



Comparing Error Correction to Errorless Learning: A Randomized Clinical Trial

Justin B. Leaf^{1,2} · Joseph H. Cihon^{1,2} · Julia L. Ferguson¹ · Christine M. Milne^{1,2} · Ronald Leaf¹ · John McEachin¹

Published online: 19 February 2020
© Association for Behavior Analysis International 2020

Abstract

Errorless learning and error correction procedures are commonly used when teaching tact relations to individuals diagnosed with autism spectrum disorder (ASD). Research has demonstrated the effectiveness of both procedures, as well as compared them. The majority of these studies have been completed through the use of single-subject experimental designs. Evaluating both procedures using a group design may contribute to the literature and help disseminate research related to the behavioral science of language to a larger audience. The purpose of the present study was to compare an errorless learning procedure to an error correction procedure to teach tact relations to 28 individuals diagnosed with ASD through a randomized clinical trial. Several variables were assessed, including the number of stimulus sets with which participants reached the mastery criterion, responding during pre- and postprobes, responding during teaching, efficiency, and the presence of aberrant behavior. The results indicated that both procedures were effective, efficient, and unlikely to evoke aberrant behavior, despite participants in the error correction condition engaging in significantly more independent correct responses and independent incorrect responses.

Keywords discrete-trial teaching · error correction · errorless learning · most-to-least prompting · tacting

Intervention programs for children diagnosed with autism spectrum disorder (ASD) commonly include teaching language skills (Barbera & Rasmussen, 2007; Greer & Ross, 2008; Petursdottir & Carr, 2011; Sundberg & Partington, 1998). Within behaviorally based intervention programs, language skills are often classified and targeted based on Skinner's (1957) conceptualization of verbal operants. This often includes

✉ Justin B. Leaf
Jblautpar@aol.com

¹ Autism Partnership Foundation, 200 Marina Drive, Seal Beach, CA 90808, USA

² Endicott College, Seal Beach, CA, USA

directly targeting a variety of tact relations through the use of discrete-trial teaching (DTT; e.g., Grow, Kodak, & Clements, 2017; Majdalany, Wilder, Smeltz, & Lipschultz, 2016), which is a common approach used in behaviorally based intervention programs (Green, 1996; Lovaas, 1987, 2003; Smith, 1999, 2001). A discrete trial involves three core components: (a) a discriminative stimulus (e.g., a teacher-delivered instruction), (b) the learner's response, and (c) a teacher-delivered consequence based on the learner's response. These three core components of DTT do not, by themselves, eliminate or prevent learner errors; therefore, approaches to prevent, reduce, or correct errors have been developed (Mueller, Palkovic, & Maynard, 2007). Generally, these approaches to address errors within DTT could be put into two broad categories: (a) errorless learning procedures and (b) error correction procedures. However, it should be noted that these categories are not mutually exclusive and can commonly occur in combination.

Errorless, or near-errorless, learning procedures involve attempting to prevent errors during all teaching sessions (Mueller et al., 2007). There are several techniques that are used to achieve near-errorless learning that have been described in the literature, including constant time delay (e.g., Ledford, Gast, Luscre, & Ayres, 2008), progressive time delay (e.g., Walker, 2008), simultaneous prompting (e.g., Leaf, Sheldon, & Sherman, 2010), stimulus shaping (e.g., Schilmoeller, Schilmoeller, Etzel, & LeBlanc, 1979), stimulus fading (e.g., Moore & Goldiamond, 1964), and most-to-least prompting (e.g., Demchak, 1990). A common theme across each of these systems is to begin teaching with the most assistive prompt in order to minimize the likelihood of an error. Prompts are faded gradually in a manner that maximizes the probability of continual correct responding while decreasing the probability of errors.

Errorless learning, more generally, has been endorsed by many professionals as the desired approach for teaching individuals diagnosed with ASD (e.g., Green, 2001; Mueller et al., 2007). For instance, Gast (2011b) stated,

In light of research supporting the effectiveness and efficiency of such “near-errorless” response-prompting procedures as CTD [constant time delay], PTD [progressive time delay], and SLP [system of least prompts], it is hard to justify the use of any trial and error procedure, [or] error correction alone. (p. 237)

When reviewing errorless learning procedures, Mueller et al. (2007) came to the same conclusion, stating, “Because there are many potential negative side effects of incorrect responding when using trial and error procedures, teaching procedures can be arranged to eliminate or reduce responding to incorrect choice stimuli” (p. 698). Furthermore, when selecting approaches to train staff in the implementation of DTT, researchers have opted for training staff to use DTT with errorless learning procedures such as most-to-least prompting, which may be due the effectiveness of most-to-least prompting over other prompting methods (e.g., Day, 1987; McDonnell & Ferguson, 1989; Miller & Test, 1989) and recommendations for practitioners to use errorless learning procedures (e.g., Kayser, Billingsley, & Neel, 1986; McDonnell & Ferguson, 1989).

Error correction, on the other hand, involves procedures that are used following an incorrect response that will increase the probability of a correct response on subsequent trials (Cariveau, La Cruz Montilla, Gonzalez, & Ball, 2018). Similar to errorless learning procedures, correct responses by the learner set the occasion for reinforcement.

There are many error correction procedures that often include vocal feedback (e.g., stating “no” following an incorrect response; Leaf et al., 2019), modeling the correct response (Carroll, Joachim, St. Peter, & Robinson, 2015), a short time-out (e.g., turning away from the learner; Carroll et al., 2015), multiple response repetition (Carroll, Owsiany, & Cheatham, 2018), or re-presentation of the trial (Carroll et al., 2015). Although these are common components of error correction procedures, research evaluating them commonly includes slight procedural deviations (Cariveau et al., 2018), and it is likely that variations and combinations of these procedures in clinical practice will occur. Some common procedures that could be classified as error correction procedures that have been empirically evaluated include, but are not limited to, no-no-prompt procedures (e.g., Leaf et al., 2010), single response repetition (e.g., Worsdell et al., 2005), and multiple response repetition (e.g., Worsdell et al., 2005). A common theme across each of these error correction procedures is providing the learner with the opportunity to respond independently prior to the instructor taking action to influence how the learner responds.

Although many studies have demonstrated the effectiveness of error correction procedures in isolation or in combination with errorless learning procedures (e.g., Carroll et al., 2015; Kodak, Fuchtman, & Paden, 2012; McGhan & Lerman, 2013; Plaisance, Lerman, Laudont, & Wu, 2016; Townley-Cochran, Leaf, Leaf, Taubman, & McEachin, 2017), some have argued against the use of procedures that allow the learner to err and are in favor of errorless learning strategies (e.g., Green, 2001; Mueller et al., 2007). For example, Gast (2011b) asked, “Why risk such a response [referring to errors] by a child if we have evidence-based instructional procedures (CTD, PTD, most to least prompting) that can and do reduce error during the early stages of learning?” (p. 237). Others have argued against the use of error correction procedures with claims that allowing the learner to err leads to more errors (Mueller et al., 2007), makes teaching sessions aversive for learners (LaVigna & Donnellan, 1986), or produces negative emotional responding (LaVigna & Donnellan, 1986). Given the common use of DTT within language instruction for individuals diagnosed with ASD and the disagreement about allowing errors, empirical investigations have included comparisons of errorless learning and error correction procedures.

Leaf et al. (2010) compared simultaneous prompting to no-no-prompt procedures to teach three children diagnosed with ASD a variety of skills (e.g., math skills, listener behavior, responses to *wh*-questions). Simultaneous prompting consisted of delivering an instruction and a controlling prompt (i.e., a prompt that resulted in the learner responding with 100% accuracy in a prior assessment) on each trial. No-no-prompt procedures consisted of allowing two consecutive incorrect responses to occur before a prompted trial occurred. After the first incorrect response, the researcher stated “no” and re-presented the trial. If the second presentation of the trial also resulted in an incorrect response, the researcher said “no” and repeated the same trial, but the instruction was accompanied by a controlling prompt. The relative effectiveness of the two approaches was evaluated using a parallel treatment design. Participants reached the mastery criterion for all of the skills taught with no-no-prompt procedures but for only 10% of the skills taught using simultaneous prompting. Participants also maintained more of the skills taught with the no-no-prompt procedure compared to the skills taught with simultaneous prompting.

Fentress and Lerman (2012) compared most-to-least prompting to no-no-prompt procedures to teach a variety of skills (e.g., listener behavior, imitation, matching) to

four children diagnosed with autism. Most-to-least prompting involved a full physical prompt (hand-over-hand prompt), followed by a partial physical prompt, and then a less intrusive prompt (Fentress & Lerman, 2012). No-no-prompt procedures occurred similarly to Leaf et al. (2010), with the inclusion of a prompt-fading system. That is, when prompts were provided on the third trial, those prompts were faded similarly to how prompts were faded in the most-to-least condition. The relative effectiveness of both procedures was evaluated using an alternating-treatments design. Results of the study indicated that both procedures were effective, but the most-to-least prompting procedure reduced the frequency of learner errors and resulted in better maintenance, whereas the no-no-prompt procedure was found to be more efficient (i.e., participants reached the mastery criterion faster with the no-no-prompt procedure).

Leaf et al. (2014a) compared an error correction procedure to most-to-least prompting to teach listener behavior (referred to as *receptive labels* in the study) to two children diagnosed with ASD. The error correction procedure consisted of the teacher presenting an instruction and providing the participant with an opportunity to respond independently. Correct responses resulted in praise and brief access to a toy. Incorrect responses resulted in the therapist saying, “No, that’s not it,” and then gesturing to the correct stimulus while saying, “This is the [correct response].” The most-to-least prompting procedure involved a three-step prompting hierarchy consisting of a point prompt, a positional prompt, and the instruction alone. The researchers used an adapted alternating-treatments design embedded in a multiple-baseline design to evaluate the effectiveness of the prompting procedures. The results indicated that both procedures were effective (i.e., both participants reached the mastery criterion for all targets), similar in terms of efficiency with respect to the number of sessions to mastery, and similar with respect to the number of incorrect responses during teaching trials.

Leaf et al. (2014b) compared error correction to most-to-least prompting to teach two children diagnosed with ASD tact relations. The error correction procedure consisted of the researcher presenting an instruction (e.g., “What is it?”) and allowing the learner the opportunity to respond independently. Correct responses resulted in praise, and incorrect responses resulted in corrective feedback and informative feedback (e.g., “No, it’s Kermit.”). The most-to-least prompting hierarchy consisted of (a) the controlling prompt (i.e., stating the correct response), (b) a multiple-choice question (e.g., “Is it Fozzie, Animal, or Beaker?”), and (c) the instruction alone (i.e., no prompt). Correct responses resulted in praise, and incorrect responses resulted in no feedback and movement to a more assistive prompt on the next trial. The results of an alternating-treatments design embedded in a multiple-baseline design demonstrated that the error correction procedure was slightly more effective (i.e., one set did not reach the mastery criterion in the most-to-least condition) and more efficient (i.e., the mastery criterion was reached in fewer sessions in the error correction condition) and resulted in a higher percentage of correct teaching trials. Mixed results were observed across both approaches with respect to maintenance of the skills taught.

Although there have been several studies documenting the effectiveness of error correction as the sole intervention (see Cariveau et al., 2018, for a review), evaluating procedural variations of error correction (e.g., Carroll et al., 2018), and comparing error correction procedures to errorless learning procedures (e.g., Leaf et al., 2014b), the debate continues as to the conditions under which each procedure should be selected or

avoided. One potential reason for the continued debate could be the use of single-subject experimental designs. Although these are a hallmark of behavior-analytic research, single-subject experimental designs have been criticized for lacking external validity and failing to demonstrate causal relationships, and some do not consider them true experimental designs (Kazdin, 2011). Nonetheless, single-subject experimental designs do offer important benefits (e.g., repeated observations across time, continuous assessment, participants serving as their own baseline) and will continue to be necessary and useful. For example, single-subject experimental designs conducted thus far help identify the conditions under which errorless learning procedures or error correction procedures are more or less effective at the individual level. However, group designs could help extend the previous research and improve our understanding of errorless learning and error correction procedures for individuals diagnosed with ASD in a variety of ways.

First, Smith et al. (2007), when describing the design of research studies for individuals diagnosed with ASD, suggested that group studies, such as randomized clinical trials, are the authoritative test of the efficacy of an intervention. If this is the case, a group study evaluating errorless learning and error correction procedures could provide an answer to the debate over which is most effective or efficient. Second, Smith et al. noted that group designs can “enable investigators to analyze variables such as subject characteristics that may be associated with favorable or unfavorable responses to intervention” (p. 361). This may be especially desirable given the common association between punishment-based procedures, such as error correction, and undesirable behavior (e.g., aggression; Ferster & DeMeyer, 1962; Gast, 2011a; Lerman & Vorndran, 2002; Sulzer-Azaroff & Mayer, 1991). Finally, group designs are commonly the standard in other related fields (Smith, 2012; Smith et al., 2007). It may be the case that conducting a group design evaluating errorless learning and error correction procedures would help applied behavior analysis, as it relates to the treatment of ASD, disseminate research related to the behavioral science of language to a larger audience. Therefore, the purpose of this study was to compare the relative effectiveness of an error correction procedure to an errorless learning procedure (i.e., most-to-least prompting) using a randomized clinical trial (RCT) to teach tact relations to 28 children diagnosed with ASD.

Method

Participants

Participants were 28 children diagnosed with ASD. To be included in this study, children had to meet the following criteria: (a) the child had an independent diagnosis of ASD; (b) the child had an Expressive One-Word Picture Vocabulary Test, Fourth Edition (EOWPVT; Martin & Brownell, 2010), standard score above 75; (c) the child had a Peabody Picture Vocabulary Test, Fourth Edition (PPVT; Dunn & Dunn, 2007), standard score above 75; (d) the child had a previous history with DTT; (e) caregivers consented for their child to participate in the study; (f) the child would benefit from instructional strategies for tact relations, or tact relations were a current goal; and (g) praise had been identified to function as a reinforcer. The inclusion criteria were

selected in an effort to recruit a large number of participants for whom the interventions employed and targets selected would likely be appropriate. Participants were recruited from the agency in which the study was conducted and through a mass e-mail service. Following obtaining caregivers' consent, the first 26 participants were randomly assigned (via [random.org](https://www.random.org)) to one of two groups: (a) error correction or (b) errorless learning (i.e., most-to-least prompting). All children had a history of receiving intervention in applied behavior analysis that most likely included the use of prompting or error correction. However, the researchers found that it was not possible to accurately and completely identify each participant's complete history with each of the procedures evaluated within this study. The scores of the first 26 children yielded statistically significant differences on the EOWPVT; therefore, the final two participants were assigned into each of the conditions so there would be no differences across the two groups.

Table 1 provides individual participant characteristics and group means. A paired *t* test was conducted to evaluate preintervention demographic variables across the two conditions. The results indicated no significant difference in the age of participants between the error correction condition ($M = 5.071$; $SD = 1.859$) and the errorless learning condition ($M = 5.786$; $SD = 1.847$), $t(26) = -1.020$, $p = .317$. There was no significant difference in the Vineland Adaptive Behavior Scales, Second Edition (VABS; Sparrow, Cicchetti, & Balla, 2005), scores of participants between the error correction condition ($M = 87.417$; $SD = 7.786$) and the errorless learning condition ($M = 83.231$; $SD = 13.361$), $t(23) = .354$; $p = .354$. There was no significant difference in the PPVT standard score of participants between the error correction condition ($M = 107$; $SD = 21.071$) and the errorless learning condition ($M = 101.357$; $SD = 16.213$), $t(26) = .794$; $p = .434$. There was no significant difference in the EOWPVT standard score of participants between the error correction condition ($M = 114.929$; $SD = 20.764$) and the errorless learning condition ($M = 104.071$; $SD = 13.826$), $t(26) = 1.628$; $p = .115$.

Setting

Sessions were conducted in rooms in a private agency that provides behavioral intervention to individuals diagnosed with ASD. The rooms consisted of a table designed for a child, two child-sized chairs, educational materials (e.g., books, toys, whiteboard), and adult furniture (e.g., desks, couch, desk chairs).

Stimuli

The same stimuli were used for all participants and conditions to best evaluate any differences between the error correction and errorless learning procedures used. That is, if different stimuli were used for each participant or each procedure, it is possible that the stimuli used could be partly, or wholly, responsible for any differences. Therefore, the same target stimuli were used in both conditions and consisted of six pictures of unknown comic book characters. The stimuli were selected based on common themes within popular media. Given the frequent release of movies and TV shows related to comic books, it was determined that comic book characters may have some immediate utility and would be maintained by the respective verbal communities of all

Table 1 Participant Demographics Across the Two Groups

Participant	Group	Age	VABS ^a	PPVT	EOWPVT
EC1	EC	9 yr 8 mo	—	98	122
EC2	EC	4 yr 7 mo	80	132	145
EC3	EC	5 yr 6 mo	84	94	114
EC4	EC	5 yr 0 mo	103	120	145
EC5	EC	5 yr 0 mo	91	120	117
EC6	EC	8 yr 7 mo	87	117	110
EC7	EC	7 yr 9 mo	95	135	132
EC8	EC	8 yr 0 mo	—	143	145
EC9	EC	8 yr 0 mo	96	92	101
EC10	EC	4 yr 2 mo	80	82	89
EC11	EC	4 yr 1 mo	79	107	100
EC12	EC	3 yr 8 mo	91	84	111
EC13	EC	4 yr 6 mo	80	97	95
EC14	EC	6 yr 8 mo	83	77	83
<i>Average</i>		<i>6 yr 1 mo</i>	<i>87</i>	<i>107</i>	<i>114</i>
EL1	EL	8 yr 7 mo	74	109	97
EL2	EL	7 yr 4 mo	—	106	118
EL3	EL	6 yr 11 mo	91	130	121
EL4	EL	5 yr 1 mo	94	78	83
EL5	EL	6 yr 5 mo	69	77	94
EL6	EL	5 yr 2 mo	79	90	106
EL7	EL	5 yr 2 mo	71	81	84
EL8	EL	6 yr 2 mo	93	118	119
EL9	EL	7 yr 6 mo	116	100	107
EL10	EL	8 yr 4 mo	70	92	104
EL11	EL	9 yr 8 mo	80	99	89
EL12	EL	9 yr 9 mo	91	111	101
EL13	EL	3 yr 10 mo	80	109	107
EL14	EL	5 yr 11 mo	74	119	127
<i>Average</i>		<i>6 yr 10 mo</i>	<i>83</i>	<i>101</i>	<i>104</i>

Note. VABS = Vineland Adaptive Behavior Scales, Second Edition; PPVT = Peabody Picture Vocabulary Test, Fourth Edition; EOWPVT = Expressive One-Word Picture Vocabulary Test, Fourth Edition.

^a Does not include scores from three participants (i.e., EC1, EC8, and EL2); however, this measure did not determine inclusion in the study.

participants. Although this could create a possibility of exposure outside of experimental sessions, we felt the social validity of the stimuli used outweighed this possible risk. The six comic book characters were randomly divided into three sets of two. The two targets in the first set were Iron Fist and Daredevil. The two targets in the second set

were Black Bolt and Hawkgire. The two targets in the third set were Yellowjacket and Carnage.

Dependent Measures

The main dependent measure was participant responding during full probes, which occurred prior to and following intervention. Participant responding was categorized as independent correct, incorrect, or no response. During pre- and postprobes, participant responding was categorized as a correct, incorrect, or no response. An *independent correct response* was defined as engaging in the vocal response that corresponded to the presented stimulus within 5 s of its presentation (e.g., saying “Iron Fist” in the presence of the picture of Iron Fist after the interventionist said, “Who is this?”). An *independent incorrect response* was defined as engaging in a vocal response that did not correspond to the presented stimulus within 5 s of its presentation. *No response* was defined as not engaging in any vocal response within 5 s of the stimulus presentation.

The second dependent measure was the number of stimulus sets for which each participant reached the mastery criterion as assessed through daily probes. The mastery criterion was defined as the learner engaging in 100% independent correct responses on all trials across three consecutive daily probes. If this mastery criterion was not met following 20 intervention sessions, the next set was introduced. If any participant did not reach the mastery criterion for a set, they were taught following the completion of the study. During daily probes, response categories and definitions were the same as in full probes (i.e., independent correct, incorrect, and no response).

The third dependent measure assessed participant responding during teaching sessions across the two conditions. Again, we used independent correct and incorrect responses as previously described during the error correction and errorless learning conditions. Additionally, during the errorless learning condition, participant responding was also categorized as prompted correct or incorrect. A *prompted correct response* was defined as engaging in the vocal response that corresponded to the presented stimulus within 5 s of its presentation following a prompt. The percentage of prompted correct responding was measured by dividing the number of trials with prompted correct responses by the total number of trials per session and multiplying by 100. A *prompted incorrect response* was defined as engaging in a vocal response that did not correspond to the presented stimulus within 5 s of its presentation following a prompt (e.g., saying “Joe” in the presence of the picture of Iron Fist after the interventionist vocally stated, “Iron”). The percentage of prompted incorrect responding was measured by dividing the number of trials with prompted incorrect responses by the total number of trials per session and multiplying by 100. All responses during teaching were scored using a datasheet that outlined the trial order. We also measured the occurrence of aberrant behavior during teaching trials. *Aberrant behavior* was defined as the participant engaging in any aggression (e.g., hitting or attempting to hit the researcher), self-injurious behavior (e.g., hitting oneself), stereotypic behavior (e.g., hand flapping), crying, yelling, or other (i.e., an aberrant behavior not predefined). If aberrant behavior occurred, the researcher marked the trial during which aberrant behavior occurred, as well as the topography of the aberrant behavior.

The final dependent measure was an evaluation of efficiency, assessed via daily probes. To determine the efficiency of each condition, we evaluated the total number of

sessions until a participant reached the mastery criterion for each set or until teaching had concluded for all three sets (i.e., if the mastery criterion had not been reached within 20 sessions). It should be noted that this number was capped at 20. Specifically, if the participant had not reached the mastery criterion for a set within 20 sessions, the next set was introduced.

Probes

Full probes Full probe sessions were used to make pre- and postcomparisons of the effectiveness of each of the conditions. Two full probe sessions for target stimuli were conducted immediately prior to and following intervention. Full probe trials began with the interventionist holding up a stimulus and providing an instruction for the participant to respond (e.g., “Who is it?”). The participant was given 5 s to respond. Neutral feedback (e.g., “Thanks.”) was provided regardless of accuracy, and no prompts were provided on any trial. Each stimulus was presented three times for a total of 18 trials. The order of the presentation of the stimuli was randomized prior to each full probe. Interventionists conducting full probes prior to teaching were kept blind to which condition the participant was assigned; however, they were not blind to which condition the participant was assigned during full probes following intervention.

Daily probes With the exception of the first intervention session, daily probes were conducted prior to each intervention session across both conditions. Daily probes were similar to full probes, but only the two stimuli currently being targeted during intervention sessions were included (e.g., Black Bolt and Hawkgirl). Each stimulus was presented in a random order three times for a total of six trials during daily probes. Neutral feedback (e.g., “Thank you,” “OK”) was provided regardless of participant responses across all trials. The mastery criterion for each stimulus set was a participant responding correctly on 100% of trials during a daily probe for three consecutive daily probe sessions. Once a participant reached the mastery criterion for one set, the next set was introduced. If the mastery criterion was not met after 20 sessions, intervention for that set was discontinued and the next set was introduced.

General Procedure

Sessions occurred once a day for 2 to 5 days a week across both conditions depending on participants’ schedules (there were no noticeable differences between the two groups with respect to scheduling). Sessions lasted between 5 and 10 min. A session consisted of either a full probe, to assess responding prior to or after intervention, or a daily probe followed by intervention based on the condition to which the participant was assigned. Across both intervention conditions, each stimulus within the targeted set was presented 10 times, in a random order, for a total of 20 trials in each intervention session.

Intervention

Error correction The error correction procedure in this study was identical to the procedure used by Leaf et al. (2014b), which combined components of less intrusive

error correction procedures (i.e., differential reinforcement for correct responses and error statements; Cariveau et al., 2018) with verbally stating the correct response. This type of error correction was selected based on its low level of intrusiveness, documented effectiveness, and common use within the agency in which this study took place. Each trial began with the researcher presenting the stimulus (i.e., picture of the comic book character) and providing an instruction (e.g., “What is his/her name?”). If the participant engaged in an independent correct response within 5 s, the researcher provided praise (e.g., “Nailed it!”). Praise was selected because conditioning praise as a reinforcer was part of all of the participants’ programming at some point, and, as a result, it had been reported by the participants’ case managers to be an effective reinforcer during the participants’ sessions. If the participant engaged in an incorrect response or no response, the researcher provided corrective and informative feedback (e.g., “No, it is Daredevil.”). Any aberrant behavior was ignored if it was nondangerous (e.g., crying, yelling) and redirected if it was dangerous (e.g., aggression, self-injurious behavior). However, at no point in time did any dangerous topographies of aberrant behavior occur.

Errorless learning Most-to-least prompting was used to minimize errors within the errorless learning condition. The researchers selected most-to-least prompting based on the research indicating that it is often more effective and efficient than other errorless prompting systems (Day, 1987; McDonnell & Ferguson, 1989; Miller & Test, 1989). The most-to-least prompting procedure used within this study consisted of a three-step prompt hierarchy: (a) a full echoic prompt (i.e., stating the full comic book character’s name; e.g., “Carnage”), (b) a partial echoic prompt (i.e., stating part of the comic book character’s name; e.g., “Car”), and (c) no prompt. The criterion to move to a less assistive prompt was a correct response on two consecutive teaching trials. The criterion to move to a more assistive prompt was an incorrect response on any teaching trial. Each stimulus within a set was on its own prompt hierarchy, and movement on the prompting hierarchy for each target was independent of responding on the other target within the set. On subsequent sessions, the researcher began with the prompt level from the previous session. If the participant engaged in an independent or prompted correct response within 5 s, the researcher provided praise (e.g., “You got it!”). If the participant engaged in an independent or prompted incorrect response or no response, the researcher provided corrective feedback (e.g., “No.”). Any aberrant behavior was ignored if it was nondangerous (e.g., crying, yelling) and redirected if it was dangerous (e.g., aggression, self-injurious behavior). However, at no point in time did any dangerous topographies of aberrant behavior occur.

Interobserver Reliability

A second researcher independently (in vivo or via videotapes) recorded participant responding during 45.8% of pre- and postprobes, 43.9% of daily probes, 38.6% of error correction teaching sessions, and 47.5% of errorless learning conditions. Interobserver agreement (IOA) was calculated by totaling the number of agreements divided by the number of agreements plus disagreements and multiplying by 100. Agreements consisted of trials in which both observers scored the same participant response (i.e., correct, incorrect, prompted correct, prompted incorrect, no response, and aberrant behavior;

e.g., if both observers scored that aggression occurred on the same trial). Disagreements consisted of trials in which both observers scored different participant responses (i.e., correct, incorrect, prompted correct, prompted incorrect, no response, and aberrant behavior; e.g., if one observer scored a correct response and the other observer scored a prompted correct response). IOA across all sessions and participants during pre- and postprobes was 100%, during daily probes IOA was 99.8% (range 67%–100% across sessions), during error correction IOA was 99.4% (range 85%–100% across sessions), and during errorless learning IOA was 99.9% (range 95%–100% across sessions).

Treatment Fidelity

Treatment fidelity was taken during 45.2% of pre- and postprobes, 44.6% of daily probes, 33% of error correction sessions, and 47.1% of errorless learning sessions. Correct researcher behaviors on each probe trial (i.e., pre-, post-, and daily probes) consisted of (a) presenting the stimulus (i.e., picture of the comic book character), (b) providing an instruction (e.g., “What is his/her name?”), (c) providing up to 5 s for the participant to respond, and (d) providing neutral feedback regardless of accuracy. Correct researcher behavior during each error correction teaching trial consisted of (a) presenting the stimulus (i.e., picture of the comic book character), (b) providing an instruction (e.g., “What is his/her name?”), (c) providing up to 5 s for the participant to respond, (d) providing praise (e.g., “You got it!”) following a correct response, and (e) providing corrective and informative feedback (e.g., “No, it is Daredevil.”) following an incorrect response or no response. Correct researcher behavior during each errorless learning teaching trial consisted of (a) presenting the stimulus (i.e., picture of the comic book character), (b) providing an instruction (e.g., “What is his/her name?”), (c) providing the correct level of prompt, (d) providing up to 5 s for the participant to respond, (e) providing praise (e.g., “You got it!”) following correct and prompted correct responses, and (f) providing corrective feedback (e.g., “Nope.”) following an incorrect response or no response. Treatment fidelity was calculated by dividing the number of trials during which the researcher engaged in all of the correct steps by the total number of trials and multiplying by 100. Treatment fidelity was 99.9% (range 94.4%–100% across sessions) during pre- and postprobes, 99.9% (range 83.3%–100% across sessions) during daily probes, and 100% during error correction teaching trials and errorless learning trials.

Design

The two intervention approaches were assessed using a two-group RCT. An RCT differs from a randomized control trial in that each group receives treatment and comparisons are made on the relative effectiveness of each treatment (Abbott, 2014). This design could be compared to an alternating-treatments design (Barlow & Hayes, 1979) in that both are designed to compare the effects of two or more interventions, procedures, or approaches.

Statistical Analyses

Initial analyses of differences between conditions were conducted via an independent-samples *t* test to determine the potential for statistical differences across participants in

the two groups in terms of independent correct responding, independent incorrect responding, prompted correct responding, prompted incorrect responding, total correct responding (i.e., independent and prompted), total incorrect responding (i.e., independent and prompted), and aberrant behavior. An independent-samples t test is designed to compare means between two groups where there are different participants in each group. If, however, the Shapiro–Wilk test of normality indicated the data were not amenable to an independent-samples t test, a Mann–Whitney rank sum test was used instead. This is a common approach if the variances are not equal, which requires a variance-stabilizing transformation or a modification of the t test. The Mann–Whitney rank sum test is a nonparametric test to compare outcomes between two independent groups. All statistical analyses were conducted using SigmaPlot, Version 13.0 (Systat Software, San Jose, CA).

Results

Full Probes

Figure 1 displays the results for the main dependent measure, the mean percentage of correct responding for all participants during full probes. Prior to intervention, the mean percentage of correct responding was 0% (as planned, to ensure no prior learning history) across the groups for the targeted stimuli with no difference between groups. Therefore, it was only necessary to evaluate postperformance data. Following intervention, the mean percentage of correct responding was 85.7% and 77.1% for participants in the error correction and errorless learning conditions, respectively. When comparing postperformance on full probes between the two conditions, a Mann–Whitney U test indicated there was a significant difference, $U = 53$, $p = .039$, in correct responding for participants in the error correction condition when compared to participants in the errorless learning condition.

Number of Stimulus Sets With Which Participants Reached the Mastery Criterion

A total of 42 sets was introduced across the participants in each condition (i.e., three sets per participant with 14 participants in each condition). All but one participant (i.e., EC12, who reached the mastery criterion on two of the three sets) reached the mastery criterion in the error correction condition for all three sets, for a total of 97.6% of sets mastered (i.e., 41 out of 42 sets). Similarly, all but one participant (i.e., EL6, who reached the mastery criterion for one of the three sets) reached the mastery criterion in the errorless teaching condition for all three sets, for a total of 95.2% of sets mastered (i.e., 40 out of 42 sets).

Responding During Teaching

Table 2 provides a summary of responding during intervention sessions across the two conditions. The mean percentage of independent correct responding was 90.8% and 75.3% across the error correction and errorless learning conditions, respectively. The results of an independent-samples t test indicated there was a significant difference in independent correct responding between the error correction condition ($M = 90.8$; $SD =$

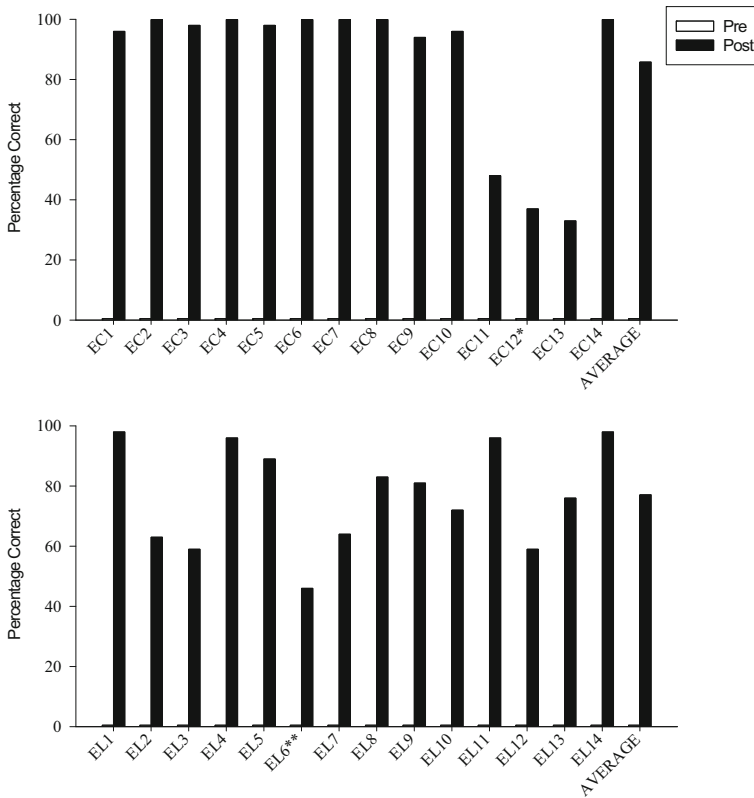


Fig. 1. The mean percentage of independent correct responding for each participant and condition during pre- and postprobes. The number of asterisks indicates the number of sets in which the participant did not reach the mastery criterion. EC = error correction; EL = errorless learning

5.575) and the errorless learning condition ($M = 75.3$; $SD = 11.786$), $t(26) = 4.456$; $p < .001$. The mean percentage of independent incorrect responding was 9.2% and 2.9% across the error correction and errorless learning conditions, respectively. A Mann–Whitney U test indicated there was a significant difference, $U = 19$, $p < .001$, in independent incorrect responding for participants in the error correction condition when compared to participants in the errorless learning condition.

The mean percentage of prompted correct and prompted incorrect responding in the errorless learning condition was 19.3% and 2.6%, respectively. Data on prompted responses in the error correction condition are not reported, given they could not occur in that condition based on the procedures within that condition. The mean percentage of overall correct responding (i.e., independent and prompted) was 90.8% and 94.6% across the error correction and errorless learning conditions, respectively. A Mann–Whitney U test indicated there was a significant difference, $U = 52$, $p = .036$, in overall correct responding for participants in the errorless learning condition when compared to participants in the error correction condition. This is not surprising, however, given that it was not possible for prompted correct responses to occur in the error correction condition. The mean percentage of overall incorrect responding was 9.2% and 5.5% across the error correction and errorless learning conditions, respectively. A Mann–

Table 2 The Percentage of Responding During Teaching Conditions Across the Two Conditions

Condition	Independent Correct	Independent Incorrect	Prompted Correct	Prompted Incorrect	Total Correct	Total Incorrect	Aberrant Behavior
Error correction	90.8%	9.2%			90.8%	9.2%	0.07%
Errorless learning	75.3%	2.9%	19.3%	2.6%	94.6%	5.5%	.55%

Whitney U test indicated there was a significant difference, $U = 52.5$, $p = .039$, in overall incorrect responding for participants in the error correction condition when compared to participants in the errorless learning condition.

The mean percentage of aberrant behavior was 0.07% and 0.55% across the error correction and errorless learning conditions, respectively. A Mann–Whitney U test indicated there was no significant difference, $U = 77.5$, $p = .169$, in aberrant behavior when comparing participants in the error correction condition to participants in the errorless learning condition.

Sessions to Mastery

Figure 2 displays the number of sessions to mastery for each set of stimuli across all participants in both conditions' groups, as well as the average number of sessions to mastery for each set of stimuli within each condition. The top panel displays the results for participants in the error correction condition, and the bottom panel displays the results for participants in the errorless learning condition. Asterisks next to a participant's number indicate the participant did not reach mastery criterion on one or more sets. The total number of sessions, then, represents the number of sessions to reach the mastery criterion or the termination criterion (i.e., 20 sessions), whichever came first. A total of 73, 68, and 89 sessions were required across the 14 participants in the error correction condition for Sets 1, 2, and 3, respectively. A total of 79, 84, and 72 sessions were required across the 14 participants in the errorless learning condition for Sets 1, 2, and 3, respectively. Overall, participants required an average of 16 sessions to reach the mastery criterion across all three stimulus sets in both conditions. A Mann–Whitney U test indicated there was no significant difference, $U = 87.5$, $p = .643$, in the average number of sessions required to reach the mastery criterion when comparing participants in the error correction condition to participants in the errorless learning condition.

Discussion

The purpose of this study was to compare the relative effectiveness of an error correction procedure to an errorless learning procedure (i.e., most-to-least prompting) using an RCT to teach tact relations to 28 children diagnosed with ASD. In general, the results of this study are consistent with previous research on error correction and errorless learning procedures. Both procedures were effective with respect to stimulus sets mastered and resulted in significant changes from pre- to postprobes. Second, both

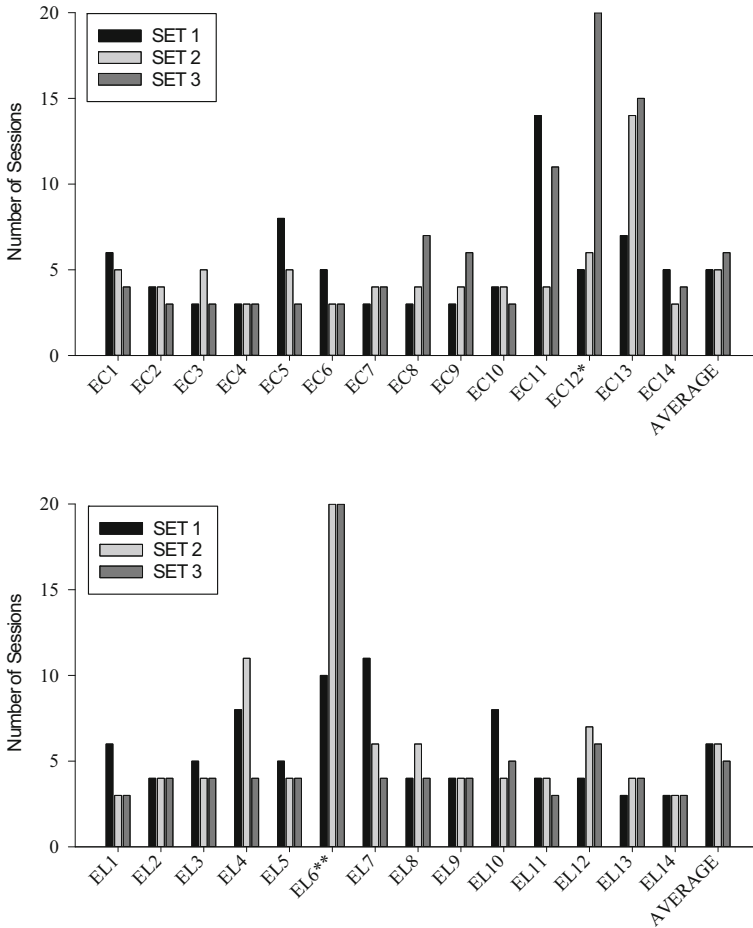


Fig. 2. The total number of sessions to reach the mastery criterion for each participant for each set and the average per set across conditions. The number of asterisks indicates the number of sets in which the participant did not reach the mastery criterion

procedures were efficient with respect to the number of sessions required to reach the mastery criterion, with no significant differences between the two conditions. Third, participants in both conditions engaged in minimal rates of aberrant behavior, with no significant differences between the two conditions.

However, there were differences worth mentioning. First, the results indicated that performance on postprobes for participants assigned to the error correction condition was significantly better when compared to the errorless learning condition. Second, the results indicated that independent correct responding during teaching for participants in the error correction condition was significantly higher than in the errorless learning condition. These results are consistent with previous research comparing error correction procedures to errorless learning procedures. Researchers have compared error correction procedures to errorless learning procedures to teach a variety of behaviors to individuals diagnosed with ASD and have found error correction procedures to be

either more effective, more efficient, or both more effective and efficient than errorless learning procedures (e.g., Fentress & Lerman, 2012; Leaf et al., 2010).

This study contributes to the existing literature on error correction and errorless learning procedures through the use of a group design. To our knowledge, this is the first study to evaluate error correction and errorless learning procedures using an RCT. Previous studies have evaluated these procedures in isolation or through comparison using single-subject experimental designs, which remains an important research endeavor. This study adds to the natural progression of research described by Smith (2012) and Smith et al. (2007), in which group research is informed through previous studies using single-subject methodology. As such, the use of an RCT can help disseminate research related to the behavioral science of language to a larger audience and contribute to the external validity of behavior-analytic procedures for teaching tact relations to individuals diagnosed with ASD (Campbell & Stanley, 1963; Kazdin, 2011).

The findings of this study, as well as other studies, have implications for clinical practice with respect to ASD intervention. First, when using similar instructional preparations (e.g., introducing targets in sets of two), the use of the error correction procedure evaluated in this study may lead to acquisition with more independent correct responses. Second, concerns about errors during the acquisition of tact relations being detrimental to the individual or to the learning process appear unsupported based on these results. Although it appears that the unacceptability of errors is accepted dogma, the results of the present study indicated that error correction procedures can result in more independent correct responding than errorless learning procedures can. Third, it appears that error correction, as a punishment procedure, can be used without evoking undesired aberrant behavior commonly associated with punishment-based procedures. In this study, the occurrence of aberrant behavior (e.g., aggression, crying, self-injury) was comparable and extremely rare across both conditions. Moreover, the rare instances of aberrant behavior involved topographies that would be considered nondangerous (e.g., yelling). Therefore, the use of error correction procedures should not be avoided simply because punishment-based procedures are commonly associated with undesired side effects, as that is inconsistent with the present study, as well as with others (e.g., Leaf et al., 2019; Lovaas & Simmons, 1969; Matson & Taras, 1989).

Despite the contributions of this study to the literature base and its clinical implications, there are several limitations that could be addressed in future evaluations. First, this study included participants with sophisticated vocal-verbal repertoires. It remains unclear whether similar results would be obtained with participants who have more limited vocal-verbal repertoires or more cognitive impairments (e.g., $IQ < 70$). Future RCTs evaluating error correction and errorless learning procedures should include individuals differing from those included in this study with respect to scores on standardized assessments, cognitive functioning (e.g., IQ), and the presence of challenging behavior (e.g., aggression).

Second, this study evaluated an error correction procedure and an errorless learning procedure to teach only one skill (i.e., tacting, sometimes referred to as *expressive labeling*). Therefore, it remains unknown if the results would be replicated when teaching other skills (e.g., listener behavior, sometimes referred to as *receptive labeling*). Despite this limitation, previous studies have compared error correction procedures to errorless learning procedures to teach other skills (e.g., Leaf et al., 2010); however, these studies used single-subject experimental designs. As such, a potentially

fruitful area for future research would be using an RCT to compare error correction procedures to errorless learning procedures to teach other skills.

Third, this study only evaluated one method of errorless learning (i.e., most-to-least prompting) and one method of error correction. As previously noted, there are several errorless learning techniques, including stimulus fading, stimulus shaping, response prevention, delayed prompting, superimposition with stimulus fading, and superimposition with stimulus shaping. Similarly, there are several error correction procedures not evaluated here, such as demonstration, remove and re-present, and multiple response repetition (Cariveau et al., 2018). There are also other prompting systems that use techniques other than the one evaluated within this study, such as CTD, PTD, simultaneous prompting, stimulus shaping, and stimulus fading. Therefore, future researchers should continue to evaluate the relative effectiveness of other error correction and errorless learning procedures using group designs.

Fourth, it is likely that within clinical settings, error correction and errorless learning procedures are not used in isolation from one another but are, rather, combined in one teaching package (e.g., flexible prompt fading; Soluaga, Leaf, Taubman, McEachin, & Leaf, 2008). Given that some have discussed that “near-errorless” learning is a more descriptive term for errorless learning (e.g., Fillingham, Hodgson, Sage, & Ralph, 2003), clinicians still need to respond in some way to errors. For instance, error correction could be used in combination with a sufficient frequency of proactive prompting to guard against the occurrence of excessive errors. Future researchers should evaluate the relative effectiveness of various combinations of errorless learning and error correction procedures.

Finally, the methods employed in this study do not permit an evaluation of the possible mechanisms responsible for the outcomes. Participants in the error correction condition could have engaged in more independent correct responses as a result of corrective feedback. It is likely that this corrective feedback functioned as a conditioned punisher. This could have increased the likelihood of correct responses to avoid the corrective feedback. In a sense, the participants engaged in signaled avoidance following contacting the contingencies surrounding incorrect responses. Alternatively, no presumed conditioned punishers were in effect contingent on incorrect responses during the errorless learning condition. Both conditions resulted in the same outcome for correct responses. Therefore, perhaps the combination of reinforcing and punishing contingencies created the ideal context for more independent correct responses. Future studies utilizing single-subject designs will likely be necessary to fully evaluate the mechanisms responsible for the effectiveness of the two procedures evaluated within this study.

Despite these limitations, this study provides one of the few group designs evaluating behavioral approaches to language interventions. Given the importance of group designs outside of the field of behavior analysis (Kazdin, 2011; Smith, 2012; Smith et al., 2007), behavior-analytic researchers should continue to provide research using designs that speak to a larger audience (Smith, 2012; Smith et al., 2007). We hope that this study functions as a motivating operation for other behavior-analytic researchers to design and conduct group research designs evaluating other procedures based on the scientific foundation of behavior analysis. The results of such studies will have many important implications for clinicians and researchers and improve the quality of services available to those in need.

Author Note Justin B. Leaf, Autism Partnership Foundation, Seal Beach, California, and Endicott College Institute for Behavioral Studies; Joseph H. Cihon, Autism Partnership

Foundation, Seal Beach, California, and Endicott College Institute for Behavioral Studies; Julia L. Ferguson, Autism Partnership Foundation, Seal Beach, California; Christine M. Milne, Autism Partnership Foundation, Seal Beach, California, and Endicott College Institute for Behavioral Studies; Ronald Leaf, Autism Partnership Foundation, Seal Beach, California; John McEachin, Autism Partnership Foundation, Seal Beach, California.

Funding No funding was received for this study.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

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

Informed Consent Informed consent was obtained from the parents of all individual participants included in the study.

References

- Abbott, J. H. (2014). The distinction between randomized clinical trials (RCTs) and preliminary feasibility and pilot studies: What they are and are not. *Journal of Orthopaedic & Sports Physical Therapy*, *44*, 555–558. <https://doi.org/10.2519/jospt.2014.0110>
- Barbera, M. L., & Rasmussen, T. (2007). *The verbal behavior approach: How to teach children with autism and related disorders*. London, UK: Kingsley.
- Barlow, D. H., & Hayes, S. C. (1979). Alternating treatments design: One strategy for comparing the effects of two treatments in a single subject. *Journal of Applied Behavior Analysis*, *12*, 199–210. <https://doi.org/10.1901/jaba.1979.12-199>
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston, MA: Houghton Mifflin Company.
- Cariveau, T., La Cruz Montilla, A., Gonzalez, E., & Ball, S. (2018). A review of error correction procedures during instruction for children with developmental disabilities. *Journal of Applied Behavior Analysis*, *3*, 217. <https://doi.org/10.1002/jaba.524>
- Carroll, R. A., Joachim, B. T., St. Peter, C. C., & Robinson, N. (2015). A comparison of error-correction procedures on skill acquisition during discrete-trial instruction. *Journal of Applied Behavior Analysis*, *48*, 257–273. <https://doi.org/10.1002/jaba.205>
- Carroll, R. A., Owsiany, J., & Cheatham, J. M. (2018). Using an abbreviated assessment to identify effective error-correction procedures for individual learners during discrete-trial instruction. *Journal of Applied Behavior Analysis*, *51*, 482–501. <https://doi.org/10.1002/jaba.460>
- Day, H. M. (1987). Comparison of two prompting procedures to facilitate skill acquisition among severely mentally retarded adolescents. *American Journal of Mental Deficiency*, *91*, 366–372.
- Demchak, M. (1990). Response prompting and fading methods: A review. *American Journal of Mental Retardation*, *94*, 603–615.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test (4th ed.; PPVT-IV)*. San Antonio, TX: Pearson. <https://doi.org/10.1037/t15144-000>
- Fentress, G. M., & Lerman, D. C. (2012). A comparison of two prompting procedures for teaching basic skills to children with autism. *Research in Autism Spectrum Disorders*, *6*, 1083–1090. <https://doi.org/10.1016/j.rasd.2012.02.006>.
- 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. <https://doi.org/10.1111/j.1939-0025.1962.tb00267.x>

- Fillingham, J. K., Hodgson, C., Sage, K., & Ralph, L. (2003). The application of errorless learning to aphasic disorders: A review of theory and practice. *Neuropsychological Rehabilitation, 13*, 337–363. <https://doi.org/10.1080/09602010343000020>
- Gast, D. L. (2011a). An experimental approach for selecting a response-prompting strategy for children with developmental disabilities. *Evidence-Based Communication Assessment and Intervention, 5*, 149–238. <https://doi.org/10.1080/17489539.2011.637358>
- Gast, D. L. (2011b). A rejoinder to Leaf: What constitutes efficient, applied, and trial and error. *Evidence-Based Communication Assessment and Intervention, 5*, 234–238. <https://doi.org/10.1080/17489539.2012.689601>
- Green, G. (1996). Early behavioral intervention for autism: What does research tell us? In C. Maurice, G. Green, & S. C. Luce (Eds.), *Behavioral intervention for young children with autism: A manual for parents and professionals* (pp. 29–44). Austin, TX: Pro-Ed.
- Green, G. (2001). Behavior analytic instruction for learners with autism advances in stimulus control technology. *Focus on Autism and Other Developmental Disabilities, 16*, 72–85. <https://doi.org/10.1177/108835760101600203>
- Greer, R. D., & Ross, D. E. (2008). *Verbal behavior analysis: Inducing and expanding complex communication in children with severe language delays*. Boston, MA: Allyn & Bacon.
- Grow, L. L., Kodak, T., & Clements, A. (2017). An evaluation of instructive feedback to teach play behavior to a child with autism spectrum disorder. *Behavior Analysis in Practice, 10*, 313–317. <https://doi.org/10.1007/s40617-016-0153-9>
- Kayser, J. E., Billingsley, F. F., & Neel, R. S. (1986). A comparison of in-context and traditional instructional approaches: Total task, single trial versus backward chaining, multiple trials. *Journal of the Association for Persons With Severe Handicaps, 11*, 28–38. <https://doi.org/10.1177/154079698601100104>
- Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). Oxford, UK: Oxford University Press.
- Kodak, T., Fuchtmann, R., & Paden, A. (2012). A comparison of intraverbal training for children with autism. *Journal of Applied Behavior Analysis, 45*, 155–160. <https://doi.org/10.1901/jaba.2012.45-155>
- LaVigna, G. W., & Donnellan, A. M. (1986). *Alternatives to punishment: Solving behavior problems with non-aversive strategies*. New York, NY: Irvington.
- Leaf, J. B., Alcalay, A., Leaf, J. A., Tsuji, K., Kassardjian, A., Dale, S., et al. (2014a). Comparison of most-to-least to error correction for teaching receptive labeling for two children diagnosed with autism. *Journal of Research in Special Education Needs, 16*, 217–225. <https://doi.org/10.1111/1471-3802.12067>
- Leaf, J. B., Leaf, J. A., Alcalay, A., Dale, S., Kassardjian, A., Tsuji, K., et al. (2014b). Comparison of most-to-least to error correction to teach tacting to two children diagnosed with autism. *Evidence-Based Communication Assessment and Intervention, 7*, 124–133. <https://doi.org/10.1080/17489539.2014.884988>
- 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. <https://doi.org/10.1901/jaba.2010.43-215>
- Leaf, J. B., Townley-Cochran, D., Cihon, J. H., Mitchell, E., Leaf, R., Taubman, M., & McEachin, J. (2019). A descriptive analysis of the use of punishment-based techniques with children diagnosed with autism spectrum disorder. *Education and Training in Autism and Developmental Disorders*. Advance online publication. <https://eric.ed.gov/?id=EJ1217376>
- Ledford, J. R., Gast, D. L., Luscre, D., & Ayres, K. M. (2008). Observational and incidental learning by children with autism during small group instruction. *Journal of Autism and Developmental Disorders, 38*, 86–103. <https://doi.org/10.1007/s10803-007-0363-7>
- Lerman, D. C., & Vomdran, C. M. (2002). On the status of knowledge for using punishment: Implications for treating behavior disorders. *Journal of Applied Behavior Analysis, 35*, 431–464. <https://doi.org/10.1901/jaba.2002.35-431>
- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*, 3–9. <https://doi.org/10.1037/0022-006X.55.1.3>
- Lovaas, O. I. (2003). *Teaching individuals with developmental delays: Basic intervention techniques*. Austin, TX: Pro-Ed.
- Lovaas, O. I., & Simmons, J. Q. (1969). Manipulation of self-destruction in three retarded children. *Journal of Applied Behavior Analysis, 2*, 143–157. <https://doi.org/10.1901/jaba.1969.2-143>
- Majdalany, L., Wilder, D. A., Smeltz, L., & Lipschultz, J. (2016). The effect of brief delays to reinforcement on the acquisition of tacts in children with autism. *Journal of Applied Behavior Analysis, 49*, 411–415. <https://doi.org/10.1002/jaba.282>

- Martin, N. A., & Brownell, R. (2010). *Expressive one-word picture vocabulary test (4th ed.; EOWPVT-4)*. Novato, CA: Academic Therapy Publications.
- Matson, J. L., & Taras, M. E. (1989). A 20 year review of punishment and alternative methods to treat problem behaviors in developmentally delayed persons. *Research in Developmental Disabilities, 10*, 85–104. [https://doi.org/10.1016/0891-4222\(89\)90031-0](https://doi.org/10.1016/0891-4222(89)90031-0).
- McDonnell, J., & Ferguson, B. (1989). A comparison of time delay and decreasing prompt hierarchy strategies in teaching banking skills to students with moderate handicaps. *Journal of Applied Behavior Analysis, 22*, 85–91. <https://doi.org/10.1901/jaba.1989.22.85>.
- McGhan, A. C., & Lerman, D. C. (2013). An assessment of error-correction procedures for learners with autism. *Journal of Applied Behavior Analysis, 46*, 626–639. <https://doi.org/10.1002/jaba.65>.
- Miller, U. C., & Test, D. W. (1989). A comparison of constant time delay and most-to-least prompting in teaching laundry skills to students with moderate retardation. *Education and Training in Mental Retardation, 24*, 363–370.
- Moore, R., & Goldiamond, I. (1964). Errorless establishment of visual discrimination using fading procedures. *Journal of the Experimental Analysis of Behavior, 7*, 269–272. <https://doi.org/10.1901/jeab.1964.7-269>.
- Mueller, M., Palkovic, C. M., & Maynard, C. S. (2007). Errorless learning: Review and practical application for teaching children with pervasive developmental disorders. *Psychology in the Schools, 44*, 691–700. <https://doi.org/10.1002/pits.20258>.
- Petersdottir, A. I., & Carr, J. E. (2011). A review of recommendations for sequencing receptive and expressive language instruction. *Journal of Applied Behavior Analysis, 44*, 859–876. <https://doi.org/10.1901/jaba.2011.44-859>.
- Plaisance, L., Lerman, D. C., Laudont, C., & Wu, W. L. (2016). Inserting mastered targets during error correction when teaching skills to children with autism. *Journal of Applied Behavior Analysis, 49*, 251–264. <https://doi.org/10.1002/jaba.292>.
- Schilmoeller, G. L., Schilmoeller, K. J., Etzel, B. C., & LeBlanc, J. M. (1979). Conditional discrimination after errorless and trial-and-error training. *Journal of the Experimental Analysis of Behavior, 31*, 405–420. <https://doi.org/10.1901/jeab.1979.31-405>.
- Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice Hall.
- Smith, T. (1999). Outcome of early intervention for children with autism. *Clinical Psychology: Science and Practice, 6*, 33–49.
- Smith, T. (2001). Discrete trial training in the treatment of autism. *Focus on Autism and Other Developmental Disabilities, 16*, 86–92. <https://doi.org/10.1177/108835760101600204>.
- Smith, T. (2012). Evolution of research on interventions for individuals with autism spectrum disorder: Implications for behavior analysts. *The Behavior Analyst Today, 35*, 101–113. <https://doi.org/10.1007/BF03392269>.
- Smith, T., Sechill, L., Dawson, G., Guthrie, D., Lord, C., Odom, S., et al. (2007). Designing research studies on psychosocial interventions in autism. *Journal of Autism and Developmental Disorders, 37*, 354–366. <https://doi.org/10.1007/s10803-006-0173-3>.
- Soluaga, D., Leaf, J. B., Taubman, M., McEachin, J., & Leaf, R. (2008). A comparison of flexible prompt fading and constant time delay for five children with autism. *Research in Autism Spectrum Disorders, 2*, 753–765. <https://doi.org/10.1016/j.rasd.2008.03.005>.
- Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland adaptive behavior scales (2nd ed.; Vineland-II)*. Circle Pines, MN: American Guidance Service. <https://doi.org/10.1037/t15164-000>.
- Sulzer-Azaroff, B., & Mayer, R. G. (1991). *Behavior analysis for lasting change*. Fort Worth, TX: Harcourt Brace College Publishers.
- Sundberg, M. L., & Partington, J. W. (1998). *Teaching language to children with autism or other developmental disabilities*. Pleasant Hill, CA: Behavior Analysts.
- Townley-Cochran, D., Leaf, J. B., Leaf, R., Taubman, M., & McEachin, J. (2017). Comparing error correction procedures for children diagnosed with autism. *Education and Training in Autism and Developmental Disabilities, 52*, 91–101.
- Walker, G. (2008). Constant and progressive time delay procedures for teaching children with autism: A literature review. *Journal of Autism and Developmental Disorders, 38*, 261–275. <https://doi.org/10.1007/s10803-007-0390-4>.
- 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. <https://doi.org/10.1901/jaba.2005.115-04>.