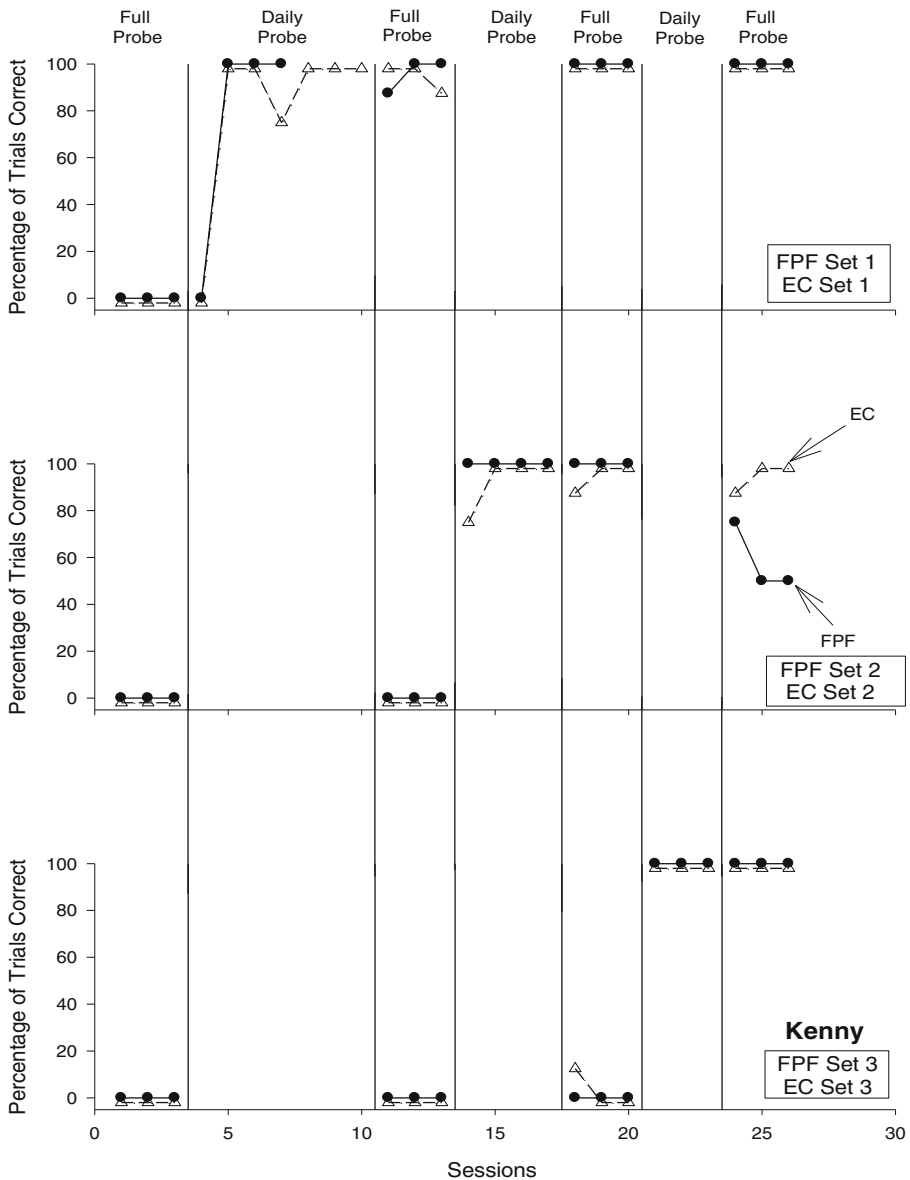


Fig. 4 Kenny probe data

of teaching time to reach mastery criterion; however, results were idiosyncratic among the participants. Billy learned skills taught with EC in fewer sessions, trials, and total amount of teaching time than skills taught with FPF. Jimmy and Kenny learned skills taught with FPF in fewer sessions, trials, and total amount of teaching time than skills taught with EC. Rob learned skills in an equivalent number of sessions and trials with both teaching conditions; however, skills with FPF required less teaching time.

Table 2 Efficiency data

Participant	Total number of sessions (FPF)	Total number of sessions (EC)	Total number of trials (FPF)	Total number of trials (EC)	Total amount of time (FPF) Min & sec	Total amount of time (EC) Min & sec
Jimmy	10	14	200	280	67:16	89:51
Rob	10	10	200	200	66:19	69:22
Billy	12	11	240	220	83:38	82:44
Kenny	11	14	220	280	93:20	130:12
Across all participants	43	49	860	980	310:33	368:08

Participant Responding During the Two Teaching Conditions

The researchers measured participant responding during teaching trials across the two teaching conditions. Figure 5 reports the data for each individual participant across the two teaching conditions. The top panel represents the percentage of independent correct trials; the second panel represents the percentage of prompted trials (FPF only); the third panel represents the percentage of incorrect trials; the fourth panel represents the percentage of remedial trials, and the bottom panel represents number of trials to mastery. Across all participants, the overall correct responding was above 90 % across both teaching conditions. Across all participants, the overall correct responding, however, was higher for skills taught with the FPF condition than skills taught with the EC condition. Thus, both procedures resulted in low rates of incorrect responding across the two conditions.

Discussion

Results of this study indicated that both flexible prompt fading (FPF) and error correction (EC) were effective in teaching four children diagnosed with ASD how to expressively label pictures of Muppet characters. In terms of efficiency, across the four participants, results indicated that FPF was more efficient than EC in terms of the total number of teaching sessions, total number of teaching trials, and total amount of instructional time; although individual differences were seen. Furthermore, FPF resulted in fewer errors during teaching. Additionally, the EC condition resulted in better maintenance across the four participants; however, this could be a result of the extra teaching sessions, teaching trials, and teaching time. Anecdotally, Kenny's incorrect responding during the final two full probe sessions was a result of giving silly answers (e.g., "Saxophone Zoot") as opposed to the correct answer (e.g., "Zoot"). Thus, the results of this study showed that both prompting procedures can be effective in teaching children with autistic disorder expressive labeling skills and further expands the research on both FPF and EC in several ways.

First, this study provides further empirical support that flexible prompt fading is an effective prompting system that can result in learning for children with ASD. Flexible

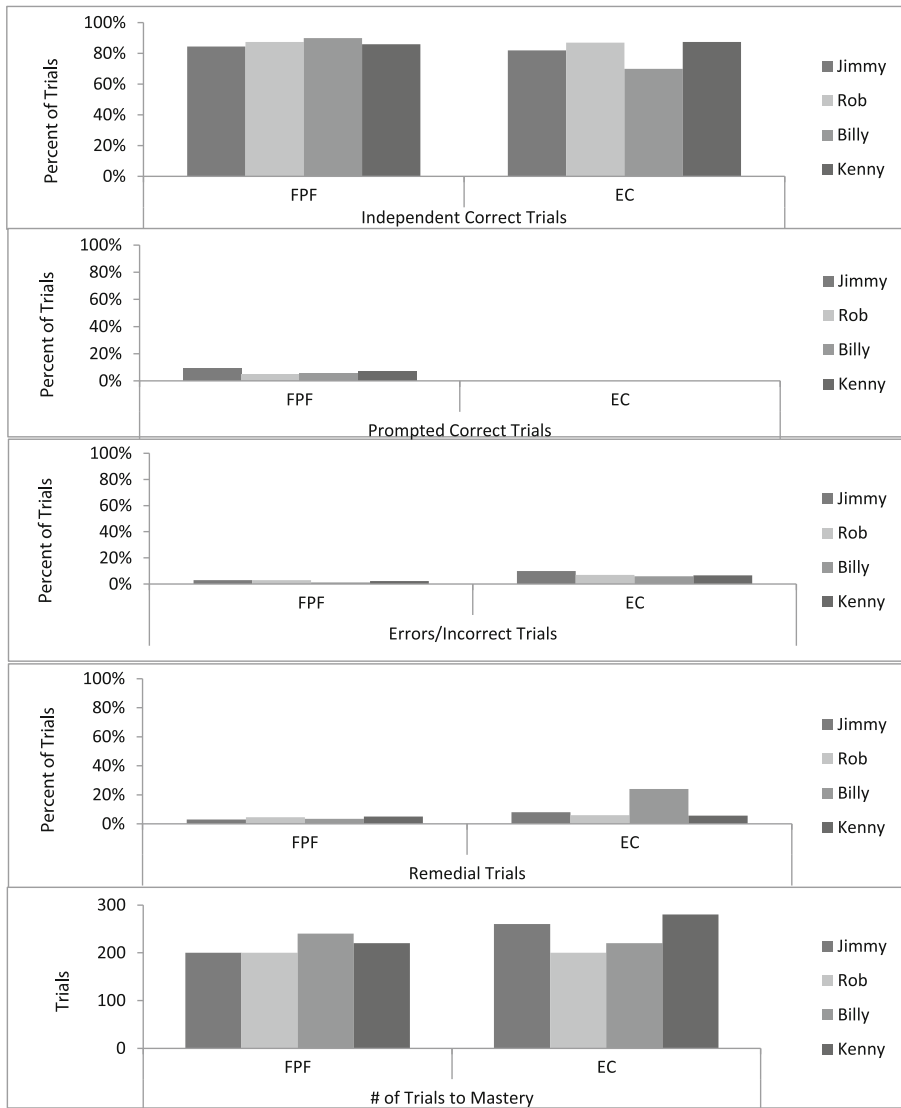


Fig. 5 Participant responding during teaching trials

prompt fading is a prompting system that has been implemented with numerous children diagnosed with ASD (e.g., Leaf et al. 2011b), yet there have been a limited number of studies that have empirically evaluated flexible prompt fading (e.g., Soluaga et al. 2008). Results of this study were similar to the previous studies in that flexible prompt fading was found to be an effective prompting method. Furthermore, the results of this study showed that FPF, which is based upon clinical judgment, can be replicated across different participants and across different research sites.

Second, many of the prompting procedures that are being implemented today require the therapist to adhere to a strict protocol. For example, in no-no prompting the teacher must always allow two independent trials before prompting the student on

the third trial. In constant time delay, the teacher must wait a preset time before providing a prompt to the learner. In flexible prompt fading, however, there is not a fixed formula that a therapist must follow and, thus, he or she is able to make in the moment interventional decisions based on guidelines and parameters throughout teaching. Use of such clinical judgment during teaching allows the teacher to make real-time assessments of and adjustments to teaching procedures based on behaviors being displayed by the learner, which may lead to accelerated rates of learning.

Third, this study demonstrates further empirical proof that error correction procedures can be effective in teaching novel skills to children with ASD. Previous researchers have demonstrated that error correction procedures can be an effective teaching methodology to learn a wide variety of skills (e.g., Leaf et al. 2010; Smith et al. 2006; Worsdell et al. 2005). This study differed from some of the previous studies that implemented error correction procedures in that the participant was not required to respond to the instructional feedback nor were multiple trials presented (e.g., Rodgers and Iwata 1991; Worsdell et al. 2005). Instead, the participant was not required to respond to the instructional feedback and only one remedial trial was provided to the participant. Yet, the results of this study still showed that EC procedures were highly effective in teaching new skills to children with autism. Despite the positive results of this study, and previous research studies, there is still a belief that EC may result in slower skill acquisition than near-errorless procedures and should not be used when first teaching a new skill to a student with ASD (e.g., Gast 2011). This study, and other recent studies (e.g., Leaf et al. 2010), have shown that (anecdotally) error correction procedures do not necessarily result in aberrant behavior or slow the rate of skill acquisition, can be implemented successfully when teaching a student a new skill, and in some cases may even result in students learning skills at a quicker rate.

Today, there are several prompting procedures being implemented to children diagnosed with ASD. One of the goals for clinicians and researchers is to identify the most effective and efficient procedures. Thus, researchers have compared several prompting systems to determine the most efficacious prompting procedures. Results of most of these comparative studies have been mixed, both in terms of effectiveness and efficiency (e.g., Berkowitz 1990; Collier and Reid 1987; Leaf et al. 2010) In this study, both teaching procedures were nearly equally effective and efficient. Therefore, when teaching simple expressive labeling to high functioning students with ASD, it may not matter which prompting/error correction procedure is implemented, since both may result in quick skill acquisition. Clinicians may elect to use an FPF approach as opposed to EC when a student displays aberrant behaviors following incorrect responding or when an error can lead to a string of errors. Thus, an antecedent based prompting strategy may be more suited.

Additionally, the FPF procedure takes a lot of rapid decision making by the clinician, where the EC procedure allows the clinician to follow a more strict protocol. In this study, no specific training was provided to the instructors; however, all of the instructors had an extensive history of implementing discrete trial teaching and applied behavior analysis. Thus, the instructors had strong clinical skills such as: (a) the ability to make moment-to-moment analysis of a participants behaviors and responding; (b) complete understanding of a variety of prompting systems and prompting types; (c) the effective use of reinforcement; and (d) an understanding of functions of behavior.

Thus, it is important that clinicians, teachers, and parents are well trained (see Leaf et al. 2011a, b) and supervised prior to implementing FPF. If clinicians are not well trained or do not display the skills described above an EC procedure may be more appropriate. It would appear that future research is warranted to explore the success of FPF if implemented by clinicians with more varied levels of experience and in different settings in order to assess the generality of these findings.

Despite the positive results of this study, there are some limitations that can be found. First, we elected to use a parallel treatment design to compare the two prompting procedures. Ideally, when utilizing a parallel treatment design, it is desirable to obtain differences in rates of acquisition between the procedures being compared. In this study, however, the participants reached mastery criterion in a near equivalent amount of sessions; thus, some experimental control was lost. One way to minimize this limitation was to place the parallel treatment design within a multiple probe design to show that acquisition of skills occurred only once the interventions were implemented. Nevertheless, some experimental control was lost due to the quick skill acquisition during both teaching conditions. Despite this being a limitation to the research study, it is still important for clinicians to know that both procedures may be equally as effective.

Second, there are several potentially significant measures that were not evaluated in this study, including: aberrant behavior, participant preference, and teacher preference. Although aberrant behavior was not measured, anecdotally there was little to no aberrant behavior throughout the study. Future researchers may wish to directly measure participants' aberrant behavior; specifically, it may be interesting to measure whether the provision of corrective feedback leads to any aberrant behaviors. It may also be of value to determine if the procedures result in differing levels of generalized prompt dependency. Additionally, future researchers may wish to use concurrent chain designs (Hanley et al. 1997, 2005) to measure participants' preference for the two instructional procedures.

A third limitation of the study is that three of the participants had a previous history with the FPF condition and had no history of EC. This previous history may lead to quicker skill acquisition for targets taught with FPF as opposed to targets taught with EC. However, Kenny had no previous history with FPF or EC and his results were similar to the three participants who had a previous history with FPF. Nevertheless, future researchers should be careful to minimize the previous history participants may have with various teaching procedures when comparing those procedures in empirical studies.

A fourth limitation of the study is in regards to the treatment fidelity taken for the FPF condition. In this study, the researchers scored if the teachers displayed correct instructor behaviors (e.g., providing the correct instruction or consequence) and if the teacher decisions resulted in the outcomes specified by the guidelines of the protocol (e.g., participant correct responding was maintained at 80 % or above). However, no measure was taken on whether the teachers used correct "clinical" judgment. Defining and evaluating correct "clinical" judgment may be difficult as one teacher may elect to provide a prompt and another teacher may elect not to provide a prompt. Future researchers may wish to further define good "clinical" judgment and find empirical measures to evaluate the use of such judgment.

Limitations regarding the difficulty obtaining procedural reliability data when treatments involve clinical judgment are not uncommon or unique to FPF. For example,

Wolery and Gast (1984) highlight this issue in the context of graduated guidance and suggest that this relative “lack of adequate procedural integrity makes effectiveness studies difficult to evaluate (p. 59)”. Other behavior change strategies, such as shaping, also rely on responsive clinical decision making. Despite these caveats, shaping and graduated guidance are clinical practices with great utility and evidence for their effectiveness. However, the question of how the current study and procedures can be systematically replicated across other learners and instructors is an empirical one.

A fifth limitation, is that during EC, when a participant is provided instructional feedback followed by a remedial trial the temporal distance is fairly close, which makes the procedure similar to FPF. Future researchers may wish to increase the time from the provision of instructional feedback to the start of the next teaching trial to see if this results in different behavioral change.

Finally, this study only demonstrated the effectiveness of FPF and EC when implemented in a one-to-one setting for a limited number of participants, all who can be considered higher functioning, and for teaching relatively simple skills. Future researchers should extend these findings by evaluating these procedures with more impacted children (e.g., lower IQ scores or higher rates of aberrant behavior), within small and large group instructional formats, and for more difficult skills. In such a manner it may be possible to evaluate the relative effectiveness and efficiency of the procedures with participants of varying characteristics and needs. In addition, in order to determine the most effective and efficient teaching procedures, FPF and EC should also be compared to other commonly implemented prompting systems (e.g., most-to-least, least-to-most, constant time delay).

References

- Berkowitz, S. (1990). A comparison of two methods of prompting in training discrimination of communication book pictures by autistic students. *Journal of Autism and Developmental Disorders*, *20*, 255–262. doi:10.1007/BF02284722.
- Bloh, C. (2008). Assessing transfer of stimulus control procedures across learners with autism. *The Analysis of Verbal Behavior*, *24*, 87–101. Retrieved from <http://www.abainternational.org/TAVB.asp>.
- Bozkurt, F., & Gursel, O. (2005). Effectiveness of constant time delay on teaching snack and drink preparation skills to children with mental retardation. *Education and Training in Developmental Disabilities*, *40*, 390–400. Retrieved from <http://www.daddccc.org/Publications/ETADDJournal.aspx>.
- Charlop, M. H., & Trashowech, J. E. (1991). Increasing autistic children daily spontaneous speech. *Journal of Applied Behavior Analysis*, *24*, 747–761. doi:10.1901/jaba.1991.24-747.
- Collier, D., & Reid, G. (1987). A comparison of two models designed to teach autistic children a motor task. *Adapted Physical Activity Quarterly*, *4*, 228–236. Retrieved from <http://www.journals.humankinetics.com/apaq>.
- Ferster, C. B., & DeMeyer, M. K. (1962). A method for the experimental analysis of behavior of autistic children. *The American Journal of Orthopsychiatry*, *32*, 89–98. doi:10.1111/j.1939-0025.1962.tb.00267.x.
- Gast, D. L. (2011). An experimental approach for selecting a response-prompting strategy for children with developmental disabilities. *Evidenced-based Communication Assessment and Intervention*, *5*, 149–155. doi:10.1080/17489539.2011.637358.
- Gast, D. L., & Wolery, M. (1988). Parallel treatments design: a nested single subject design for comparing instructional procedures. *Education and Treatment of Children*, *11*, 270–285. Retrieved from <http://www.educationandtreatmentofchildren.net/>.
- Hanley, G. P., Piazza, C. C., Fisher, W. W., Contrucci, S. A., & Maglieri, K. A. (1997). Evaluation of client preference for function-based treatment packages. *Journal of Applied Behavior Analysis*, *30*, 459–473. doi:10.1901/jaba.1997.30-459.

- Hanley, G. P., Piazza, C. C., Fisher, W. W., & Maglieri, K. A. (2005). On the effectiveness of and preference for punishment and extinction components of function-based interventions. *Journal of Applied Behavior Analysis*, 38, 51–65. doi:10.1901/jaba.2005.6-04.
- Leaf, R. B., & McEachin, J. J. (1999). *A work in progress: Behavior management strategies and a curriculum for intensive behavioral treatment of autism*. New York: Different Roads to Learning. Retrieved from <http://www.difflearn.com/>.
- Leaf, J. B., Sheldon, J. B., & Sherman, J. A. (2010). Comparison of simultaneous prompting and no-no prompting in two choice discrimination learning with children with autism. *Journal of Applied Behavior Analysis*, 43, 215–228. doi:10.1901/jaba.2010.43-215.
- Leaf, J. B., Oppenheim, M., Dotson, W., Johnson, V. A., Courtemanche, A. B., Sherman, J. A., et al. (2011a). Effects of no-no prompting on teaching expressive labeling of facial expressions to children with and without a pervasive developmental disorder. *Education and Training in Developmental Disabilities*, 46, 186–203. Retrieved from <http://www.daddccc.org/Publications/ETADDJournal.aspx>.
- Leaf, R. B., Taubman, M., McEachin, J. J., Leaf, J. B., & Tsuji, K. H. (2011b). A program description of a community-based intensive behavioral intervention program for individuals with autism spectrum disorders. *Education and Treatment of Children*, 34, 259–285. doi:10.1353/etc.2011.0012.
- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Clinical and Consulting Psychology*, 55, 3–9. doi:10.1037/0022-006x.55.1.3.
- MacDuff, G. S., Krantz, P. J., & McClannahan, L. E. (1993). Teaching children with autism to use photographic activity schedules: maintenance and generalization of complex response chains. *Journal of Applied Behavior Analysis*, 26, 89–97.
- Morse, T. E., & Schuster, J. W. (2000). Teaching elementary students with moderate intellectual disabilities how to shop for groceries. *Exceptional Children*, 66, 273–288. Retrieved from <http://www.ccc.sped/org/exceptionalchildren/>.
- Rodgers, T. A., & Iwata, B. A. (1991). An analysis of error-correction procedures during discrimination training. *Journal of Applied Behavior Analysis*, 24, 775–781. doi:10.1901/jaba.1991.24-775.
- Schreibman, L. (1975). Effects of within-stimulus and extra-stimulus prompting on discrimination learning in autistic children. *Journal of Applied Behavior Analysis*, 8, 91–112. doi:10.1901/jaba.1975.8-91.
- Schumaker, J., & Sherman, J. A. (1970). Training generative verb usage by imitation and reinforcement procedures. *Journal of Applied Behavior Analysis*, 3, 273–287. doi:10.1901/jaba.1970.3-273.
- Smith, T. (2001). Discrete trial training in the treatment of autism. *Focus on Autism and Other Developmental Disabilities*, 16, 86–92. doi:10.1177/108835760101600204.
- Smith, T., Mruzek, D. W., Wheat, L. A., & Hughes, C. (2006). Error correction in discrimination training for children with autism. *Behavioral Interventions*, 21, 245–263. doi:10.1002/bin.223.
- Soluaga, D., Leaf, J. B., Taubman, M., McEachin, J., & Leaf, R. B. (2008). A comparison of flexible prompt fading and constant time delay for five children with autism. *Research in Autism Spectrum Disorders*, 2, 753–765. doi:10.1016/j.rasd.2008.03.005.
- Tarbox, R. S., Wallace, M. D., Penrod, B., & Tarbox, J. (2007). Effects of three-step prompting on compliance with caregiver requests. *Journal of Applied Behavior Analysis*, 40, 703–706. doi:10.1901/jaba.2007.703-706.
- Wolery, M., & Gast, D. L. (1984). Effective and efficient procedures for the transfer of stimulus control. *Topics in Early Childhood Special Education*, 4(3), 52–77.
- Worsdell, A. S., Iwata, B. A., Dozier, C. L., Johnson, A. D., Neidert, P. L., & Thomason, J. L. (2005). Analysis of response repetition as an error-correction strategy during sight-word reading. *Journal of Applied Behavior Analysis*, 38, 511–527. doi:10.1901/jaba.2005-115-04.