The Use of Matrix Training to Teach Color-Shape Tacts to Children with Autism

Sarah E. Frampton¹ · Taylor M. Thompson² · Brittany L. Bartlett² · Bethany Hansen^{2,3} · M. Alice Shillingsburg¹

Published online: 24 September 2018
© Association for Behavior Analysis International 2018

Abstract

Matrix training consists of preplanning instruction by arranging components of desired skills across a minimum of two axes. In the current study, three matrices were developed for each participant (e.g., Matrix 1, Generalization Matrix 1, and Generalization Matrix 2) with known color and shape components. Following baseline, nonoverlapping (i.e., diagonal) training was conducted with Matrix 1. Results of posttests were used to determine the extent of emergence of untrained color-shape combinations across all matrices. Results from all six participants indicated that mastery criteria were eventually met for Matrix 1. For five participants, mastery criteria were also eventually met for generalization matrices. Results replicate findings from prior studies and offer a simple approach for both testing emergence of untrained skills and remediating responding.

Keywords Autism · Matrix training · Recombinative generalization · Tact · Verbal behavior

Considering the degree of language delays exhibited by some individuals with autism spectrum disorder (ASD), interventions that are designed to promote demonstrations of skills beyond the content and context of training are highly desirable. For practitioners serving individuals with ASD, initial

programming tends to focus on developing the most effective and efficient procedures for establishing new responses (e.g., LaFrance & Miguel, 2014; Sundberg & Michael, 2001). This could include teaching basic verbal operants through transfer of stimulus control procedures, refining schedules of reinforcement for teaching and maintaining skills, shaping the articulation of target responses, and so on. Once reliable procedures have been established to train new responses, practitioners can begin to focus on generalization of these responses, including emergence of responses that have not been directly trained and controlled by novel stimuli (e.g., Carr & Miguel, 2013). Although a multitude of resources are available to help practitioners address initial programs (e.g., Leaf & McEachin, 1999; Lovaas, 2002; Sundberg & Partington, 1998), only more recently have resources become available to assist with the latter goals (e.g., Dixon, 2014; Rehfeldt & Barnes-Holmes, 2009). As noted by Axe (2016), three main approaches to producing untrained or generative responding have been highlighted in recent research: equivalence-based instruction, derived relational responding instruction, and matrix training.

Matrix training has been used for decades to establish untrained skills with individuals with disabilities (e.g., Axe & Sainato, 2010; Frampton, Wymer, Hansen, & Shillingsburg, 2016; Goldstein, Angelo, & Moussetis, 1987; Goldstein & Brown, 1989; Goldstein & Moussetis, 1989; Karlan et al., 1982; Kohler & Malott, 2014; Light, Watson, & Remington, 1990; Pauwels, Ahearn, & Cohen, 2015; Remington, Watson, & Light, 1990; Striefel, Wetherby, & Karlan, 1978). The

Research Highlights

- Children with autism may require specific instruction to combine learned words into phrases.
- Matrix training is a useful strategy for programming, as it lends itself to very systematic instruction and analysis.
- Recombinative generalization of known color-shape components was demonstrated with targets from the trained matrix for all participants in the current study.
- Emergence of color-shape tacts for targets in untrained matrices may be due to refined stimulus control over verbal responses.

Brittany Bartlett and Taylor Thompson are now affiliated with Village Autism Center, Marietta, Georgia. Bethany Hansen is now affiliated with University of Nebraska Medical Center Munroe-Meyer Institute, Omaha, Nebraska.

✉ Sarah E. Frampton
sframpton@mayinstitute.org

¹ May Institute, Inc., 41 Pacella Park Drive, Randolph, MA 02368, USA

² Language and Learning Clinic, Marcus Autism Center, Atlanta, GA, USA

³ Emory University School of Medicine, Atlanta, GA, USA

specific procedures used vary along several key dimensions but generally consist of identifying relevant component skills and selecting targets for intervention that are composed of combinations of these skills. The components are displayed across two to three axes, depending on the skills to be targeted. For example, Kohler and Malott (2014) included $3 \times 3 \times 3$ matrices, consisting of subjects, verbs, and objects to produce subject-verb-object tacts (e.g., “Chase spills milk.”). The component skills may be known to the participants (i.e., mastered) prior to training, or they may be unknown. The subjects, verbs, and objects included in the Kohler and Malott (2014) study were mastered by the participants prior to training. However, Pauwels et al. (2015) included objects and prepositions in matrices that were not mastered prior to training.

Matrix training typically begins with training for the *diagonal targets* within one matrix. This type of training is considered nonoverlapping, as each target selected for training is a unique combination of the component skills included in the matrix. However, overlapping training can be employed, which would suggest that across the targets selected for training, some or all of the components are included multiple times. For example, Pauwels et al. (2015) initially trained the four diagonal targets within each matrix and then trained four additional targets, which included one additional exemplar of each component skill. Training may be conducted simultaneously, which suggests that all targets are in training at the same time, or sequentially, which suggests that each target is trained one at a time. Kohler and Malott (2014) applied simultaneous training, whereas Pauwels et al. (2015) used sequential training.

Matrix training is designed to promote efficiency in instruction and ongoing assessment of responding to untrained targets. These untrained targets may not be limited to the *nondiagonal* targets but could also include targets assigned to other matrices. For example, Kohler and Malott (2014) included a total of six matrices and assessed results of intervention with targets within one matrix on responding within other matrices.

Of the applications of matrix training to teach skills to individuals with disabilities, only a few have focused on teaching vocal tact skills to individuals with ASD (Frampton et al., 2016; Kohler & Malott, 2014; Pauwels et al., 2015). Of these studies, only Frampton et al. (2016) and Kohler and Malott (2014) included multiple matrices so that results of intervention could be evaluated beyond the trained matrix. These results are of particular interest, as they allow for broader evaluations of generative responding.

As noted previously, Kohler and Malott (2014) applied matrix training to teach subject-verb-object tacts to two 5-year-old children diagnosed with ASD. For each participant, six matrices were developed consisting of mastered subject, verb, and object names. In baseline, participants were shown

videos depicting different target subjects engaging in target verbs with target objects and asked “What?” to evoke a tact response. Every response combination in all six matrices was evaluated, and when no correct responses were observed, diagonal training (i.e., nonoverlapping training) with Matrix 1 was initiated. The three diagonal targets were simultaneously trained to mastery, and subsequent probes were conducted to evaluate if correct responses occurred for untrained (i.e., nondiagonal) targets within Matrix 1. If this was observed, as it was for one participant, probes were conducted across the remaining untrained matrices. If not observed, as it was for the second participant, additional training with Matrix 1 targets was provided. Following the additional training, probes across untrained matrices were conducted. In order to minimize trial presentations, initially only the diagonal targets of the other matrices were evaluated. If correct responding was obtained for two thirds of the diagonal targets within an untrained matrix, the remaining nondiagonal targets within that matrix were probed. If responding was below mastery levels, diagonal training was conducted with that matrix. The probe-and-train sequence was repeated until untrained responses were obtained across all six matrices, following the logic of a multiple-probe design across stimulus sets (matrices).

Results from Kohler and Malott (2014) indicate that diagonal training within one matrix may be sufficient for generalization to occur across several untrained matrices. For the participant that demonstrated this outcome, varied success was observed across matrices though she had supposedly equal familiarity with the component skills prior to intervention. For this same participant, additional training across matrices led to untrained responding within and across matrices. Overall, matrix training led to mastery of 148 untrained tact responses. For the second participant, responding was less robust, and additional training was required for targets in three of the matrices and varied probe procedures were required in two of the remaining three matrices. Despite the additional training, this approach still led to mastery of 84 untrained tact responses. The procedures used in Kohler and Malott (2014) allowed for a thorough investigation of the effects of matrix training on untrained responses. However, the high number of targets included across the six matrices (162 total tacts) may be a barrier for implementation in a school or clinical setting due to limited time designated for material creation. Additionally, the level of oversight required to collect, monitor, analyze, and interpret these findings may prevent widespread clinical application. Finally, although occurrence of responses in untrained matrices is optimal from a clinical perspective, the experiment employed a multiple-probe design across matrices. Thus, the occurrence of these correct responses in baseline ultimately weakened experimental control.

Frampton et al. (2016) addressed these limitations by employing a multiple-probe design across participants and

reducing the number of matrices evaluated. Mastered nouns and verbs were assigned to two matrices at the beginning of the study, Matrix 1 and the Generalization Matrix. Following baselines of all targets in both matrices, diagonal training was conducted with Matrix 1. Posttests were conducted for Matrix 1 and then the Generalization Matrix. Results indicated that for four of the five participants, this approach was sufficient to produce untrained responses in both matrices. For the fifth participant, high levels of correct responding were observed with targets in Matrix 1 but not in the Generalization Matrix. A new matrix, Matrix 2, was developed for this participant and the same baseline, diagonal training, and posttest procedures were employed. High levels of correct responding were again observed for the targets in Matrix 2, whereas only low levels of correct responding were observed with the Generalization Matrix. A third matrix was developed, and the same procedures employed, this time resulting in high levels of correct responding in both Matrix 3 and the Generalization Matrix.

Taken together, findings from the matrix training literature with children with ASD suggest that considerations related to the number of matrices included and sequence of testing require specific attention. Though the inclusion of too many matrices may be a barrier to implementation from the perspective of a clinician, including too few may also be a barrier. Results from Kohler and Malott (2014) suggest that evidence of untrained responding in one matrix does not necessarily indicate that untrained responding will now occur across all untrained targets. Additionally, results from Frampton et al. (2016) indicate that training across matrices may be necessary as a remedial procedure. However, identifying new targets and conducting pretests could slow the implementation of the procedures overall. Thus, inclusion of one additional matrix from the beginning of the procedures would allow for a broader assessment of untrained responding and a more systematic progression through remedial procedures if they are warranted. Frampton et al. (2016) proposed that the sequence of posttests (i.e., trained matrix before untrained matrix) may also require consideration, as there may be effects on responding that ultimately differentiate matrix training from multiple-exemplar training. The authors noted that following training with three targets (i.e., Matrix 1 diagonal targets), the participants had an opportunity to respond to recombinations of the trained stimuli (i.e., Matrix 1 nondiagonal targets) prior to exposure to completely novel targets (i.e., Generalization Matrix targets). This sequence differs from a multiple-exemplar training preparation in which the presentation of novel targets would immediately follow training. The authors hypothesized that the sequence of the posttests may have been an active variable in their study and suggested that future studies manipulate the order of posttests to illuminate this mechanism.

Thus, the purpose of the current study was to extend results from Frampton et al. (2016) in several ways. First, the current

study applied matrix training to color-shape tacts. Previous research has utilized matrix training to examine color-shape tacts with nonsense words (Esper, 1925; Foss, 1968) and did not include individuals with ASD. Second, the current study incorporated an additional generalization matrix to streamline the introduction of remedial procedures. Third, the current study randomized the order of pre- and posttests during all phases of the study to minimize the potential sequencing effects that may be exerted by Matrix 1 targets.

Method

Participants

These data were collected as part of a 16-week pilot study of children between the ages of 4 years, 0 months, and 5 years, 11 months. Participants were eligible if they primarily mandated (requested) using single words, tacted (labeled) common objects using single words, but exhibited deficits in answering questions. English was required as a primary language spoken in the home. Children without a formal, documented diagnosis of ASD were excluded from the study. Children displaying significant problem behavior that could interfere with the structured intervention were also excluded from participation.

Participants were recruited using a combination of methods, including a review of the language clinic's waitlist, a flier in common areas of the language clinic, and referral from the affiliated assessment team. Seventy-one families were screened by phone and nine met the eligibility criteria. One participant was eligible after the phone screen but did not meet our inclusion criteria during the in-person screen. Two participants withdrew from the study following enrollment due to family events. In total, six participants completed the entire 16 weeks of the study.

All six participants were scheduled for 48 treatment sessions (3 days per week for 16 weeks). All six participants provided a diagnostic report confirming a formal diagnosis of ASD and/or obtained a total raw score of 15 or greater on the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003), which is highly suggestive of ASD. Table 1 displays results from diagnostic measures, as well as total raw scores on the SCQ. Additionally, Table 1 displays results from the MacArthur–Bates Communicative Development Inventory (CDI; Fenson et al., 2007), a questionnaire completed at the start of the study that provided a parent report of their child's language. Participants were randomly assigned by an online application to be in the first, second, or third intervention cohort; two participants were assigned to each cohort. These participants would receive intervention during the first 16 weeks, followed by the next pair, and then the final pair. Cohort 1 participants were exposed to

Table 1 Participant characteristics

Participant	Age (months)	Ethnic Group	SCQ Score	CDI Scores		
				No. of Words	No. of Word Forms	No. of Word Endings
George	54 m	African American	18	148	1	0
Oscar	57 m	African American	9	132	2	1
Carl	62 m	Ethiopian	17	211	3	0
Stefan	54 m	Not reported	16	236	7	4
Tony	64 m	African American	24	274	1	0
Wayne	49 m	Caucasian	23	305	3	0

the fewest baseline probes, followed by Cohort 2, and finally Cohort 3.

Setting and Materials

All sessions took place in a therapy room at a language clinic. The participant and therapist were seated at a table with materials needed for the session (i.e., toys, edible items, and data sheets). Materials used consisted of 7.62 cm × 12.7 cm index cards displaying a colored shape glued onto the center of each card. Each card consisted of a different shape created by the “insert shape” feature (e.g., triangle, square, rectangle) in Microsoft Word. The shapes on each index card were approximately 5.08 cm × 10.16 cm and were filled in with a solid color (e.g., red, yellow, blue) using the fill feature in Microsoft Word. Participants were all exposed to the same color combinations, shown in Table 2. However, target groups were varied in their matrix assignments (i.e., Matrix 1, Generalization Matrix 1, and Generalization Matrix 2) to increase experimental rigor. Thus, the targets assigned to each matrix varied across participants.

Response Measurement and Interobserver Agreement

The primary dependent measure was the percentage of correct color-shape tact responses across all matrices. For a correct response, the participants had to emit the response within 5 s of the vocal question “What is it?” when the therapist presented a card with the targeted shape and color. The participant had to tact both the correct color and shape. Tacts with an incorrect color or shape response, a missing color or shape, or an addition of another irrelevant response or a failure to respond were scored as incorrect.

A second observer collected trial-by-trial reliability data in vivo using paper and pencil. Interobserver agreement was calculated for each session by using the number of agreements between two observers and dividing by the total number of

agreements plus disagreements then multiplying by 100. For George, reliability data were collected on 82.98% of sessions, with a mean agreement of 98.29% (range 88.89%–100%). For Oscar, reliability data were collected on 72% of sessions, with a mean agreement of 98.15% (range 77.78%–100%). For Carl, reliability data were collected on 16.67% of sessions, with a mean agreement of 100%. For Stefan, reliability data were collected on 90% of sessions, with a mean agreement of 98.77% (range 88.89%–100%). For Tony, reliability data were collected for 45.45% of sessions, with a mean agreement of 98.89% (range 88.89%–100%). For Wayne, reliability data were collected for 64.58% of sessions, with a mean agreement of 98.5% (range 73.33%–100%).

Procedural Fidelity

Procedural fidelity was scored by a trained second observer in vivo using paper and pencil. A checklist outlining each step of the procedures was created. A “yes” was recorded if the step was completed as outlined. If any deviations from the procedures were observed, the second observer recorded “no.” The procedural fidelity score was calculated by dividing the number of steps completed with fidelity (scored as a “yes”) by the total number of steps and then converted into a percentage. For George, fidelity data were collected for 44.68% of sessions, with

Table 2 Target groupings

Colors	Shapes
purple	crescent
red	triangle
orange	arrow
blue	pentagon
green	circle
yellow	square
pink	star
white	rectangle
black	heart

a mean score of 99.89% (range 97.78%–100%). For Oscar, fidelity data were collected for 32% of sessions, with a mean score of 100%. For Carl, fidelity data were not collected. For Stefan, fidelity data were collected for 20% of sessions, with a mean score of 100%. For Tony, fidelity data were collected for 25% of sessions, with a mean score of 100%. For Wayne, fidelity data were collected for 41.67% of sessions, with a mean score of 99.72% (range 97.06%–100%).

Target Selection

Colors were arranged on the vertical axis and shapes were arranged on the horizontal axis within 3×3 matrices. Three matrices were created, and their designations were randomly assigned for each participant as either Matrix 1, Generalization Matrix 1, or Generalization Matrix 2. Before intervention began, the colors and shapes were mastered as tacts in isolation. A target was considered mastered when the participant provided correct responses on three consecutive probe trials. To assess the colors, the therapist presented a card entirely composed of a single color and asked, “What is it?” To assess the shapes, the therapist presented a white card with a black outline of a shape and asked, “What is it?” Participants were given 5 s to respond to the vocal questions. Once all colors and shapes met mastery criteria, baseline procedures began.

Design

A nonconcurrent multiple-baseline design across participants (Watson & Workman, 1981) was used to assess the effects of diagonal training on responding to nondiagonal targets and targets in the generalization matrices. The sequence of baselines and posttests was randomized to control for any possible ordering effects. Thus, during baselines and posttests, participants were exposed to sessions for Matrix 1, Generalization Matrix 1, and Generalization Matrix 2 in randomized order. Posttest probes for all three matrices were conducted after diagonal training with Matrix 1. Progression through the next steps of the protocol varied depending upon performance in these posttest sessions. If mastery criteria were met for all three matrices, the protocol was considered complete. If responding was below mastery level for diagonal targets within Matrix 1, booster sessions were conducted for those targets and posttests were repeated.

Baseline

Procedures in baseline were similar to those used in Frampton et al. (2016). Caregivers of each participant reported to the therapist toys and edibles that were typically highly preferred. These items were made available in each session. To begin each experimental session, the therapist identified items to be

used as reinforcers during teaching by asking the participant “What do you want?” or by arranging items and waiting for a response in the form of a reach or a point toward a preferred item. The items in the array varied based on these responses and vocal mands emitted by the participant. If the participant ceased to emit these responses or vocal mands for the available items, new items were selected from the toy shelf and the procedures were repeated.

Prior to a baseline trial, the therapist presented one to two mastered demands to gain momentum and provided praise contingent on correct responses. All mastered demands were unrelated to the matrix targets. Probe trials were conducted following compliance with mastered tasks. For each probe trial, the therapist presented a card with the target color and shape and asked, “What is it?” The participant was given 5 s to respond to the vocal question. Neutral statements (e.g., “okay,” “interesting”) were provided for correct or incorrect responses. The therapist presented two to three baseline trials before interspersing a mastered demand. Praise and access to preferred items were provided following correct responses to mastered demands. Baseline sessions included a total of nine trials, during which the nine targets within the matrix were each presented once. Participants were preassigned, based on their cohort, to a minimum number of sessions to ensure that introduction of training would be staggered across participants. Additional baseline sessions were also conducted if stability was not observed.

Diagonal Training Matrix 1

Procedures used in training were similar to Frampton et al. (2016). Training was conducted with the three diagonal targets in Matrix 1. Each of the three targets was presented three times per session in randomized order for a total of nine trials in a session. For each training trial, the therapist presented a card with the target color and shape and asked, “What is it?” The participant was given 3 s to respond to the vocal question. Correct responding resulted in praise and access to preferred items. If the participant did not respond or provided an incorrect response, the therapist began error correction. Error-correction procedures began with the therapist re-presenting the same discriminative stimuli (e.g., the color-shape card plus vocal question). The therapist immediately provided a vocal prompt (e.g., “What is it? Red triangle.”). If the participant echoed the vocal prompt, the therapist provided another independent opportunity (a transfer trial). Contingent upon correct responding for the transfer trial, the therapist interspersed a mastered demand followed by an additional independent opportunity (a spaced trial). Correct responding on a spaced trial resulted in praise and access to preferred items. Correct responses during the prompted or transfer trials in the error-correction sequence resulted in praise only. The participant was given 3 s to respond to the vocal question during transfer

and spaced trials. If an error occurred at any point during the correction procedure, the therapist re-presented the discriminative stimuli and then provided an immediate vocal prompt and terminated the trial following a correct response to the vocal prompt. Praise and access to preferred items were not provided.

Mastery criteria were two consecutive sessions with a minimum of 89% (eight of nine) correct independent responses. It should be noted that only responses to the first presentation of the discriminative stimuli were counted toward this total. Responses during the error-correction sequence (e.g., transfer and spaced trials) were not included in calculations of mastery.

Posttest

Procedures were identical to baseline. Following mastery of diagonal training with Matrix 1, posttests were conducted in randomized order for Matrix 1, Generalization Matrix 1, and Generalization Matrix 2. Each matrix was probed for a minimum of three sessions. Mastery criterion was 89% (eight of nine) correct responses for two of the three posttest sessions with no evidence of a decreasing trend. If a decreasing trend was detected, additional posttest sessions were conducted until stability in responding was observed. Results of posttests were used to determine which, if any, remedial procedures would be conducted.

Maintenance

A maintenance phase was included 4 weeks after the final posttest session to assess the durability of responding over time. Procedures were identical to the baseline and posttest conditions. No practice was provided related to the color-shape combinations during the interim between the last posttest and the maintenance sessions.

Remedial Procedures: Booster Diagonal Training Matrix 1

All procedures were identical to diagonal training. Participants were required to meet the same mastery criteria before posttests were repeated.

Results

Results for the color-shape matrices are displayed for all participants in Fig. 1. Across participants, few correct responses were observed during baseline. George and Carl emitted no correct responses, Oscar and Wayne each emitted only one correct response, and Tony and Stefan each emitted a total of three correct responses. Diagonal training with Matrix 1 was completed in an average of 3.67 sessions (range 3–7)

across participants. Posttests indicated that mastery criteria were met across all three matrices for Oscar, Stefan, and Carl. For Wayne, George, and Tony, results of posttests showed varied levels of correct responding across matrices, but no participant met mastery criteria for any matrix. Wayne emitted variable correct responses with the diagonal and nondiagonal targets in Matrix 1 but no correct responses for any targets in the generalization matrices. George emitted initially high levels of correct responses for targets in Matrix 1 and Generalization Matrix 1. However, over subsequent probes, correct responding decreased to 0% correct. No correct responses were observed for Generalization Matrix 2. Tony also emitted initially high levels of correct responding for targets in Matrix 1 and Generalization Matrix 1, but these also decreased over subsequent probes. Tony also emitted few correct responses for Generalization Matrix 2.

Booster training for the diagonal targets in Matrix 1 was completed in an average of six sessions (range 4–9) for these three participants. Posttests indicated that mastery criteria were met across all three matrices for George and Tony. Wayne continued to emit variable correct responses with the diagonal and nondiagonal targets in Matrix 1, suggesting the booster training had little effect upon responding for these targets. However, booster training appeared to lead to improved responding for targets in the generalization matrices; moderate levels of correct responding were observed for targets in Generalization Matrix 1, and one correct response was observed for Generalization Matrix 2. Wayne reached the end of his participation period, so no additional remedial procedures were employed. Maintenance probes were completed 4 weeks after the completion of the final posttests with Oscar, Tony, Stefan, and Carl. Responding continued at high levels across matrices across participants.

Discussion

Results of matrix training with the color-shape targets show that diagonal training with Matrix 1 alone was sufficient for high levels of correct responding to be observed with all untrained targets (i.e., the nondiagonal targets in Matrix 1 and in the untrained generalization matrices) for three of the six participants. For two participants, only booster training for the Matrix 1 targets was required before untrained responding was observed. Wayne emitted higher levels of correct responding across all matrices following booster sessions, but not quite at mastery levels. Additional intervention would have been warranted in the form of training for the nondiagonal targets in Matrix 1.

Results from the current study replicate and extend results from Frampton et al. (2016) in several important ways. First, our results highlight the utility of the inclusion of a second generalization matrix. In some cases, this revealed differential

Fig. 1 The percentage of correct color-shape tact responses to targets within all matrices shown across participants

success with targets in one generalization matrix as opposed to another (e.g., George and Wayne). Had only the matrix with higher levels of responding been included, we would have overestimated the success of the intervention and perhaps ended prematurely. The inclusion of a second generalization matrix would have allowed us to immediately begin remedial training with another matrix if warranted. Overall, it seems that the benefits of including an additional generalization matrix outweigh the potential drawbacks, the primary drawback being the time used to conduct additional sessions. However, when a participant was immediately successful across all matrices (e.g., color-shape tacts for Oscar, Carl, and Stefan) and no remedial training was needed, at most nine additional pre- and posttest sessions were conducted. As research has yet to identify definitive predictors of success following matrix training, the inclusion of a second generalization matrix seems to be an efficient and conservative means of programming.

Second, the current study extended results from Frampton et al. (2016) by randomizing the order of the pre- and posttest sessions. The authors posited that the general sequence of

testing Matrix 1 (for which training of diagonal targets was conducted) prior to a generalization matrix (for which no training was conducted) may be an active variable related to the success of matrix training approaches. Specifically, the experimental history created in this sequence includes the opportunity to emit responses in the presence of recombinations of the trained stimuli (i.e., the nondiagonal targets) prior to exposure to novel combinations of stimuli (i.e., the generalization matrix targets). Results from the current study demonstrate that a specific test order sequence is not required for success following matrix training, though results from the present study still suggest it may be useful. Thus, responding with Generalization Matrix 2 targets seemed to improve for George and Wayne following exposure to posttest sessions for Matrix 1 and Generalization Matrix 1. This may be taken as more evidence to support Frampton et al.'s (2016) hypothesis related to the effects of posttest sequencing. Future studies could consider conducting posttests with generalization matrices until stability in responding is observed. If high levels of correct responding are observed, no additional intervention is

warranted and the procedures used would fairly closely resemble multiple-exemplar training. However, if incorrect responses occur with targets assigned to the generalization matrices, the Matrix 1 posttests could be introduced, and once high levels of responding are observed, the posttests for the generalization matrices could be repeated. This would allow for analysis of specific effect of exposure to the Matrix 1 posttests upon responding to targets within the generalization matrices.

In the present study, George, Wayne, and Tony showed differential levels of success between the trained matrices and one or both generalization matrices. After multiple posttest sessions, correct responding across matrices decreased for George and Tony, though maintenance was observed for Matrix 1 targets for Wayne. For George and Tony, the absence of reinforcement following probe trials may be responsible for decrements in responding across sessions. Reinstatement of the reinforcement for responding to the Matrix 1 diagonal targets may have been responsible for the increased correct responding in the second round of posttests for these participants. For Wayne, improvements in correct responding were observed but to a lesser extent. Some alternative approaches to conducting tests for emergence of untrained skills may be considered to potentially avoid this effect on responding, such as providing reinforcement contingent upon correct responses (Ribeiro, Miguel, & Goyos, 2015; Velasco & Tomanari, 2014) or contingent upon all responses (LeBlanc, Miguel, Cummings, Goldsmith, & Carr, 2003).

Upon further analysis of our results and the mechanisms related to the general success of matrix training, we wonder if a clarification in terminology is warranted. Generally, the effectiveness of matrix training has been attributed to *recombinative generalization*. As stated by Goldstein and Moussetis (1989), “Recombinative generalization describes the process of producing or responding to novel utterances; when familiar stimuli are recombined in novel ways” (p. 246). The key word “recombining” suggests that particular stimuli were previously combined and the novel utterances are evoked by new combinations of these same stimuli. Thus, the correct responses to nondiagonal targets demonstrated following training for the diagonal targets in the same matrix seem appropriately attributed to recombinative generalization. However, the correct responses to targets in the generalization matrices cannot be accounted for in the same way, as the stimuli are not recombinations of trained targets.

These responses seem best accounted for by an overall increase in the precision of stimulus-response relations, such that the color portion of the stimulus now evokes the “color” element of the response and the shape portion of the stimulus now evokes the “shape” element of the response. This change in response repertoire overall seems more in line with Palmer’s (2012) description of atomic repertoires, “a set of

fine-grained units of behavior, each under control of a distinctive stimulus, that can be evoked in any permutation by the arrangement of corresponding stimuli” (p. 61). Although distinguishing between the emergence of untrained targets within trained matrices as opposed to untrained matrices on its face appears to have little practical value, we believe that developing a more thorough account of these mechanisms is necessary to support the development of systematized remedial procedures when positive results are not initially observed. Clarifying these distinctions may help practitioners diagnose the exact nature of the faulty stimulus-response relations and then prescribe the correct remedial procedures.

For the current study we created a decision-making model with specific remedial steps matched to different patterns of responding. If diagonal training with Matrix 1 resulted in low levels of responding across matrices, we hypothesized that the lack of contingent reinforcement for correct responding during the posttests was at fault. Thus, additional training for the Matrix 1 diagonal targets was prescribed to make these responses more resistant to extinction. If diagonal training with Matrix 1 resulted in high levels of responding with only the diagonal targets, we hypothesized recombinative generalization had not occurred and overlapping training (i.e., training some nondiagonal targets) would be warranted (e.g., Pauwels et al., 2015). If diagonal training with Matrix 1 resulted in high levels of responding with only the targets in Matrix 1, we hypothesized that recombinative generalization was observed but responding had yet to become “fine-grained” enough for novel responses to occur. Thus, presenting more exemplars and training across matrices was suggested (e.g., Frampton et al., 2016; Kohler & Malott, 2014).

There were several practical considerations that influenced our decision to include three different 3×3 matrices rather than expand or combine existing matrices. We could have conceptualized Matrix 1 (e.g., Colors 1–3 and Shapes 1–3) and Generalization Matrix 1 (e.g., Colors 4–6 and Shapes 4–6) as components of a larger 6×6 matrix. This would allow testing of 18 additional targets (e.g., Color 3 and Shape 4, or Color 5 and Shape 1) without adding new colors and shapes (e.g., Generalization Matrix 2 with Colors 7–9 and Shapes 7–9). First, the burden of material creation would have been substantially increased. In a 6×6 matrix, 36 stimulus cards must be created. With three 3×3 matrices, this number is reduced to 27. Although this seems inconsequential, in a clinical setting with multiple participants these time savings can add up quickly. Second, precise organization is required to ensure that only the very specific color-shape combinations are trained or tested at each particular time. We found that with the inclusion of three distinct matrices, the cards should be organized very easily in three separate groups and easily presented according to our pre-session randomization on our data sheets. As a conceptual consideration, we also hypothesized that if correct responding was observed with targets in the

generalization matrices, then correct responding for all combinations of mastered colors and shapes would now occur. Essentially, we assumed that our three different matrices ultimately formed a larger 9×9 matrix, and that correct responses to any one of those color-shape combinations should now occur. We did not test this assumption, as we noted the practical considerations of the creation of 81 stimulus cards per participant and precise sequencing of instruction were not suitable for a clinical setting. However, future research may examine these assumptions and determine whether there are more efficient ways of arranging instruction.

Some limitations of our findings should be noted. Wayne was not able to complete all recommended treatment sessions due to the time limitations on his participation. Thus, the actual utility of the prescribed remedial procedures was not tested to completion. It is our hope that future studies will follow the proposed model of decision making and provide a fuller analysis of results. Additionally, the current study employed a nonconcurrent design across participants, as our participants received intervention at different time points. This limitation is minimized, as participants were randomly assigned a time point at the outset of the study. Despite these limitations, our study replicates findings from prior matrix studies (e.g., Frampton et al., 2016; Kohler & Malott, 2014; Pauwels et al., 2015) and includes a proposed method of synthesizing these findings into a model that can be applied on a broader scale.

On a theoretical note, results from the current study and others in this line of research may also be interpreted using relational frame theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001). Generally, RFT analyzes verbal behavior as generalized operant classes of behavior. RFT accounts of language and grammar theorize that the patterns of verbal behavior labeled as “grammar” by the general community consist of generalized repertoires of relational verbal operants that come under the antecedent control of stimuli, due to a history of multiple-exemplar training in the presence of those stimuli. Such antecedent stimuli are referred to as “contextual cues” in the RFT literature because, by definition, they set the occasion for untrained or generative responding, as opposed to discriminative stimuli, which control behavior through a history of direct training (Barnes-Holmes, Barnes-Holmes, & Cullinan, 2000). In the case of noun-verb or adjective-noun utterances, Barnes-Holmes et al. (2000) proposed that generalized repertoires of tacting objects (e.g., “nouns”) and generalized repertoires of tacting other physical properties (e.g., color “adjectives”), both of which are frames of coordination, may come to be controlled simultaneously by the same antecedent contextual cue. In the case of the current study, training across exemplars in the presence of the contextual cue (i.e., “What is it?” and a color-shape stimulus) could establish “What is it?” to simultaneously cue the untrained coordinative relational responding of tacting

the shape and color of the nonverbal stimulus. Put simply, a history of multiple-exemplar training for both shape-tacting and color-tacting operants in the presence of “What is it?” makes both generalized operants likely, even with novel stimuli, in the presence of the instruction “What is it?” However, in the present study, we observed correct responding in the presence of stimuli from some matrices (i.e., Matrix 1 and Generalization Matrix 1) but not in the presence of others (i.e., Generalization Matrix 2) for some participants. From an RFT perspective, this result suggests that the function of “What is it?” to control both tacting repertoires was not fully established, at least for those participants for whom generalization was challenging.

Skinner’s (1957) account of grammar and language as a type of autoclitic frame with multiple variables affecting the emission of a response is another conceptual perspective from which to consider the current data. In Chapter 13 of *Verbal Behavior*, Skinner noted that particular patterns of responding to stimuli in multiple ways (e.g., tacting both object and color properties at the same time) get reinforced in the past, so that now, when confronted with a stimulus for which one has only been trained to tact in one way (e.g., only object or color), one might be more likely to tact both properties, compared to a stimulus with which one has no history. This could mean that color-shape stimuli that were included in the training context (i.e., Matrix 1 stimuli) may be more likely to evoke the intraverbal pattern, or autoclitic frame. For example, if “red square” was trained as a diagonal target, the emission of the response “red” may be likely to evoke a subsequent shape response (i.e., “circle”) when presented with untrained, nondiagonal targets from the same matrix. In contrast, color-shape stimuli that have not been presented in a training context (i.e., generalization matrices) may be less likely to evoke this intraverbal pattern. For example, the response “black” had never been explicitly trained in sequence with a “shape” response. If error patterns indicate that the participants continue to tact the shape (rather than color) in posttests, this could account for differences in performance across generalization matrices. If the shape tact is emitted first (i.e., “star”), there is no particular training history to evoke an additional color response. Alternatively, if the color tact is evoked first, this response may exert some additional intraverbal control over the emission of the shape tact. These hypotheses are highly speculative and would benefit from error-pattern analysis in future studies. For example, do the participants err by failing to emit any responses, by emitting incorrect responses, or by emitting incomplete responses (only color or only shape)? Results from the present study cannot clarify these conceptual questions, though we hope findings from our study will be useful to spur future research and increase the precision and efficiency of teaching for individuals with ASD.

Acknowledgements The study was conducted at the Marcus Autism Center and Emory University and was supported by a grant from the Marcus Core Pilot Funding Award.

Compliance with Ethical Standards

Conflict of Interest The authors have no declared financial conflicts of interest.

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

Informed Consent Informed consent was obtained for all individuals in the study.

References

- Axe, J. B. (2016). Combining concepts from verbal behavior and derived relational responding produces efficient language instruction for children with autism. *Evidence-Based Communication Assessment and Intervention*, 9(3), 106–112. <https://doi.org/10.1080/17489539.2016.1153813>
- Axe, J. B., & Sainato, D. M. (2010). Matrix training of preliterate skills with preschoolers with autism. *Journal of Applied Behavior Analysis*, 43, 635–652. <https://doi.org/10.1901/jaba.2010.43-635>
- Barnes-Holmes, D., Barnes-Holmes, Y., & Cullinan, V. (2000). Relational frame theory and Skinner's *Verbal behavior*: A possible synthesis. *The Behavior Analyst*, 23(1), 69–84. <https://doi.org/10.1007/bf03392000>
- Carr, J. E., & Miguel, C. F. (2013). The analysis of verbal behavior and its therapeutic applications. In G. J. Madden (Ed.), *APA handbook of behavior analysis* (Vol. 2). Washington, DC: American Psychological Association.
- Dixon, M. R. (2014). *The PEAK relational training system: Direct training module*. Carbondale, IL: Shawnee Scientific Press.
- Esper, E. A. (1925). A technique for the experimental investigation of associative interference in artificial linguistic material. *Language Monographs*, 1, 5–46.
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-bates communicative development inventories: User's guide and technical manual* (2nd ed.). Baltimore, MD: Brookes Publishing.
- Foss, D. J. (1968). An analysis of learning in a miniature linguistic system. *Journal of Experimental Psychology*, 76, 450–459. <https://doi.org/10.1037/h0025506>
- Frampton, S. E., Wymer, S. C., Hansen, B., & Shillingsburg, M. A. (2016). The use of matrix training to promote generative language with children with autism. *Journal of Applied Behavior Analysis*, 49(4), 869–883. <https://doi.org/10.1002/jaba.340>
- Goldstein, H., Angelo, D., & Moussetis, L. (1987). Acquisition and extension of syntactic repertoires by severely mentally retarded youth. *Research in Developmental Disabilities*, 8, 549–574. [https://doi.org/10.1016/0891-4222\(87\)90054-0](https://doi.org/10.1016/0891-4222(87)90054-0)
- Goldstein, H., & Brown, W. H. (1989). Observational learning of receptive and expressive language by handicapped preschool children. *Education and Treatment of Children*, 12, 5–37.
- Goldstein, H., & Moussetis, L. (1989). Generalized language learning by children with severe mental retardation: Effects of peers' expressive modeling. *Journal of Applied Behavior Analysis*, 22, 245–259. <https://doi.org/10.1901/jaba.1989.22-245>
- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York, NY: Plenum Press.
- Karlan, G. R., Brenn-White, B., Lentz, A., Hodur, P., Egger, D., & Frankoff, D. (1982). Establishing generalized, productive verb-noun phrase usage in a manual language system with moderately handicapped children. *The Journal of Speech and Hearing Disorders*, 47, 31–42. <https://doi.org/10.1044/jshd.4701.31>
- Kohler, K. T., & Malott, R. W. (2014). Matrix training and verbal generativity in children with autism. *The Analysis of Verbal Behavior*, 30, 170–177. <https://doi.org/10.1007/s40616-014-0016-9>
- LaFrance, D. L., & Miguel, C. F. (2014). Teaching language to children with autism spectrum disorder. In P. Sturmey, J. Tarbox, D. R. Dixon, & J. L. Matson (Eds.), *Handbook of early intervention for autism spectrum disorders: Research, practice, and policy* (pp. 403–436). New York, NY: Springer.
- Leaf, R., & McEachin, J. (1999). *A work in progress: Behavior management strategies and a curriculum for intensive behavioral treatment of autism*. New York, NY: DRL Books.
- Leblanc, L. A., Miguel, C. F., Cummings, A. R., Goldsmith, T. R., & Carr, A. E. (2003). The effects of three stimulus-equivalence testing conditions on emergent US geography relations of children diagnosed with autism. *Behavioral Interventions*, 18(4), 279–289. <https://doi.org/10.1002/bin.144>
- Light, P., Watson, J., & Remington, B. (1990). Beyond the single sign: The significance of sign order in a matrix-based approach to teaching productive sign combinations. *Mental Handicap Research*, 3, 161–178. <https://doi.org/10.1111/j.1468-3148.1990.tb00034.x>
- Lovaas, O. I. (2002). *Teaching individuals with developmental delays: Basic intervention techniques*. Austin, TX: Pro-Ed.
- Palmer, D. C. (2012). The role of atomic repertoires in complex behavior. *The Behavior Analyst*, 35(1), 59–73.
- Pauwels, A. A., Ahearn, W. H., & Cohen, S. J. (2015). Recombinative generalization of tacts through matrix training with individuals with autism spectrum disorder. *The Analysis of Verbal Behavior*, 31, 200–214. <https://doi.org/10.1007/s40616-015-0038-y>
- Rehfeldt, R. A., & Barnes-Holmes, Y. (2009). *Derived relational responding applications for learners with autism and other developmental disabilities: A progressive guide to change*. Oakland, CA: New Harbinger.
- Remington, B., Watson, J., & Light, P. (1990). Beyond the single sign: A matrix-based approach to teaching productive sign combinations. *Mental Handicap Research*, 3, 33–50. <https://doi.org/10.1111/j.1468-3148.1990.tb00079.x>
- Ribeiro, D. M., Miguel, C. F., & Goyos, C. (2015). The effects of listener training on discriminative control by elements of compound stimuli in children with disabilities. *Journal of the Experimental Analysis of Behavior*, 104(1), 48–62. <https://doi.org/10.1002/jeab.161>
- Rutter, M., Bailey, A., & Lord, C. (2003). *SCQ: The social communication questionnaire*. Torrance, CA: Western Psychological Services.
- Skinner, B. F. (1957). *Verbal behavior*. New York, NY: Appleton-Century-Crofts.
- Striefel, S., Wetherby, B., & Karlan, G. R. (1978). Developing generalized instruction-following behavior in severely retarded people. In C. E. Meyers (Ed.), *Quality of life in severely and profoundly mentally retarded people: Research foundations for improvement* (pp. 267–326). Washington, DC: American Association on Mental Deficiency.

- Sundberg, M. L., & Michael, J. (2001). The value of Skinner's analysis of verbal behavior for teaching children with autism. *Behavior Modification*, 25, 698–724.
- Sundberg, M. L., & Partington, J. W. (1998). *Teaching language to children with autism or other developmental disabilities*. Pleasant Hills, CA: Behavior Analysts.
- Velasco, S. M., & Tomanari, G. Y. (2014). Assessing derived conditional relations under reinforcement conditions. *The Psychological Record*, 64(3), 551–558. <https://doi.org/10.1007/s40732-014-0056-5>
- Watson, P. J., & Workman, E. A. (1981). The non-concurrent multiple baseline across-individuals design: An extension of the traditional multiple baseline design. *Journal of Behavior Therapy and Experimental Psychiatry*, 12(3), 257–259. [https://doi.org/10.1016/0005-7916\(81\)90055-0](https://doi.org/10.1016/0005-7916(81)90055-0)