Promoting Generalized Advanced Language Skills of Children in Intensive Behavioral Intervention with Promoting the Emergence of Advanced Knowledge Generalization Module (PEAK–G)

Sarah M. Dunkel-Jackson^{1,2} · Mark R. Dixon³

Published online: 26 January 2018

© Association for Behavior Analysis International 2018

Abstract

The Promoting the Emergence of Advanced Knowledge Relational Training System Generalization Module (PEAK–G; Dixon, 2014) was used within a community-based intensive behavioral intervention (IBI) program to enhance the generalized advanced language skills of children diagnosed with autism spectrum disorder. Four multiple-baseline, across-behaviors designs were used to assess the effectiveness of the PEAK–G “train and test” discrete trial training (DTT) method on directly trained and generalized language skills. After implementing PEAK–G, directly trained language skills increased for all participants, and generalization to test stimuli was observed for two participants, with some generalization observed for a third participant. These data suggest that PEAK–G may be a viable option for community-based practitioners to promote more advanced generalized language skills to children who have mastered early language skills.

Keywords Generalization · Autism · Train and test · Verbal behavior

- The typically enhanced discrimination skills of children with autism spectrum disorder (ASD) can lead to obvious strengths, but overselectivity to stimuli may hinder generalization of skills.
- PEAK–G promotes directly trained and generalized advanced language skills based on Skinner’s (1957) analysis of verbal behavior.
- The PEAK–G train and test DTT method provides important information about generalization of skills across a variety of stimuli targets.
- PEAK–G may be a viable option for community-based IBI and applied behavior analysis (ABA) practitioners.

Approximately 1 out of every 68 children in the United States is diagnosed with autism spectrum disorder (ASD; Centers for Disease Control and Prevention [CDC], 2014), a developmental disability that affects social communication and social interaction and is characterized by restricted and repetitive patterns of behavior (American Psychiatric Association, 2013). These behavioral excesses and skill deficits can be improved through effective interventions based on the principles of applied behavior analysis (ABA) that have proven to be the most useful for developing skills of children diagnosed with ASD (e.g., Granpeesheh, Dixon, Tarbox, Kaplan, & Wilke, 2009; Leaf, Taubman, McEachin, Leaf, & Tsuji, 2011; Lovaas, 1987; McEachin, Smith, & Lovaas, 1993; National Autism Center, 2009; Sallows & Graupner, 2005). Large-scale, community-based intensive behavioral intervention (IBI) programs for children with ASD have also demonstrated effectiveness and integrity. Overall, outcome studies suggest that children who receive community-based IBI make significant improvements in cognitive skills, significant reductions in autism symptoms, and improvements in adaptive functioning (e.g., Blacklock, Perry, & Dunn Geier, 2014; Freeman & Perry, 2010; Perry et al., 2008a; Perry et al., 2008b).

ABA and IBI programs have certainly demonstrated great success with significantly increasing the skills and learning

✉ Sarah M. Dunkel-Jackson
sdunkel@siu.edu

¹ Rehabilitation Institute, Southern Illinois University and Kinark Child and Family Services, Markham, ON, Canada

² Present address: Behavioural Psychology Program, Seneca College, King Campus, 13990 Dufferin Street, King City, ON L7B 1B3, Canada

³ Rehabilitation Institute, Southern Illinois University, Carbondale, IL, USA

trajectories of children diagnosed with ASD. However, these programs are not without critiques—especially those highlighting the limitations of preventing rote responding in lieu of producing generalized repertoires. In the seminal article on ABA, Baer, Wolf, and Risley (1968) emphasized that “generality” of behavior change was a “valuable characteristic of applied behavior analysis” (p. 96). Generalization is said to occur when specific behaviors are emitted under different, untrained conditions (e.g., across other people, places, behaviors, and time) without programmed contingencies for those untrained behaviors (Cooper, Heron, & Heward, 2007; Stokes & Baer, 1977). Response generalization occurs when behaviors are reinforced in the presence of specific stimuli and later, functionally equivalent, and topographically similar behaviors are emitted in the presence of those same stimuli. Stimulus generalization occurs when behaviors that have been reinforced in the presence of specific stimuli now occur in the presence of other, similar stimuli under extinction conditions.

Generalization is an especially critical focus when developing interventions for children with ASD because the goal is to teach basic skills, including generalized repertoires (e.g., generalized imitation), so that learning may continue under natural, untrained situations (e.g., typical classroom, employment, new relationships). The typically enhanced discrimination skills of children with ASD can lead to obvious strengths (e.g., visual search tasks), but overselectivity to stimuli (e.g., Lovaas, Koegel, & Schreibman, 1979) may actually hinder generalization of skills to other learning situations (Brown & Bebko, 2012). Froelich et al. (2012) suggested that individuals with ASD may have difficulties with generalization of skills to novel stimuli. For example, children with ASD may require additional training (Betz, Higbee, & Pollard, 2010) or more complex, higher order conditional discrimination training (Okuyama & Isawa, 2010) before responding generalizes to the natural environment. In a comprehensive review of literature on “overselectivity,” Matthews (1994) found that in 18 out of 20 studies, participants diagnosed with ASD were found to meet criteria for overselectivity much more frequently than participants in other groups. Furthermore, children who demonstrated overselective stimulus control included those with the highest scores on the Autism Diagnostic Observation Schedule (ADOS; Dickson, Wang, Lombard, & Dube, 2006) and a mental age below 6 years (Bailey, 1981; Lovaas et al., 1979; Wilhelm & Lovaas, 1976). Interestingly, the children who often meet eligibility criteria for community-based IBI programs include those children with moderate to severe autism symptoms and lower scores on various cognitive assessments—those who are most likely to require additional training to promote generalization.

Fortunately, many applied researchers and practitioners who work with children diagnosed with ASD have adequately assessed and programmed for generalization (see Stokes &

Baer, 1977; Stokes & Osnes, 1989). For example, by teaching skills under one set of stimulus conditions (e.g., IBI setting, Material Set A) and later testing skills under another set of stimulus conditions (e.g., classroom, Material Set B) without providing prompts and/or programmed reinforcement (i.e., extinction), practitioners can identify the extent to which teaching has resulted in generalized outcomes of new responses to similar stimuli or similar responses to new stimuli (see Cuvo, 2003; Guttman & Kalish, 1956; Skinner, 1953). If generalization does not occur, stimulus control analyses (e.g., Rincover & Koegel, 1975) and further interventions (e.g., Bass & Mulick, 2007; Day-Watkins, Murray, & Connell, 2014; Handleman & Harris, 1980; Lovett & Rehfeldt, 2014; Marzullo-Kerth, Reeve, Reeve, & Townsend, 2011; Moes & Frea, 2002; Persicke, Tarbox, Ranick, & St. Clair, 2013; Stokes & Baer, 1977; Stokes & Osnes, 1989) are conducted to enhance the probability of generalization.

There is no doubt that analyzing stimulus control issues (e.g., generalization gradients) and promoting response generalization and stimulus generalization (e.g., Stokes & Baer, 1977; Stokes & Osnes, 1989) are important when implementing behavior change procedures. However, it is much more vital to take these steps when providing services to those populations that demonstrate characteristic challenges with generalization—namely, children diagnosed with ASD who present with a profile correlated with generalization issues (e.g., high scores on the ADOS, low scores on cognitive assessments) and are therefore eligible to receive intensive services. What is needed for these children is a comprehensive and efficacious assessment and intervention program that not only teaches specific skills but also generates additional responses to the child’s world without direct teaching.

Gould, Dixon, Najdowski, Smith, and Tarbox (2011) reported that an early intensive behavioral intervention curriculum should (a) be comprehensive and address the major domains of human functioning, (b) target early childhood development starting at age 6 months through 8 years, (c) consider behavior function as opposed to topography, (d) provide a direct link from assessment to specific intervention, and (e) track progress over time. Gould and colleagues reviewed several commonly used resources and reported that only four assessments included the critical components to aid in the development of interventions. The Vineland Adaptive Behavior Scale (VABS-II; Sparrow, Cicchetti, & Balla, 2005) and Verbal Behavior Milestones Assessment and Placement Program (VB-MAPP; Sundberg, 2008) were among the four assessments that met Gould and colleagues’ requirements and, interestingly, are among the most commonly used assessments for children with ASD receiving community-based IBI.

After the Gould et al. (2011) early intensive behavioral intervention (EIBI) assessment review, another curriculum assessment was published that may include these “critical

components.” The Promoting the Emergence of Advanced Knowledge (PEAK) Relational Training System (Dixon, 2014, 2015, 2017) is a comprehensive, functional tracking and assessment system that guides practitioners to specific curriculum targets of early childhood skills.

Peak

PEAK (Dixon, 2014, 2015, 2017) is a new criterion-referenced curriculum assessment and intervention guide system based on a contemporary behavior-analytic approach. PEAK aims to teach basic and advanced language skills by promoting generativity and relational responding for more abstract and complex language use. PEAK includes four modules: Direct Training (PEAK-DT; Dixon, 2014), Generalization (PEAK-G; Dixon, 2014), Equivalence (PEAK-E; Dixon, 2015), and Transformation (PEAK-T; Dixon, 2017). Each PEAK module includes a performance matrix, a program assessment, a discrete trial training (DTT) flowchart, a data sheet, a program list, 185 program instruction sheets (184 sample programs, one blank program), and a PEAK report card. The program assessment includes a checklist of 184 skills used to indirectly and/or directly assess the current repertoire of a child and help identify programs for instruction. The Performance Matrix Triangle allows intervention teams to visually organize and analyze results of the program assessment by coloring in cells that pertain to mastered skills across assessment periods. The DTT flowchart and accompanying instructional pages outline the process of conducting instructional trials, starting from preparing materials, presenting discriminative stimuli, and praising (correct responses) and prompting (incorrect responses) according to the PEAK error correction procedures. The data sheet includes five trial blocks of 10 trials, each with a Notes section. The program list can be placed in a client’s three-ring binder for quick reference and indexing of programs. One program instruction sheet is available for each of the skills outlined in the program assessment and performance matrix and includes a goal, materials, instructions (i.e., discriminative stimulus [SD]), typical stimuli, space to enter a target list, dates of introduction and mastery, and the prompt hierarchy and scoring procedures. Each PEAK manual also provides a written introduction to PEAK; instructions for completing each of the PEAK components; and tips for setting up the learning environment, managing the program book, selecting programs for instruction, and making adjustments to the programs.

PEAK-G The PEAK-G module (Dixon, 2014) expands upon the direct training methods of PEAK-DT by “assessing the emergence of generalized responding” (Dixon et al., 2015, p. 225) and by providing a curriculum that is designed to promote the emergence of a flexible and generalized language

repertoire (Dixon, 2014). Within a single program, instructors directly train only some responses and simply test other responses using the “train and test” strategy, therefore assessing whether children can demonstrate generalized responding to untrained stimuli. PEAK-G provides a means of assessing and subsequently programming for generalized repertoires of motor imitation; vocal imitation; categorization; simple and complex mathematics; and complex, language-based problem solving, among many other skill areas and verbal operants (Dixon, 2014; Dixon et al., 2015; Skinner, 1957).

Clinicians and educators with a basic understanding of linguistic structures, behavior analysis (and its application), and Skinner’s (1957) verbal operants conduct the PEAK-G program assessment through a combination of caregiver report, records review, observations, and direct testing. As noted by Dixon (2014), a PEAK manual is not a substitute for good clinical programming and expert consultation.

Peer-Reviewed Support for PEAK There are several peer-reviewed, published articles examining the effectiveness and psychometric properties of PEAK-DT, PEAK-G, PEAK-E, and PEAK-T (e.g., Belisle, Dixon, Stanley, Munoz, & Daar, 2016; Dixon, Belisle, Whiting, & Rowsey, 2014a; Dixon et al., 2014b; Dixon, Whiting, Rowsey, & Belisle, 2014c; McKeel, Dixon, Daar, Rowsey, & Szekely, 2015a; McKeel, Rowsey, Belisle, Dixon, & Szekely, 2015b; McKeel, Rowsey, Dixon, & Daar, 2014; Rowsey, Belisle, & Dixon, 2014). PEAK has demonstrated reliability (e.g., PEAK-DT: Dixon, Stanley, Belisle, & Rowsey, 2016), internal validity (e.g., PEAK-DT: Rowsey et al., 2014; PEAK-G: Dixon et al., 2017b), and external validity (PEAK-DT: Dixon et al., 2014b; Dixon et al., 2014b; Malkin, Dixon, Speelman, & Luke, 2017; McKeel et al., 2014; PEAK-G: Dixon et al., 2015; Dunkel-Jackson, 2016). Further research from Dixon and colleagues has explored the relationship between age and PEAK scores of both normative and ASD samples for PEAK-DT (Dixon et al., 2014a) and PEAK-G (Dixon et al., 2017b). The application of the PEAK Relational Training System has also been assessed via randomized control trial (McKeel et al., 2015a) and several single-subject designs (e.g., Belisle et al., 2016; Daar, Negrelli, & Dixon, 2015; Dixon et al., 2015).

The effectiveness of the PEAK-G module, specifically, was examined by Dixon, Peach, Daar, and Penrod (2017a), whereby researchers enhanced the generalization of autoclitic mands, distorted tacts, and creative path finding by three 4- and 5-year-old children diagnosed with ASD participating in an early childhood education program in a public school. Dixon, Peach, et al. highlighted that the potential utility of the train and test methods of PEAK-G is intriguing and that these methods require further research to determine their utility in clinical and educational settings. Dixon, Peach, et al. noted that the targeted train and test stimuli varied only

slightly in terms of stimulus or response generalization, suggesting that future research across more of the generalization gradient may provide additional support for the use of PEAK–G. Furthermore, maintenance of verbal behavior skills across time was not assessed in the study by Dixon, Peach, et al. Therefore, the purpose of the current study was to further explore the utility of PEAK–G to enhance the generalization of advanced verbal operants for four children diagnosed with ASD participating in a community-based IBI program.

Method

Participants

Participants included four 6- to 7-year-old children (Doyle, Derek, Eric, and JJ) diagnosed with ASD receiving community-based IBI from a regional service provider in Ontario, Canada. Participant demographics, including adaptive behavior scores, are provided in Table 1. All learners received primarily small-group instruction with some one-on-one instruction. Each learner was due to be discharged from IBI within 4 to 9 months with transition plans that included receiving consultative Connections for Students (CFS) services to help guide a successful transition to a mainstream or small-class special education placement in his public or Catholic school district. Derek, Eric, and JJ received 24+ hr of IBI across 4 to 5 days per week, and Doyle received 18 h of service across 3 days per week, consistent with his transition plan.

In general, all learners required intervention for generalized advanced language and social skills. All learners communicated vocally and had mastered early language skills, including echoic, mand, tact, and intraverbal skills under contrived and naturally occurring instructional situations controlled by motivating operations (e.g., completed Level 1, Level 2, and most of Level 3 on the VB–MAPP Milestones Assessment). Derek, Eric, and JJ were below and Doyle was near grade level on the Ontario Academic School Curriculum.

Each participant exhibited challenging behaviors (e.g., vocal protesting, flopping to the floor, playing with and/or

destroying materials) of mild intensity that were sensitive to escape contingencies (e.g., during new, challenging tasks) and were managed by general behavior management strategies not requiring a formal behavior intervention plan (BIP). Toward the end of the study, Eric required a formal BIP for aggression toward peers (e.g., lightly hitting peer), which was maintained by escape from aversive (e.g., “annoying”) peer behavior. Consistent with recommendations from the PEAK–G manual (Dixon, 2014), participants were chosen because they regularly responded to conditioned and generalized reinforcement in the form of praise and tokens, with JJ receiving more frequent primary reinforcers during learning trials of high response effort. All learners in the current study also participated in a differential reinforcement (of other behavior [DRO] or of alternative behavior [DRA]) “goals program,” whereby individualized goals (e.g., “I kept my hands to myself,” “I listened to my teacher”) were reviewed prior to the session and “checked” by both the learner and the instructor when a timer indicated that an individualized goal duration had passed (i.e., 5 to 20 min of instructional time). The learner gained access to a chosen backup reinforcer (e.g., play time, Nintendo Wii, Nintendo DS) for a duration based on the number of goals met (e.g., meeting three goals led to 3 min of access to the reinforcer).

Setting and Materials

All sessions were conducted in the IBI setting, a 6.7 m × 10.7 m space within a public school with cubicles for one-on-one work, tables and chairs for small-group instruction, a staff desk area, and a computer area. DTT sessions were conducted in the cubicles and small-group instruction areas with relevant PEAK–G programs, stimuli, preferred items, and PEAK–G data sheets. To summarize, graph, and analyze PEAK data, the IBI team used the PEAK Electronic Data System (PEAK–EDS), a standardized Microsoft Excel workbook (i.e., electronic client binder) created by the first author based on suggestions from the PEAK manuals (e.g., Dixon, 2014) with embedded prompts (i.e., comments, colored text) for accurate completion of specific elements (e.g., worksheet for performance matrix, program assessment, data summary, figure).

Table 1 Participant demographics for experiment 3

Participant	Age (year, months)	Gender	IBI setting	Challenging behavior	VABS–II ABC (C, DL, S subscale scores)
Derek	7, 6	Male	Center/CFS	Vocal protesting	82 (82, 85, 85)
Doyle	6, 3	Male	Center/CFS	Vocal protesting	95 (108, 101, 97)
Eric	7, 4	Male	Center/CFS	Vocal protesting, aggression to staff and peers	74 (74, 76, 78)
JJ	7, 3	Male	Center/CFS	Vocal protesting, flopping, property destruction	70 (70, 65, 78)

IBI intensive behavioral intervention *VABS–II* Vineland Adaptive Behavior Scale, *ABC* adaptive behavior composite, *C* communication, *DL* daily living, *S* socialization, *CFS* Connections for Students

Experimental Design

A concurrent, multiple-baseline design across programs was used for each participant. To minimize the adverse effects of having no intervention in place due to the requirement of lengthy baselines, an embedded multiple-probe design was used for Doyle.

Procedure

Program Assessment The PEAK–G program assessment is a criterion-referenced checklist of 184 generalization skills that correspond to 184 PEAK–G programs. To conduct the program assessment, the author met with the supervisor and/or the staff assigned as each learner’s primary instructor. Staff indicated a “Y” for skills that the learner had engaged in previously but was not directly taught and an “N” for skills that the learner had never engaged in or was only directly taught. Staff indicated “?” for skills that they were unsure about. For skills without a “Y” or “N,” staff reviewed other sources of information (e.g., report cards, individualized education program [IEP] goals, previous curriculum plans) and/or directly tested the skills with the learner. During direct assessments of PEAK–G skills, staff systematically presented an array of 10 discrete trials per skill (Dixon, 2014) and provided no programmed consequences for correct or incorrect responses but did provide frequent breaks (see Dixon et al., 2015) and/or reinforcement for interspersed mastered skills. For skills with multiple responses (e.g., staff “?” and direct test “Y”), the clinical team reviewed the information and determined which response most correctly matched the learner’s language repertoire, and this was used as the final assessment score.

Performance Matrix The PEAK performance matrix allows the intervention team to visually organize and analyze the results of the program assessment. Staff in the current study entered program assessment data into the PEAK–EDS and, by clicking a macro-enabled button, filled the performance matrix across each assessment period.

Program Selection A corresponding PEAK–G program is available for each of the 184 items on the program assessment. The IBI team first met with the learner’s family to identify potential skill domains (e.g., communication, social skills, school skills) and specific targets (e.g., working in groups, not being bullied). The IBI team then reviewed the PEAK performance matrix and program assessment for each learner and selected three to four specific programs, as outlined in the following sections. Stimuli lists for each program were collaboratively developed with the intervention team. Levels for each set of stimuli were established to modify response effort and stimuli dissimilarity along the generalization gradient based on each learner’s profile and previous data (example

stimuli used and descriptions of the types of modifications along the gradient are available from the first author).

Intraverbal Substitution by Function—2A. The IBI team selected 2A for Eric and JJ to increase their identification of multiple uses of objects. Staff would ask, “What can you do with ____?” (e.g., a pencil) and after a correct response (e.g., “write with it”), staff would ask, “What else can you do with ____?” Eric and JJ were expected to respond intraverbally with an additional, nonrote function of the object (e.g., “play drums”). Train and test stimuli pairs were identified based on their formal and functional similarity (e.g., pen and pencil, golf club and hockey stick).

Flexible Textual Behavior—3A. The IBI team selected 3A for all learners to expand their textual behavior (see Skinner, 1957; e.g., reading by finding different words in all directions [e.g., forward, backward, diagonal] on various word searches). Additionally, this program incidentally targeted fine motor skills, handwriting, and independent seat work during both one-on-one and small-group work settings. Train and test stimuli pairs were identified based on similar response effort (e.g., inclusion of pictures, number of letters in words, direction of words in word search, number of columns and rows).

I Spy: Tolerating Failure—3C. The IBI team selected 3C for all learners to increase their scanning and guessing of objects in their environment when given clues. For Derek and Doyle specifically, this program also targeted reacting appropriately if they did not guess correctly (i.e., tolerating failure). Staff would select an object in the environment and say, “I spy with my little eye something that is ____ and ____.” Clues were based on features (e.g., colors, has fur), functions (e.g., used for keeping food cold), classes (e.g., is an animal), and letter sounds (e.g., “starts with /l/”). This program was conducted individually and within a small group to assess tolerating loss if a peer responded correctly first, as well as some observational learning of the rules of the game.

Exclusion: Feature—4C. The IBI team selected 4C for Derek to increase his selection of a picture that did not share the same features of the other pictures in an array. The program also targeted Derek’s description of why a particular stimulus did not belong. Staff laid out three pictures, two of which shared a similar feature (e.g., striped pants, zebra) and one that did not (e.g., yellow cat). Staff would ask, “Which one does not belong?” After Derek selected the correct picture, staff asked, “Why?” and Derek was expected to say, “Because it ____” (e.g., doesn’t have stripes). Train and test stimuli pairs were similar by type of feature (e.g., color, parts of the whole, pattern, material).

Asking What (Pardon)—4F. The IBI team selected 4F for JJ and Doyle to increase their appropriate (i.e., polite, “cool,” assertive) manding to hear a statement again if it was not heard the first time. The program was selected to increase the variety of Doyle’s responses to include polite manding with adults (e.g., “Pardon?” “Sorry, can you say that again?”) and more

socially appropriate manding with peers (e.g., “What?” “Huh?”). This program was selected to increase JJ’s assertive manding if a teacher’s instructions were not heard—or listened to—the first time. Staff would mumble part or all of a statement (e.g., “JJ, write *cat* [coughs]”; “Oh cool, it’s a *mshmlsh*”) and only repeat the statement clearly if the learner appropriately manded. Train and test stimuli pairs were identified based on observations of the typical verbal stimuli that Doyle or JJ ignored (e.g., “[instruction], JJ”).

Identifying Sarcasm—5H. The IBI team selected 5H for Doyle to increase his identification of whether a statement was genuine or sarcastic. This program and specific stimuli were selected to target the subtle facial expressions and intonation of a sarcastic (or genuine) speaker instead of solely the content of his or her statement. That is, Doyle followed most rules unflinchingly and would not comply with an instruction if it was “against the rules.” However, if told to do something that he generally preferred to do (e.g., “Go get your lunch,” “Go outside for recess”) or that was similar to a teacher’s instruction (e.g., “Stand up,” “Go get the bubbles”), he would comply with the instruction despite the speaker’s subtle sarcasm. This proved to be an essential skill at school because Doyle was being bullied in his class and reportedly did not know when other students “were kidding.”

PEAK-G DTT Each day, staff prepared data sheets and stimuli according to the PEAK–G program and the learner’s progress. Staff conducted at least one PEAK–G DTT trial block of 10 trials per program each day for each learner in the IBI setting. Staff followed the DTT procedures outlined in the PEAK–G procedural integrity (PI) checklist (see Dunkel-Jackson, 2016).

Data Summarization and Analysis The staff reviewed data sheets daily and made simple adjustments and revisions that were not modifications to the integrity of PEAK (e.g., change reinforcers, change specific prompt topography). The author and/or staff summarized and graphed data approximately every two days in the PEAK–EDS (i.e., when access to the computer network was available). General mastery criteria of receiving a PEAK score of at least 80 across at least 2 out of 3 consecutive trials were used. Based on visual analysis of the data and observation notes (e.g., error analyses), the author made further adjustments (e.g., changing stimuli or targets, introducing more train stimuli or exemplars, simplifying stimuli across the generalization gradient).

Dependent Variables

Daily PEAK–G scores for both train and test stimuli, PEAK–G program assessment scores, and corresponding PEAK performance matrix scores were used to monitor overall progress. PEAK–G program assessments were administered at the start of the study to guide program selection and at discharge from

IBI to determine progress. Daily PEAK–G total scores for each PEAK–G program were collected by trained staff responsible for delivering IBI. The percentages of the daily PEAK–G total scores for both train and test stimuli were summarized and graphed in a separate series for review and analysis by the intervention team (i.e., parents, IBI staff). The percentage of the daily PEAK–G total score was calculated by dividing the total score for train (or test) stimuli by the total score and multiplying by 100.

Documentation of novel, generalized responses that occurred throughout the day outside of the direct DTT instructional period included permanent products (e.g., homework assignments), direct observation from ASD consultants providing CFS in the school, and anecdotal reports from the parents at clinic meetings. Scores on the VB–MAPP Barriers Assessment and Transition Assessment related to generalization before and after PEAK were also included. These data served as evidence of additional generalized responding.

Interobserver Agreement

Compared to other research settings, community-based IBI settings cannot financially support staff to conduct interobserver agreement (IOA) sessions on 25% of trials due to the intensity of the service they provide. However, in the current study, the author, supervisors, and the IBI staff independently collected PEAK score data during at least three trials during at least one 10-trial block each week for each learner. If IOA had been below 80% on more than three observations, the author would have provided additional monitoring and training (this did not occur). IOA per trial was calculated by dividing the smaller PEAK–G score (e.g., the instructor’s record of 4) by the larger PEAK–G score (e.g., the supervisor’s record of 8) and multiplying by 100 to obtain a percentage. The mean IOA per trial block was calculated by adding all IOA data across all IOA trials within the trial block and dividing by the number of trials with IOA within the trial block. IOA was assessed during 13.3% of all trial blocks (79 trial blocks with IOA out of 593 total trial blocks) with 99.9% IOA (range of 90%–100%).

Procedural Integrity

Similar to IOA, compared to other research settings, community-based IBI settings cannot financially support staff to conduct PI sessions on 25% of trials due to the intensity of the service they provide. However, in the current study, each staff member was required to have successfully completed PI training by meeting and/or exceeding mastery criteria for PEAK–G DTT PI (i.e., 90%). The author and the supervisors (e.g., clinical supervisor, senior therapist) also collected PI data for each staff member at least once each week for

PEAK–G DTT by using a task analysis data sheet created by the first author based on instructional pages in the manual (see Table 2; Dunkel-Jackson, 2016). The percentage of PI task analysis steps conducted correctly during a trial was calculated by dividing the sum of correctly performed steps by the sum of correctly performed, incorrectly performed, and prompted steps and multiplying by 100. The mean percentage of PI task analysis steps conducted correctly during a trial block was calculated by dividing the sum of all correctly performed steps across trials within the trial block by the sum of all correctly performed, incorrectly performed, and prompted steps and multiplying by 100. PI was formally assessed during 13.8% of trial blocks (82 trial blocks with PI out of 593 total trial blocks) with a mean of 95.4% fidelity (range of 65.2%–100%). Of the trials within the trial blocks with PI, 64.7% were trials in which the learner responded incorrectly, which required staff to implement the more complex error correction and data collection procedure.

A small number of additional PI errors were identified indirectly (e.g., team meetings, supervision, binder review). These errors were related to DTT and decision making (e.g., meeting mastery criteria but not moving to the next level) and are outlined in the Results section.

Table 2 PEAK Discrete Trial Training (DTT) Procedural Integrity (PI) steps

DTT PI step
Preparation
1. Prepare data sheet
2. Prepare stimuli
3. Establish motivation
4. Present SD (stimuli)
5. Allow 3 s to respond
Correct response
6. Provide praise
7. Provide reinforcer
8. Collect data
9. Begin next trial
Incorrect response
10. Re-present SD
11. Provide prompt (score 8)
12. Allow 3 s to respond
13. Re-present SD
14. Provide prompt (score 4)
15. Allow 3 s to respond
16. Re-present SD
17. Provide prompt (score 2)
18. Allow 3 s to respond
19. No additional stimuli

SD discriminative stimulus

Results

Figures 1, 2, 3 and 4 display the PEAK–G scores for Doyle, Derek, Eric, and JJ, respectively. During baseline, PEAK scores for the 14 programs were at 0 for four programs and variable with a later decreasing or stabilizing trend for 10 programs. In general, after introducing PEAK–G programs in a staggered manner, total PEAK scores increased for all learners on all programs compared to baseline with variable degrees of mastery (i.e., generalization). The results specific to each learner's progress are outlined in the following sections and include separate analyses for train and test stimuli.

Doyle In general, Doyle's baseline demonstrated low yet variable scores across programs. After introducing PEAK–G for each program, Doyle's correct responses to train stimuli quickly increased, and his correct responses to test stimuli gradually or quickly increased on specific programs. Doyle mastered 2 out of 4 programs before graduating from IBI and transitioning to school. Despite challenging behavior interfering with consistent performance, Doyle made gains in the other two programs compared to baseline.

Asking What—4F. During baseline for 4F, Doyle received a total PEAK score between 40 and 60. Level 1 was quickly mastered within four trials. Staff introduced Level 2 stimuli, which required Doyle to provide more variable responses (e.g., polite “Pardon?”; cool “Huh?”). Doyle mastered Level 2 in nine trials. Generalization was quite evident with Level 3 stimuli and novel stimuli, with Doyle reaching mastery after three and two trials, respectively. The stimulus similarity of train and test stimuli (e.g., “What is your first name?” vs. “What is your last name?”) as well as the limited range of response topographies taught may have proven to be effective for Doyle.

I Spy: Tolerating Failure—3C. For 3C, Doyle received a total PEAK score between 10 and 30 during baseline. Staff introduced the first stimulus set, and the percentage of the total PEAK score gradually increased for both train and test stimuli, with the rate of acquisition faster for train stimuli. Specifically, Doyle responded with 100% accuracy to train stimuli by the 10th trial and 100% accuracy to test stimuli by the 14th trial, with Doyle reaching mastery of Level 1 after 21 trials. Doyle participated in a game of “I Spy” with another peer and individually throughout Level 1. Additional social interaction was observed during games with peers (e.g., “[Peer], do you know that one? I can help you”; “Now it is my turn. I spy with my little eye [train stimuli]. ... Yes, you are right!”). Level 2 stimuli were introduced during Doyle's final week in IBI, with Doyle increasing responses compared to baseline but not quite reaching mastery of Level 2 before his final day, when a novel set of Level 3 stimuli was introduced. The team's choice of train and test stimuli that were similar in stimulus features (e.g., pink and purple art on cubby, purple and blue art on

Fig. 1 PEAK scores across PEAK programs for Doyle. PEAK-G = Promoting the Emergence of Advanced Knowledge (PEAK) relational training system generalization module; BL = baseline; FU = follow-up

cubby) that gradually increased in complexity across the gradient (e.g., door and window are somewhat similar in their features but differ by function) may have proven to be effective for Doyle.

Flexible Textual Behavior—3A. For 3A, Doyle completed word searches with variable PEAK scores of 0 to 40 during baseline. Staff introduced the first stimulus set and observed increases in the percentage of total PEAK scores for train and

test stimuli with some maintaining variability. Doyle completed word searches during a mixture of one-on-one and small-group instructional sessions. Generalization of finding words never previously taught did occur, but due to vocal protesting (e.g., “I don’t want to,” “This one is too small [of font]”) and illegible handwriting, staff provided additional prompts and therefore lower scores. Alterations to the reinforcement schedule (e.g., more frequent reviews of the goals program, a higher

quality reinforcer) did not produce increases in the behaviors associated with flexible textual behavior (i.e., handwriting, attending to task instead of peers). Doyle did not technically reach mastery of Level 1 stimuli, but based on an error analysis (e.g., prompts for focusing on work, not peers, or focusing on writing neatly), the author initiated Level 2 stimuli during Doyle's last 2 weeks in service. The percentage of PEAK scores increased for both train and test stimuli despite continued vocal protesting (e.g., "I don't want to. I'm done [with IBI] in 3 days").

For Doyle, the choice of stimuli pairs for train and test stimuli were matched on physical similarity (e.g., forward words only for train and test stimuli) and complexity (e.g., three- to five-letter words for both train and test stimuli) and gradually increased in physical dissimilarity and complexity across levels. Doyle's accuracy with both train and test stimuli improved, but it is hypothesized that his challenging behavior influenced scores (i.e., prompting on-task behavior more often

than specific flexible textual behavior) and not necessarily generalization of flexible textual behavior.

Identifying Sarcasm—5H. Across baseline for 5H, Doyle received a total PEAK score between 20 and 40 for stimuli involving adult instructions and preferred activities. For trials involving stimuli that were "against the rules," Doyle responded with a total PEAK score of 90 and 100. Staff therefore introduced only stimuli requiring intervention. After one trial block that included primarily verbal prompts (i.e., score of 8 per trial), Doyle quickly mastered each level within three, two, and four trials, respectively, as well as accurate responses to novel stimuli during follow-up. Doyle also began to spontaneously tact the staff's behaviors that were associated with sarcastic and genuine instructions by using words previously learned in other contexts and applying them to this new context (e.g., "That is your mysterious voice," "Those are your crazy eyes," "That is sarcasm").

Fig. 2 PEAK scores across PEAK programs for Derek. PEAK-G = Promoting the Emergence of Advanced Knowledge (PEAK) relational training system generalization module; BL = baseline; FU = follow-up

Derek Overall, Derek demonstrated low scores across programs during baseline. After introducing PEAK–G for each program, Derek’s correct responses to train stimuli quickly increased, and his correct responses to test stimuli gradually increased. Derek mastered 2 out of 3 programs and made significant gains in the third program before graduating from IBI and transitioning to school.

Flexible Textual Behavior—3A. For 3A, Derek completed zero word searches correctly during baseline. After the introduction of the first stimulus set of five train and five test stimuli, the percentage of the total PEAK score increased for both train and test stimuli rather simultaneously. Derek reached mastery of Level 1 after nine trials. Staff then selected new stimuli with a higher response effort for a second stimulus set (Level 2). PEAK scores gradually increased for both train and test stimuli in the second stimulus set, with Derek reaching mastery of Level 2 after 23 trials. Toward the end of Level 2, Derek completed word

searches during a mixture of one-on-one and small-group instructional sessions. Staff chose Level 3 stimuli to be closer to Derek’s academic level, and this resulted in Derek requiring only two trials to reach mastery for Level 3. During follow-up, Derek received a total PEAK score between 60 and 100 on a selection of novel word searches. Derek’s parents also noted that he “loved” doing word searches at home and at school, where he completed 100% (15 of 15) of novel word searches independently.

The matched train and test stimuli based on gradually increasing physical dissimilarity and complexity is likely what promoted Derek’s acquisition and generalization within this program. For example, during the first three PEAK–G trial blocks after baseline, staff prompted 11 of the 15 train stimuli at a Level 2 prompt with limited generalization and increases for corresponding test stimuli. However, across the next six trials, Derek required fewer and fewer prompts on the train stimuli (3 of the 30 train stimuli required a prompt at a level of

Fig. 3 PEAK scores across PEAK programs for Eric. PEAK–G = Promoting the Emergence of Advanced Knowledge (PEAK) relational training system generalization module; BL = baseline; FU = follow-up; R = program revision

Fig. 4 PEAK scores across PEAK programs for JJ. PEAK-G = Promoting the Emergence of Advanced Knowledge (PEAK) relational training system generalization module; BL = baseline; FU = follow-up; R = program revision

2 or 4) and gradually increased accurate flexible textual behavior with the test stimuli.

I Spy: Tolerating Failure—3C. For 3C, Derek received a total PEAK score between 10 and 40 during baseline. After the introduction of the first stimulus set, the percentage of the total PEAK score increased for both train and test stimuli, with the rate of acquisition faster for train stimuli. Specifically, Derek responded with 100% accuracy to train stimuli by the 16th trial and 100% accuracy to test stimuli by the 24th trial,

with Derek reaching mastery of Level 1 after 35 trials. Toward the end of Level 1, Derek successfully participated in a game of “I Spy” with another peer and individually. Staff then introduced five additional train and five additional test stimuli for Level 2. Derek gradually increased accuracy to train stimuli (100% by the 10th trial) and responded similarly to test stimuli across Trials 3 through 10, where he made the same mistake each time. Derek engaged in challenging behavior during test trials of the 11th trial block in Level 2. On

Derek's final day in IBI, he successfully participated in a game of "I Spy" with other peers with novel stimuli in a novel environment (i.e., playground).

When conducting an error analysis for Level 1, Derek initially responded to the same stimulus across trial blocks in a variable manner for both train stimuli (e.g., scoring 2s, 4s, 8s, and 10s across the first 10 trial blocks in response to discriminative stimuli for microwave) and test stimuli (e.g., scoring 0s and 10s across the first 4 days before scoring only 10s in response to discriminative stimuli for fridge). However, as Level 1 continued, Derek began to respond with less variability and increased his independent correct responding (i.e., score of 10).

Exclusion by Feature—4C. For 4C, Derek received a total PEAK score between 10 and 30 during baseline. After the introduction of the first stimulus set, the percentage of the total PEAK score increased for both train and test stimuli, with Derek reaching mastery of Level 1 after 13 trials. Staff then introduced Level 2 and then Level 3 stimuli, with Derek reaching mastery after 11 and eight trials, respectively. For novel stimuli, Derek received total PEAK scores of 80 to 90. Although not all of Derek's specific responses were documented, instructors occasionally noted when Derek provided a stimulus feature that was not chosen and/or taught by the instructor (e.g., trained to respond "TV is square" in the presence of TV, lollipop, and ball, but Derek once noted "TV needs a plug").

Eric Overall, baseline for Eric was low, with little variability across programs. After introducing PEAK-G for each program, Eric's correct responses to train stimuli gradually increased, but limited generalization was observed. Although Eric did not master any programs and despite challenging behavior interfering with on-task behavior, Eric increased his verbal repertoire through several directly trained skills before graduating from IBI and transitioning to school. Eric also was ill during part of the latter portion of the study and attended partial hours or no hours. Compared to the other learners, Eric had the lengthiest documented history of not maintaining or generalizing skills "mastered" in his IBI programming. As such, the author referred Eric to an occupational therapist for further assessment of his fine and gross motor movements as well as a psychologist for further diagnostic assessment.

Flexible Textual Behavior—3A. For 3A, Eric accurately completed zero word searches during baseline. Staff introduced the first stimulus set and observed slight increases in the percentage of total PEAK scores for train stimuli but not test stimuli. Eric improved his performance on all stimuli (e.g., found at least some words), but for most train worksheets, Eric required a prompt level of 2, which included verbal, gestural, and full physical prompting on at least some portion of the word search (i.e., 47 of the first 55 trials required a Level 2 prompt). Due to an increase in vocal protesting (e.g., "I can't

do this!"), staff provided additional prompts and conducted more frequent preference assessments. After the 12th trial, the IBI team reviewed the data and identified a significant revision that included (a) more frequently reviewing Eric's goals program throughout the worksheet trial block (e.g., "Remember your goals; if you try your best, you get the iPad"; "Good job doing your work, it is almost time for *Tom and Jerry*") and (b) modifying the 10-trial block to include two smaller trial blocks across two instructional periods to reduce the response effort. Stimuli were not changed because Eric had found most words at least once across the various trials.

After the revision, the percentage of the total PEAK score increased for train stimuli between 20% and 100%, with some increases for test stimuli as well. However, test stimuli responses stabilized and train stimuli responses began to decrease, so the IBI team identified a second revision that included training more exemplars with similar stimulus features. After a week of illness, Eric returned to IBI and increased his correct responses to the additional train stimuli but performed variably on the remaining test stimuli. Staff also noted variable but lower rates of on-task behavior during group sessions (mean of 62.1% of time-sampling intervals, range of 0% to 100%) compared to one-on-one sessions (mean of 96.9% of time-sampling intervals, range of 87.5% to 100%). The IBI team then identified a third revision that included training all stimuli and conducting trial blocks during one-on-one sessions only. This was followed by Eric reaching modified mastery of Level 1. Level 2 stimuli were introduced during Eric's last week before returning to school, with results demonstrating a low level with an upward trend for both train and test stimuli. School staff later reported that Eric completed simple word searches in his classroom with some prompting to stay on task.

Intraverbal Substitution by Function—2A. During baseline for 2A, Eric received a PEAK score of 0 to 10. After introducing the first stimulus set, the percentage of the total PEAK score for train and test stimuli increased to 20% to 70% and 0% to 40%, respectively, without reaching mastery. Staff reviewed the errors and noted that although Eric responded with previously learned rote answers that may be grammatically incorrect (e.g., SD: "What can you do with a pencil?" Eric: "For writing"), Eric had made some progress but required numerous intrusive prompts (i.e., Level 2 prompt) to respond with an alternate function. Therefore, based on recommendations from the PEAK-G manual, the team decided to train all stimuli to provide more exemplars that may enhance generalization. With continued prompting, the total PEAK score increased to between 50 and 80. During the final two trial blocks, staff provided an immediate partial verbal prompt instead of allowing Eric 3 s to respond in a rote manner. This revision was followed by similar scores but fewer rote errors.

I Spy: Tolerating Failure—3C. For 3C, Eric received a total PEAK score between 0 and 30 during baseline. Staff introduced the first stimulus set and the percentage of the total PEAK score gradually increased for both train and test stimuli, with the rate of acquisition faster for train stimuli. Specifically, Eric responded with 100% accuracy to train stimuli by the ninth trial and with as high as 60% accuracy to test stimuli. For train stimuli, Eric's errors included finding the correct item but tacting it incorrectly or finding an incorrect item despite correctly finding it during prior trial blocks. Eric did not generalize across train and test stimuli despite the stimuli similarity (e.g., broom and mop) and mastering previous tacting and intraverbal programs with similar stimuli. After 23 trials, the IBI team made a significant revision by adding more train stimuli with similar features as additional exemplars, which was followed by a large increase in the total PEAK score to mastery. Staff then introduced Level 2 stimuli, and Eric immediately responded at a higher level on both train and test trials than during the first few trial blocks of Level 1. Although Eric did not fully master 3C, he successfully participated in a game of "I Spy" with other peers with novel stimuli in a novel environment (i.e., playground) on the day before returning to school. Additionally, his participation in a game of "I Spy" with siblings greatly improved. For example, during a family observation at the IBI center, Eric's stepsibling participated in a game of "I Spy" with Eric, which reportedly impressed the parents because Eric could now play a game with his typically developing stepsibling in multiple settings.

JJ Similar to Eric, JJ's baseline was low, with little variability across programs. After introducing PEAK-G for each program, JJ's correct responses to train stimuli gradually increased for all programs with some observed generalization. Although, like Eric, JJ did not master any programs and despite challenging behavior interfering with on-task behavior, JJ enhanced his verbal repertoire through several directly trained skills before graduating from IBI and transitioning to school. The author also referred JJ to a school occupational therapist for further assessment of his fine and gross motor movements, a school behavior intervention team due to his challenging behavior, and the school psychologist for further diagnostic assessment.

Asking What—4F. During baseline for 4F, JJ received a total PEAK score of 0. Level 1 stimuli were introduced with gradual acquisition of train stimuli and no generalization. After the seventh trial, the IBI team reviewed success across staff who used slightly different verbal prompts (e.g., full verbal, partial verbal) within the Level 8 and Level 4 prompts and reviewed which prompts eventually produced a response during a Level 2 prompt (25 of the first 35 trials required a Level 2 prompt). Based on the staff's observation that JJ had begun to repeat the SD with the prompt or only mumble a rote "What?" the IBI team made an adjustment that included (a) two staff,

one as a communicative partner and one as a back prompt, and (b) a variety of textual prompts (i.e., red "What?" card, yellow "Sorry, can you say that again?" card) for a Level 8 prompt that could be faded across trials. The adjustment resulted in increased correct responses to train stimuli and limited generalization. However, staff reported that contingent upon the delivery of the SD, JJ had consistently begun to immediately engage in echoic behavior, thus repeating the SD before a prompt could be delivered.

A second adjustment included training additional exemplars, which resulted in an increase of PEAK scores. However, because JJ continued to engage in immediate echoic behavior, the IBI team decided to make a significant revision to the PEAK prompt hierarchy and included an immediate prompt (i.e., errorless teaching) instead of requiring staff to wait 3 s after presenting the SD. Across Phases 3, 4, 5, and 6 of the errorless teaching, JJ responded to nearly all stimuli accurately while staff gradually and systematically faded prompt intrusiveness (i.e., after two trials at 100% accuracy, staff faded from a textual, gestural, and verbal prompt [R3] to a textual and verbal prompt [R4] to a 1-s delayed textual prompt [R5] and then to a 3-s delayed textual prompt [R6]; see Touchette & Howard, 1984). JJ responded without prompts on the last phase (R6), which more closely mirrored the PEAK prompt procedures. Staff then reintroduced the original train stimuli (reinforced and prompted) and test stimuli (recently reinforced and prompted during errorless teaching) to assess maintenance (not generalization because responses to test stimuli were reinforced). JJ gradually decreased his accurate responses to test stimuli and train stimuli across the phase. JJ also did not respond appropriately to staff or peers during play breaks despite modified stimuli (e.g., "JJ, want some Smarties? Go over to the [mumble]"). Instead, JJ either engaged in echoic behavior of the SD or turned away from the staff member or peer. Unfortunately, further revisions and adjustments could not be made due to JJ's transition to school.

Flexible Textual Behavior—3A. For 3A, JJ accurately completed zero word searches during baseline. After introducing the first stimulus set, staff observed gradual increases in the percentage of total PEAK scores for train stimuli up to 90% and modest increases for test stimuli. The trial block where the test stimuli data point is at 100% indicates an unintended, shortened trial block of four train stimuli and one test stimulus. Similar to Eric, JJ's performance improved across stimuli but often required a prompt level of 4 or 8 to stay on task. In fact, toward the end of Level 1, JJ began to engage in more frequent vocal protesting (e.g., "This is too hard!"), flopping, and minor property destruction during both group and one-on-one sessions. On-task data were similarly variable for group (mean of 72% of time-sampling intervals, range of 10% to 100%) and one-on-one (mean of 68.8% of time-sampling intervals, range of 50% to 100%) sessions.

After 33 trials, the IBI team decided to make a more significant revision that included (a) more frequently reviewing JJ's goals program throughout the worksheet trial block (e.g., "Remember your goals; if you do your work, we'll go for a walk"), (b) more frequent primary reinforcers (e.g., tickles, chips) throughout the trial block, and (c) modifying the 10-trial block to include two smaller trial blocks across two instructional periods to reduce the response effort (presented as one data point). Similar to Eric, stimuli were not changed because JJ had found most words at least once across the various trials. After the revision, the percentage of the total PEAK score increased for train stimuli to between 60% and 90% with the remaining test stimuli at 100%, thus meeting mastery criteria.

Staff introduced Level 2 stimuli during JJ's final weeks in IBI and noted increased responding to train stimuli (17 trials with a Level 2 prompt, three trials with a Level 4 prompt, and four trials with no prompts) and regular refusal to complete test stimuli (19 of 20 trials at a score of 0). After reviewing the data sheet notes, interviewing staff, and observing JJ, it seemed that JJ had discriminated between the contingencies of train and test stimuli, thereby engaging in textual behavior with train stimuli, which included staff praise and prompts, and immediately refusing to engage in textual behavior with test stimuli, which included no praise, no prompts, and a shorter delay to the next trial and ultimate reinforcer break. Therefore, the author made a revision that attempted to keep trial block durations as similar as possible to minimize adventitious differences in delay to reinforcement. The revision for test stimuli included prompting on-task behavior (e.g., "You have to try your best before we can go play"; "Do the frog word search [test] like you did the bat word search [train]") and praising on-task behavior (e.g., "See! You *can* do it when you try your best!") but not specific textual behavior (e.g., no full physical prompt to circle a word). Following the revision, JJ increased responding to both train and test stimuli to 80% and 60% of the total PEAK score, respectively.

Intraverbal Substitution by Function—2A. During baseline for 2A, JJ received a PEAK score of 0 to 20. After introducing the first stimulus set, the percentage of the total PEAK score for train and test stimuli increased to 20% to 100% and 0% to 40%, respectively, without reaching mastery. The IBI team reviewed JJ's errors and noted that JJ initially responded with previously learned rote answers and then began to respond with alternate functions for train stimuli only, resulting in no generalization across similar stimuli. Therefore, the team decided to train additional exemplars, which resulted in increased PEAK scores to nearly 100%. Staff introduced Level 2 stimuli, which was followed by increases in the percentage of total PEAK scores for both train and test stimuli up to 80%.

I Spy: Tolerating Failure—3C. For 3C, JJ received a total PEAK score between 0 and 10 during baseline. After

introducing the first stimulus set, the percentage of the total PEAK score gradually increased for both train and test stimuli, with the rate of acquisition faster for train stimuli. Specifically, JJ responded with 100% accuracy to train stimuli by the 13th trial and to test stimuli by the 22nd trial, with mastery of Level 1 by the 23rd trial (although staff conducted an additional trial block). JJ successfully participated in a game of "I Spy" individually and as part of a small group during three trial blocks in Level 1. Staff then introduced Level 2 stimuli. Unfortunately, staff trained all stimuli during the majority of Level 2 due to an error in data sheet setup when transferring data from a completed data sheet to a new data sheet (i.e., staff did not cross out Prompt Levels 2, 4, and 6 on the new data sheet for test stimuli) while the author and supervisor were away from the IBI site. Despite the error, JJ increased responding to Level 2 stimuli and "maintained" responding to test stimuli under baseline conditions (i.e., no reinforcement, no prompting) for the final trial block. JJ's parents reported that after beginning PEAK-G, JJ began playing a game of "I Spy" at home with his siblings.

Further Generalization and Social Validity Near the end of the study, all learners received homework of various trained (JJ, Eric) and novel (Derek, Doyle) word searches that were reportedly completed with minimal (JJ, Eric) to no prompting (Derek, Doyle) at home with their families and at school. Additionally, scores on the VB-MAPP Barriers Assessment: Failure to Generalize item and the VB-MAPP Transition Assessment: Generalization of Skills Across Time, Settings, Behaviors, Materials, and People item also suggested further generalization after PEAK was introduced, as scores remained at criterion (Doyle), improved (Derek, Eric), or remained the same (JJ; see Tables 3 and 4).

Discussion

All learners in the current study enhanced their verbal repertoires through the use of PEAK-G, a curriculum that allowed the IBI team to target unique advanced language skills that are rarely examined in the literature. Two learners, Derek and Doyle, demonstrated mastery of and generalization within their programs. Eric and JJ demonstrated limited generalization but substantial increases in responding via direct training compared to baseline. Although this may indicate that PEAK-G was unsuccessful for Eric and JJ, their performance improved compared to baseline, and the information obtained through PEAK-G's train and test methodology is invaluable for future program planning. For example, during their transition to school, staff informed the school team that skills may need to be taught using more direct contingencies and more exemplars for Eric and JJ, whereas Derek and Doyle may

Table 3 VB–MAPP barriers assessment: failure to generalize

Participant	Preassessment	Postassessment	Change
Doyle	0 (typically demonstrates stimulus and response generalization at a level commensurate with other skills)	0	Stayed at criterion since PEAK
Derek	Not testable (start of IBI) 1 (shows some kind of difficulty with any type of stimulus generalization or with any skill)	0	Improved since PEAK
Eric	Not testable (start of IBI) 2 (requires formal generalization training on most skills)	1	Improved since PEAK
JJ	Not testable (start of IBI) 1	1	Stayed the same since PEAK

VB–MAPP Verbal Behavior Milestones Assessment and Placement Program *PEAK* Promoting the Emergence of Advanced Knowledge (PEAK) Relational Training System, *IBI* intensive behavioral intervention

require fewer exemplars and can more quickly generalize skills across more of the generalization gradient.

Derek and Doyle demonstrated generalization across increasingly dissimilar stimuli (e.g., Derek’s Exclusion: Feature—4C simple color or shape of train and test stimuli to type of material and parts of the whole, which are more complex; Derek and Doyle’s Flexible Textual Behavior—3A forward-only word search to forward, backward, and diagonal word search). Eric and JJ required significant prompting to increase responding to train stimuli, which did not allow the IBI team to make significant changes to the stimuli features to increase accuracy with test stimuli. For example, Eric and JJ made similar mistakes on Intraverbal Substitution by Function—2A (e.g., rote, previously learned response of “for writing” for pen and pencil stimuli); when examining their learning history and previous intraverbal programs across feature, function, and class, it was found that they had received reinforcement for these rote responses. For Eric and JJ, it is hypothesized that the salient feature of the PEAK discriminative stimulus was the object (i.e., pencil or pen) and not the other stimuli (i.e., “What can you do with a ___?” or “What else can you do with a ___?”), as the object had previously signaled the availability of reinforcement. It is possible that the response effort for train stimuli was too high

(e.g., multiple word searches, each with a different complexity within a single trial block) or the stimuli were too different (e.g., pencil or pen could have been red pen or blue pen). However, it is also possible that their previous intervention and degree of tight stimulus control interfered with flexible responding to PEAK programs. Yet another possible interpretation of these data is that Derek and Doyle may have had sufficient histories with multiple-exemplar training and could now be considered “generalized generalizers.” In contrast, Eric and JJ may not yet have had a history with a sufficient number and type of multiple exemplars to establish such a repertoire. Future studies could more closely examine the extent to which prior instructional histories influence responding to train and test stimuli, acquisition, and generalization.

The outcomes of a train and test method may provide practitioners and teachers with specific information about their student’s individual learning history, including the probability of generalization, which allows them to more specifically program for generalization (e.g., program common stimuli, train sufficient exemplars, provide multiple-exemplar training; Stokes & Baer, 1977) and request additional resources as needed (e.g., educational assistants to provide individualized, direct instruction).

Table 4 VB–MAPP transition assessment: generalization of skills across time, settings, behaviors, materials, and people

Participant	Preassessment	Postassessment	Change
Doyle	5 (consistently demonstrates both stimulus and response generalization on the first or second trial)	5	Stayed at criterion since PEAK
Derek	Not testable (start of IBI) 4 (demonstrates spontaneous response generalization in the natural environment on 10 occasions)	5	Improved since PEAK
Eric	Not testable (start of IBI) 2 (generalizes to new materials, but only after extensive generalization [multiple exemplar] training)	3 (demonstrates spontaneous stimulus generalization in the natural environment on 10 occasions)	Improved since PEAK
JJ	Not testable (start of IBI) 3	3	Stayed the same since PEAK

VB–MAPP Verbal Behavior Milestones Assessment and Placement Program, *PEAK* Promoting the Emergence of Advanced Knowledge (PEAK) Relational Training System, *IBI* intensive behavioral intervention

The limitations of the current study include errors of commission during some trial blocks (e.g., one trial block for Eric's 3A, Level 3 for JJ's 3C) where staff inadvertently provided reinforcement and prompting during some of the test trials. This is likely not an uncommon mistake when using a train and test method, but it can be mitigated by better organizing stimuli, which the IBI team did expertly (e.g., a three-pronged binder for laminated, double-sided word searches with stimuli numbers written in the corner of each; stimuli cards denoting "test" or "train" rather than only the data sheet and program page), and more carefully transferring stimuli or notes about adjustments from one completed data sheet to a new data sheet (e.g., crossing out Prompt Levels 2, 4, and 8 for test trials on the new data sheet; enhancing the existing data sheet to include a Notes section per trial block). Errors like this may occur when direct supervisors are unavailable and another supervisor without significant PEAK experience provides coverage of the team. Another potential limitation is that some learners who attend school may also be working on similar programs, which highlights a potential confounding variable. Although plausible, parents reported that these programs had never been targeted before, and it certainly seems unlikely that school staff addressed these specific targets in the last 2 months of school before the summer break (i.e., the period when much of the experiment was conducted). Additionally, although children were enrolled in IBI, these results may not be fully applicable to all IBI learners with varying skill profiles.

Future research should examine the applicability of these results to children who are younger (i.e., participate in EIBI), have additional diagnoses (e.g., intellectual disability, motor deficits), exhibit lower skill profiles (e.g., children in home or one-on-one clinic settings), and/or exhibit advanced skill profiles (e.g., children receiving individualized intensities of service, attending school). Also, future researchers may wish to examine specific learner profiles and outcomes that may allow practitioners to better identify for which learners PEAK-G will be a more appropriate curriculum compared to PEAK-DT, which directly trains responses to stimuli. Preliminary results of the current study demonstrate that PEAK-G was most successful with the two children who had higher scores on PEAK-DT, PEAK-G, and VABS-II and shorter instructional histories with dense reinforcement schedules and errorless prompting commonly associated with DTT. Coupled with the results of Dunkel-Jackson (2016) and those of Malkin et al. (2017), it is plausible that the lower a VABS-II score, the more likely direct contingencies (i.e., PEAK-DT) or extremely similar stimuli (e.g., PEAK-G with simplified stimuli) will be required to demonstrate skill acquisition and generalization. Practitioners would benefit greatly from a more thorough, and empirically based, set of guidelines for selection and sequencing of programs across all PEAK modules.

Applied replications and extensions of this study should be conducted by additional researchers and laboratories to provide further information about the utility of PEAK. Future applied researchers may also wish to examine the extension of this curriculum to other services (e.g., ABA therapy, schools, day treatment) and other populations (e.g., children with acquired brain injury, adults with developmental disabilities). Comparing the outcomes of children who receive instruction informed by PEAK versus another curriculum guide (e.g., VB-MAPP, Assessment of Basic Language and Learning Skills—Revised [ABLLS-R], state and/or provincial school academic curricula) may also be valuable to practitioners responsible for choosing curricula. The impact of simply reviewing a new curriculum such as PEAK may also provide the opportunity for practitioners to more creatively program for their clients in both targets (e.g., expanded verbal operants) and methods used (e.g., train and test, stimulus equivalence, transformation of stimulus function, naturalistic teaching strategies). Finally, although learners in the current study clearly made gains, a longitudinal study of the utility of introducing all PEAK modules across several learners' time in IBI and their transition to an appropriate ABA-based service and/or school would help confirm the view that PEAK is valuable for IBI.

In sum, the enhanced discrimination skills of children with ASD may hinder generalization of advanced verbal behavior and language skills. The PEAK-G (Dixon, 2014) curriculum and train and test DTT method provide a promising instructional package to enhance generalization of skills across a variety of stimuli targets. The current study demonstrated specific gains in acquired and generalized advanced language skills of children with ASD participating in a community-based, clinical setting. This successful real-life application of the PEAK instructional package suggests to practitioners that PEAK-G may be a viable option for community-based IBI and ABA programs.

Acknowledgements The authors wish to thank Shannon Borch, Jenn Watts, Jen Hocking, Donna MacDonald, Frances Donovan, Ginny Carriere, Melecia Washington, Renee Ruddock, Sandie Hoskin, Sarah Burch, and Simone Bégin.

Compliance with Ethical Standards 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 Declaration of Helsinki and its later amendments or comparable ethical standards.

Informed Consent Informed consent was obtained from the parent or legal guardian of all individual participants included in the study. Additional informed consent was obtained from the parent or legal guardian of all individual participants for whom identifying information is included in this article. This research is based on data used in Sarah M. Dunkel-Jackson's doctoral dissertation.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 1, 91–97.
- Bailey, S. L. (1981). Stimulus overselectivity in learning disabled children. *Journal of Applied Behavior Analysis*, 14, 239–248.
- Bass, J. D., & Mulick, J. A. (2007). Social play skill enhancement of children with autism using peers and siblings as therapists. *Psychology in the Schools*, 44, 727–735.
- Belisle, J., Dixon, M. R., Stanley, C. R., Munoz, B., & Daar, J. H. (2016). Teaching foundational perspective-taking skills to children with autism using the PEAK-T curriculum: single-reversal “I–you” deictic frames. *Journal of Applied Behavior Analysis*, 49, 965–969. <https://doi.org/10.1002/jaba.324>.
- Betz, A. M., Higbee, T. S., & Pollard, J. S. (2010). Promoting generalization of mands for information used by young children with autism. *Research in Autism Spectrum Disorders*, 4, 501–504.
- Blacklock, K., Perry, A., & Dunn Geier, J. (2014). Examining the effectiveness of intensive behavioural intervention in children with autism aged 6 and older. *Journal on Developmental Disabilities*, 20, 37–49.
- Brown, S. M., & Bebko, J. M. (2012). Generalization, overselectivity, and discrimination in the autism phenotype: a review. *Research in Autism Spectrum Disorders*, 6, 733–740.
- National Autism Center. (2009). Results of the National Standards Project, Phase I. Retrieved from <http://www.nationalautismcenter.org/reports/>.
- Centers for Disease Control and Prevention (CDC). (2014). Community report on autism from the Autism and Developmental Disabilities Monitoring (ADDM) Network. Retrieved from http://www.cdc.gov/ncbddd/autism/states/comm_report_autism_2014.pdf.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson.
- Cuvo, A. J. (2003). On stimulus generalization and stimulus classes. *Journal of Behavioral Education*, 12, 77–83.
- Daar, J. H., Negrelli, S., & Dixon, M. R. (2015). Derived emergence of WH question–answers in children with autism. *Research in Autism Spectrum Disorders*, 19, 59–71. <https://doi.org/10.1016/j.rasd.2015.06.004>.
- Day-Watkins, J., Murray, R., & Connell, J. E. (2014). Teaching helping to adolescents with autism. *Journal of Applied Behavior Analysis*, 47, 850–855. <https://doi.org/10.1002/jaba.156>.
- Dickson, C. A., Wang, S. S., Lombard, K. M., & Dube, W. V. (2006). Overselective stimulus control in residential school students with intellectual disabilities. *Research in Developmental Disability*, 27, 618–631.
- Dixon, M. R. (2014). *PEAK relational training system: direct training module*. Carbondale, IL: Shawnee Scientific Press.
- Dixon, M. R. (2015). *PEAK relational training system: Equivalence module*. Carbondale, IL: Shawnee Scientific Press.
- Dixon, M. R. (2017). *PEAK relational training system: Transformation module*. Carbondale, IL: Shawnee Scientific Press.
- Dixon, M. R., Belisle, J., Whiting, S. W., & Rowsey, K. E. (2014a). Normative sample of the PEAK relational training system: direct training module and subsequent comparisons to individuals with autism. *Research in Autism Spectrum Disorders*, 8, 1597–1606.
- Dixon, M. R., Carman, J., Tyler, P. A., Whiting, S. W., Enoch, M., & Daar, J. H. (2014b). PEAK relational training system for children with autism and developmental disabilities: correlations with Peabody picture vocabulary test and assessment reliability. *Journal of Developmental and Physical Disabilities*, 26, 603–614. <https://doi.org/10.1007/s10882-014-9384-2>.
- Dixon, M. R., Whiting, S. W., Rowsey, K., & Belisle, J. (2014c). Assessing the relationship between intelligence and the PEAK relational training system. *Research in Autism Spectrum Disorders*, 8, 1208–1213.
- Dixon, M. R., Belisle, J., Stanley, C. R., Rowsey, K. E., Daar, J. H., & Szekeley, S. (2015). Toward a behavior analysis of complex language for children with autism: evaluating the relationship between PEAK and the VB-MAPP. *Journal of Developmental and Physical Disabilities*, 27, 223–233. <https://doi.org/10.1007/s10882-014-9410-4>.
- Dixon, M. R., Stanley, C. R., Belisle, J., & Rowsey, K. E. (2016). The test-retest and interrater reliability of the promoting the emergence of advanced knowledge-direct training assessment for use with individuals with autism and related disabilities. *Behavior Analysis: Research and Practice*, 16, 34–40. <https://doi.org/10.1037/bar0000027>.
- Dixon, M. R., Peach, J., Daar, J. H., & Penrod, C. (2017a). Teaching complex verbal operants to children with autism and establishing generalization using the PEAK curriculum. *Journal of Applied Behavior Analysis*, 50, 317–331.
- Dixon, M. R., Rowsey, K. E., Gunnarsson, K. F., Belisle, J., Stanley, C. R., & Daar, J. H. (2017b). Normative sample of the PEAK relational training system: generalization module with comparison to individuals with autism. *Journal of Behavioral Education*, 26, 101–122. <https://doi.org/10.1007/s10864-016-9261-4>.
- Dunkel-Jackson, S. M. (2016). *PEAK for IBI: Examination of the utility of the PEAK Relational Training System within intensive behavioral intervention* (unpublished doctoral dissertation). Southern Illinois University, Carbondale, IL.
- Freeman, N. L., & Perry, A. (2010). Outcomes of intensive behavioural intervention in the Toronto preschool autism service. *Journal on Developmental Disabilities*, 16, 17–32.
- Froelich, A. L., Anderson, J. S., Bigler, E. D., Miller, J. S., Lange, N. T., Dubray, M. B., et al. (2012). Intact prototype formation but impaired generalization in autism. *Research in Autism Spectrum Disorders*, 6, 921–930. <https://doi.org/10.1016/j.rasd.2011.12.006>.
- Gould, E., Dixon, D. R., Najdowski, A. C., Smith, M. N., & Tarbox, J. (2011). A review of assessments for determining the content of early intensive behavioral intervention programs for autism spectrum disorders. *Research in Autism Spectrum Disorders*, 5, 990–1002.
- Granpeesheh, D., Dixon, D. R., Tarbox, J., Kaplan, A. M., & Wilke, A. E. (2009). The effects of age and treatment intensity on behavioral intervention outcomes for children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 3, 1014–1022.
- Guttman, N., & Kalish, H. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, 51, 79–88.
- Handleman, J. S., & Harris, S. L. (1980). Generalization from school to home with autistic children. *Journal of Autism and Developmental Disorders*, 10, 323–333.
- Leaf, R. B., Taubman, M. T., McEachin, J. J., Leaf, J. B., & Tsuji, K. H. (2011). A program description of a community-based intensive behavioral intervention program for individuals with autism spectrum disorders. *Education and Treatment of Children*, 34, 259–285.
- 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.
- Lovaas, O. I., Koegel, R. L., & Schreibman, L. (1979). Stimulus overselectivity in autism: a review of research. *Psychological Bulletin*, 86, 1236–1254.
- Lovett, S., & Rehfeldt, R. A. (2014). An evaluation of multiple exemplar instruction to teach perspective-taking skills to adolescents with Asperger syndrome. *Behavioral Development Bulletin*, 19, 22–36.
- Malkin, A., Dixon, M. R., Speelman, R., & Luke, N. (2017). Evaluating the relationships between the PEAK relational training system – direct training module, assessment of basic language and learning skills – revised, and the Vineland adaptive behavior scales – II.

- Journal of Developmental and Physical Disabilities*, 29, 341–351. <https://doi.org/10.1007/s10882-016-9527-8>.
- Marzullo-Kerth, D., Reeve, S. A., Reeve, K. F., & Townsend, D. B. (2011). Using multiple exemplar training to teach a generalized repertoire of sharing to children with autism. *Journal of Applied Behavior Analysis*, 44, 279–294.
- Matthews, B. (1994). *Stimulus overselectivity, stimulus generalization, and visual context in adults with autism* (Master's thesis). Retrieved from <http://seds.sturt.flinders.edu.au/matthews/thesis/>.
- McEachin, J. J., Smith, T., & Lovaas, O. I. (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal on Mental Retardation*, 97, 359–372.
- McKeel, A. N., Rowsey, K., Dixon, M. R., & Daar, J. H. (2014). Correlation between PEAK relational training system and one-word picture vocabulary tests. *Research in Autism Spectrum Disorders*, 12, 34–39. <https://doi.org/10.1016/j.rasd.2014.12.007>.
- McKeel, A. N., Dixon, M. R., Daar, J. H., Rowsey, K. E., & Szekely, S. (2015a). Evaluating the efficacy of the PEAK relational training system using a randomized controlled trial of children with autism. *Journal of Behavioral Education*, 24, 230–241. <https://doi.org/10.1007/s10864-015-9219-y>.
- McKeel, A. N., Rowsey, K. E., Belisle, J., Dixon, M. R., & Szekely, S. (2015b). Teaching complex verbal operants with the PEAK relational training system. *Behavior Analysis in Practice*, 8, 241–244. <https://doi.org/10.1007/s40617-015-0067-y>.
- Moes, D. R., & Frea, W. D. (2002). Contextualized behavioral support in early intervention for children with autism and their families. *Journal of Autism and Developmental Disorders*, 32, 519–533.
- Okuyama, T., & Isawa, S. (2010). Right-left discrimination from one's own and another person's viewpoint in children with autism: higher-order conditional discrimination and generalization of viewpoint. *Japanese Journal of Behavior Analysis*, 24, 2–16.
- Perry, A., Cummings, A., Dunn Geier, J., Freeman, N., Hughes, S., LaRose, L., et al. (2008a). Effectiveness of intensive behavioral intervention in a large, community-based program. *Research in Autism Spectrum Disorders*, 2, 621–642.
- Perry, A., Cummings, A., Dunn Geier, J., Freeman, N. L., Hughes, S., Managhan, T., et al. (2008b). Predictors of outcome for children receiving intensive behavioral intervention in a large, community-based program. *Research in Autism Spectrum Disorders*, 5, 592–603.
- Persicke, A. M., Tarbox, J., Ranick, J., & St. Clair, M. (2013). Teaching children with autism to detect and respond to sarcasm. *Research in Autism Spectrum Disorders*, 7, 193–198.
- Rincover, A., & Koegel, R. (1975). Setting generality and stimulus control in autistic children. *Journal of Applied Behavior Analysis*, 8, 235–246.
- Rowsey, K. R., Belisle, J., & Dixon, M. R. (2014). Principal component analysis of the PEAK relational training system. *Journal of Developmental and Physical Disabilities*, 27, 15–23. <https://doi.org/10.1007/s10882-014-9398-9>.
- Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioral treatment for children with autism: four-year outcome and predictors. *American Journal of Mental Retardation*, 110, 417–438.
- Skinner, B. F. (1953). *Science and human behavior*. New York: Free Press.
- Skinner, B. F. (1957). *Verbal behavior*. New York: Appleton-Century-Crofts.
- Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland adaptive behavior scales, second edition: survey forms manual*. Bloomington: Pearson.
- Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis*, 10, 349–367.
- Stokes, T. F., & Osnes, P. G. (1989). An operant pursuit of generalization. *Behavior Therapy*, 20, 337–355.
- Sundberg, M. L. (2008). *Verbal behavior milestones assessment and placement program*. Concord, CA: AVB Press.
- Touchette, P. E., & Howard, J. S. (1984). Errorless learning: reinforcement contingencies and stimulus control transfer in delayed prompting. *Journal of Applied Behavior Analysis*, 17, 175–188.
- Wilhelm, H., & Lovaas, O. I. (1976). Stimulus overselectivity: a common feature in autism and mental retardation. *American Journal of Mental Deficiency*, 81, 26–31.