A|B|A|I
Association for Behavior Analysis International

Check for updates

Evaluation of Broad Skill Repertoire Changes Over 3 Months of PEAK Implementation in a Special Education Classroom

Jordan Belisle[1] · Leah Clark[2] · Kayla Welch[1,2] · Nicole McDonald[1] · Lindsey Dennis[1] · Elana Sickman[1]

Abstract

The Promoting the Emergence of Advanced Knowledge (PEAK) Relational Training System provides a comprehensive language and cognition training strategy for use with children with disabilities. In the present study, we conducted PEAK within a special education classroom over the course of 3 months, and language and cognitive skills were assessed using the PEAK Comprehensive Assessment (PCA). On average, each of the five participants mastered 11.2 target programs across PEAK modules with an average of 186.4 new verbal behaviors or supporting behaviors. All participants showed an increase on the PCA following the intervention, with an average increase of 39.6 correct responses, and parents and staff reported improvements in participants' use of language at home and at school relative to prior semesters.

Keywords Autism · Comprehensive programming · PEAK · Relational frame theory

Comprehensive programs may involve multiple skills that relate to broader performance repertoires (Granpeesheh et al., 2009). Notably, the University of California at Los Angeles's Young Autism Project (Lovaas & Smith, 1989) demonstrated marked gains in intelligence test scores following intensive behavior-analytic intervention, and this result has since been replicated (e.g., Smith & Lovaas, 1998). Promoting the Emergence of Advanced Knowledge Relational Training System is one comprehensive assessment to intervention approach that has generated a large amount of empirical research (PEAK; Dixon et al., 2017). PEAK contains four training modules that target specific modalities of learning to improve language and cognitive skills in children with disabilities (e.g.,

verbal operants, relational framing, and supporting behaviors such as eye contact and waiting). PEAK Direct Training (PEAK-DT) uses discrete-trial training to directly teach new language skills. PEAK Generalization (PEAK-G) contains more advanced targets and tests for the generalization of new skills across behaviors and contexts. PEAK Equivalence (PEAK-E) integrates equivalence-based instruction to target the learning processes of reflexivity, symmetry, transitivity, and equivalence. Finally, PEAK Transformation (PEAK-T) integrates other relational frames and tests for transformations of stimulus function.

Assessment research suggests that the PEAK-DT and PEAK-G assessments produce scores that are related to other available assessments of language, as well as intelligence tests, in multiple studies (Ackley et al., 2019). Earlier research has also supported PEAK intervention when targeting select programs, such as tacting metaphorical emotions and foundational perspective taking, among others (Dixon et al., 2017). Although teaching these and related skills is undoubtedly important, the greatest advantage afforded by the use of comprehensive programming may be improvements in broader skill repertoires that can occur when multiple programs are conducted over a more extended period of time.

Some early research on PEAK has emerged to support these broader skill repertoire improvements. For example, McKeel et al. (2015) conducted a randomized control trial evaluating PEAK-DT using the PEAK-DT indirect

Research Highlights
- The intervention took place within a special education setting.
- PEAK Comprehensive Assessment results show broad repertoire growth across participants.
- Social validity results suggest changes in participants' use of language at school and at home.
- Procedures were individualized across participants.

✉ Jordan Belisle
jbelisle@missouristate.edu

1 Psychology Department, Missouri State University, 901 S. National Ave, Springfield, MO 65897, USA

2 Pender Public School, Pendragon Academy, Pender, Nebraska, USA

assessment as a pre-post measure. Results showed that participants in the experimental group mastered more target skills compared to the control group over 1 month. Dixon et al. (2019) were the first to implement the full PEAK curriculum with all modules and showed an increase in intelligence test scores only when all four modules were implemented for 4 hr per week over 3 months. Despite positive results, these earlier studies did not include measures of social validity to determine whether the training led to meaningful and perceivable changes in children's use of language at school or at home, as producing such changes is a stated purpose of the curriculum (Dixon, 2016).

The present study attempted to replicate the results of the previous research using the recently developed PEAK Comprehensive Assessment (PCA) as a measure of broader changes in language learning. Unlike prior PEAK assessments, the PCA is standardized and directly administered. The PCA may also provide a more direct analysis of these four learning repertoires compared to alternative measures such as intelligence test scores. An additional benefit of the PCA is that the items are not directly targeted within the PEAK curriculum, which is a notable limitation in the studies reported by McKeel et al. (2015). In addition, a social validity measure was constructed for the purpose of the present study to evaluate whether perceptible changes in language or challenging behavior were observed by staff and parents relative to prior semesters, when the PEAK curriculum was not implemented.

Method

Participants and Setting

Participants were five students with autism (one female and four males, with ages ranging from 5 to 13 years). Stephanie was 10 years old, and her score on the indirect PEAK-DT assessment fell within the normative range of a typically developing 5- or 6-year-old using normative data reported by Dixon et al. (2014). Her initial PCA score was 234. Jim was 5 years old, and his indirect PEAK-DT assessment score fell within the normative range of a typically developing 3- or 4-year-old. He scored a 48 on the PCA. Jeremiah was 8 years old, and his indirect PEAK-DT score fell within the normative range of a typically developing 3- or 4-year-old. He scored a 58 on the PCA. David was 6 years old, and his indirect PEAK-DT score also fell within the normative range of a typically developing 3- or 4-year-old. He scored a 60 on the PCA. Finally, Thomas was 13 years old, and his indirect PEAK-DT fell within the normative range of a typically developing 3- or 4-year-old. He scored a 37 on the PCA.

The study took place over a 3-month period within a special education classroom. The classroom was in a public

school and had three separate rooms. PEAK trials were run in the classroom in one of the three rooms away from the other participants, either in a separate room or in a room separated by a room divider. PEAK sessions were administered on average three times per day over the course of the 3 months with an average training dosage of 1.5 hr per day. This represents a greater overall dosage than that reported in the Dixon et al. (2019) and McKeel et al. (2015) randomized control trials. The duration of the intervention was equal to the study conducted by Dixon et al. (2019).

Materials

PEAK Comprehensive Assessment

Initial assessments included the PCA and the PEAK-DT and PEAK-G indirect assessments. The PCA includes preassessments from each of the four PEAK modules. The PEAK-DT and PEAK-G subtests each contain 64 items. For these subtests, the assessor presents a stimulus in a standardized flip book along with verbal instructions. If the learner engages in the correct response, they receive one point in the PCA. Direct-training items are exemplar items from the full PEAK-DT indirect assessment that would have likely been learned through direct reinforcement. Generalization items are exemplar items from the full PEAK-G indirect assessment that represent generalized skills.

The PEAK-E subtest contains 24 items taken directly from the original 48-item PEAK-E preassessment used by Dixon, Belisle, and Stanley (2018). In the original preassessment, each item was assessed twice, and this is reduced to one administration in the PCA. All items involve the presentation of arbitrary or unfamiliar relations where a subset of relations is stated and other relations are tested that increase in complexity from reflexive relations through equivalence relations. To ensure our scores aligned with previous work on the PEAK-E preassessment, we multiplied scores on the PEAK-E subtest by 2 within our reported PCA total score. The PEAK-T subtest contains 96 receptive items and 96 expressive items for a total of 192 items. The frame families assessed include coordination, distinction, opposition, comparison, hierarchical, and deictic relations. We then summed all correct responses on the PCA to achieve a total score that was used for pre-post analysis.

PEAK Indirect Assessment (DT + G) and Factor Scoring Grids

Whereas the PEAK-E and PEAK-T subtests can directly inform program selection, the PEAK-DT and PEAK-G subtests only provide a directly tested estimate of performance. Each indirect assessment contains the 184 skills that are targeted within both modules. The list ranges from simple to complex skills, and this sequence is further refined within the PEAK

factor scoring grids. To complete the indirect assessments, we asked the teacher and a paraprofessional who were most familiar with the participant to complete the assessment by indicating whether the participants were capable of performing the skill. Interobserver agreement between the two assessors was evaluated by dividing the number of agreements by the total number of items. Mean interobserver agreement was 92%. Then, PCA subtest scores and indirect assessment scores were used along with the factor scoring grid to select target items.

Intervention Materials

We used the PCA subtests and the PEAK-DT and PEAK-G indirect assessment to select the best-fit 20 target programs from each module that would be appropriate during the intervention. All PEAK programs are implemented using a discrete-trial training arrangement. In PEAK-DT, all correct responses are directly reinforced. In PEAK-G, a subset of trials can result in direct reinforcement of correct responses, and other trials (i.e., generalization test trials) are never reinforced. In PEAK-E, trained relation trials provide direct reinforcement of correct responses, and derived relational test responses are never reinforced. PEAK-T operates similarly to PEAK-E with the addition of transformation test trials that require the participants to engage in new behavior that involves the target relations without reinforcement. PEAK programs correspond directly to the assessment results and provide a goal statement, a list of materials, a task analysis for implementation, and a list of exemplar stimuli. Implementers then record 1 to 30 potential stimuli that are targeted within the program. Implementers can also add stimuli to the program in levels. For example, 10 Level 1 stimuli may be introduced, and once the stimuli are mastered, an additional 10 Level 2 stimuli may be subsequently introduced. In general, 10 Level 1 stimuli were introduced for PEAK-DT programs, followed by 10 Level 2 stimuli once mastered. For generalization, five training stimuli and five testing stimuli were initially introduced (Level 1), followed by a new set of stimuli (Level 2). For equivalence, a class could contain multiple relations (e.g., train A = B and test B = A). In general, we introduced four classes in Level 1 and an additional four classes in Level 2. The number of new "verbal behavior targets" was equal to the number of classes multiplied by the number of relations (e.g., train A = B and test B = A across four classes equals eight target behaviors). This same strategy was used for PEAK-T programs.

In addition to the PEAK programs, materials were developed specific to each program. Stimuli were also individualized based on the age and interests of each participant. All participants underwent the same token economy reinforcement system that involved a 5-token board. Tokens were exchanged for access to preferred items, including access to classroom toys, the swing, and classroom tablets, as well as edible stimuli provided in the classroom or by the caregivers. Reinforcers varied considerably across participants and over time. Implementation fidelity was assessed using the fidelity measure reported by Dixon et al. (2019). For all implementers, we assessed five 10-trial blocks by reviewing videos of staff implementing PEAK programming. Implementation fidelity was 94% (range 88%–100%).

Procedure

At the beginning of the academic semester, we completed the PCA with all participants. The initial assessment served as a pretest measure. Over the course of 3 months, we conducted multiple PEAK programs within a multiple-probe design across PEAK programs as described by Belisle et al. (2021). First, identified programs were tested under baseline conditions (i.e., no prompts, feedback, or reinforcers were used within a 10-trial block). If the participants scored below 80% correct responding, a second probe was conducted prior to implementing the program. If the score was again below 80%, then the program was part of the intervention. Staggering of the intervention as required within multiple-baseline logic was achieved by testing new programs once earlier programs were mastered. Our goal was to conduct four to eight programs, evenly distributed across the four PEAK modules, simultaneously. The total number of programs targeted throughout the intervention varied as a function of the rate of program mastery, and the total number of skills targeted within each program varied as a function of the module that the program came from, as well as additional considerations from the teacher and the paraprofessionals. A summary of the target programs is provided in the Supplemental Materials. Training was conducted in 10-trial blocks until mastery was achieved for training and testing targets within Level 1 and again for Level 2. The mastery criterion was three consecutive trial blocks with 90% or greater independent correct responding. Training involved the delivery of tokens for correct or prompted responses with prompt fading specific to each program. Five tokens were exchanged for access to preferred reinforcers.

The multiple-probe strategy was conducted over the course of 3 months, followed by a second administration of the PCA. Therefore, the totality of this research design was a multiple probe across skills, replicated across five participants, with a pre-post evaluation of broad repertoire changes as assessed using the PCA. We also had parents and staff complete a social validity measure (see the Supplemental Materials). The questions were related to improvements in language and challenging behavior (parents and staff) and ease of implementation (staff only). It is important to note that the language within the assessment contains "compared to prior semesters," where a score of 5 represented *neutral*, a score of 10 represented *significantly improved*, and a score of 0 represented

significantly less improved. Scores were submitted anonymously to reduce reactivity.

Results and Discussion

The primary results of the study are summarized in Fig. 1. On average, the participants mastered 11.2 programs and gained 186.4 new target skills (all programs contained multiple skill targets) over the course of the study. Stephanie mastered 11 programs with 255 target skills, David mastered 14 programs with 312 target skills, Jim mastered 11 programs with 177 target skills, Thomas mastered 10 programs with 125 target skills, and Jeremiah mastered 10 programs with 123 target skills. The number of days to master new programs varied considerably across participants. Although baseline data are not shown in the figure to reduce the overall number of graphs for interpretation, consider that all programs had baseline scores below 80% across two trial probes. Mastery was achieved across two levels for all programs.

Pre-post PCA results are also shown in Fig. 2. As can be seen, all participants showed an improvement on the PCA from pretest to posttest. Stephanie's score increased to 302 on the postassessment, representing a 29% increase. David scored a 94 on the postassessment, representing a 56.6% increase. Jim scored a 96 on the postassessment, representing a 100% increase. Thomas scored a 78 on the postassessment, representing 110.8% increase. Finally, Jeremiah scored a 65 on the postassessment, representing a 12% increase. Because this is the first study conducted with pre-post PCA performance, we do not know how these results compare to other research; however, one advantage of the PCA is that it provides a standardized metric of skill acquisition that can be
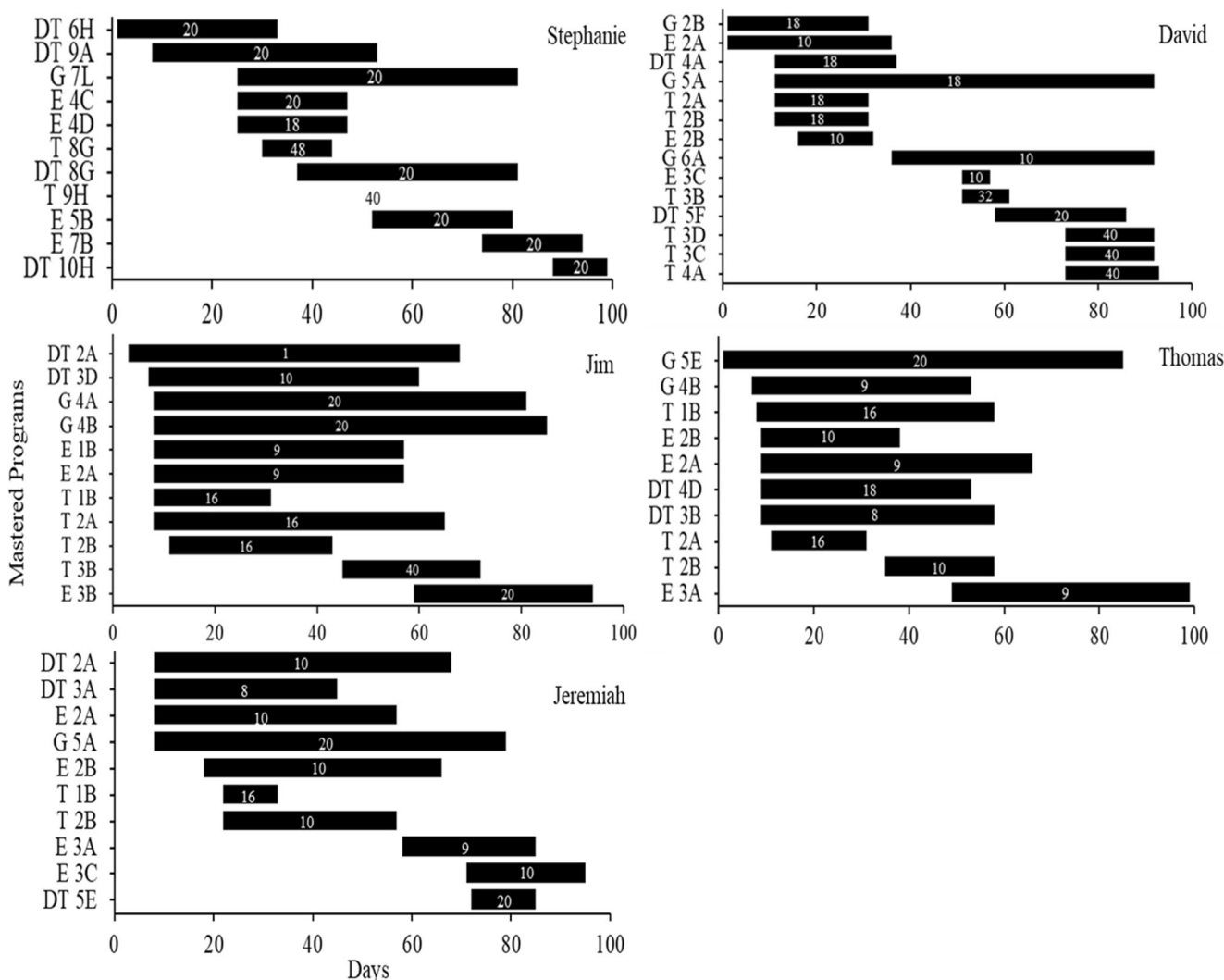
Fig. 1 Multiple-probe across-skills design replicated with the five participants. *Note.* Bars represent the duration from the onset of training to program mastery. Numbers indicate the total number of stimuli or relations targeted within the program. DT = PEAK Direct Training Module; G = PEAK Generalization Module; E = PEAK Equivalence Module; T = PEAK Transformation Module
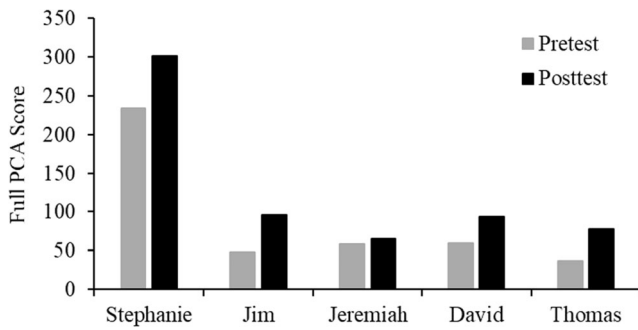
Fig. 2 Pre-post PEAK comprehensive assessment (PCA) scores across the participants in the study

compared across studies, intervention strategies, and implementation differences such as dosage and duration.

Social validity scores also supported the intervention in promoting language performance as reported by parents and staff

relative to prior semesters, and these results are summarized in Fig. 3. The mean rating across the five language items for parents was 8 out of 10, where a score of 5 would suggest no perceptible change. Therefore, we can conclude that parents felt that their children's language skills improved as a result of the use of the curriculum. Staff also reported a similar mean rating, 7.7, suggesting that language skills may have increased in the school setting as well compared to prior semesters. Only parents reported a significant improvement in challenging behavior, with a mean rating of 7.2. One potential explanation for this finding is that the programming may have evoked challenging behavior within the learning environment. It is important to note that reducing challenging behavior is not a stated purpose of the PEAK system. Staff additionally supported that the intervention was easy to administer compared to strategies used in prior semesters. We believe this finding is important because PEAK uses advanced strategies such as equivalence-based instruction. These
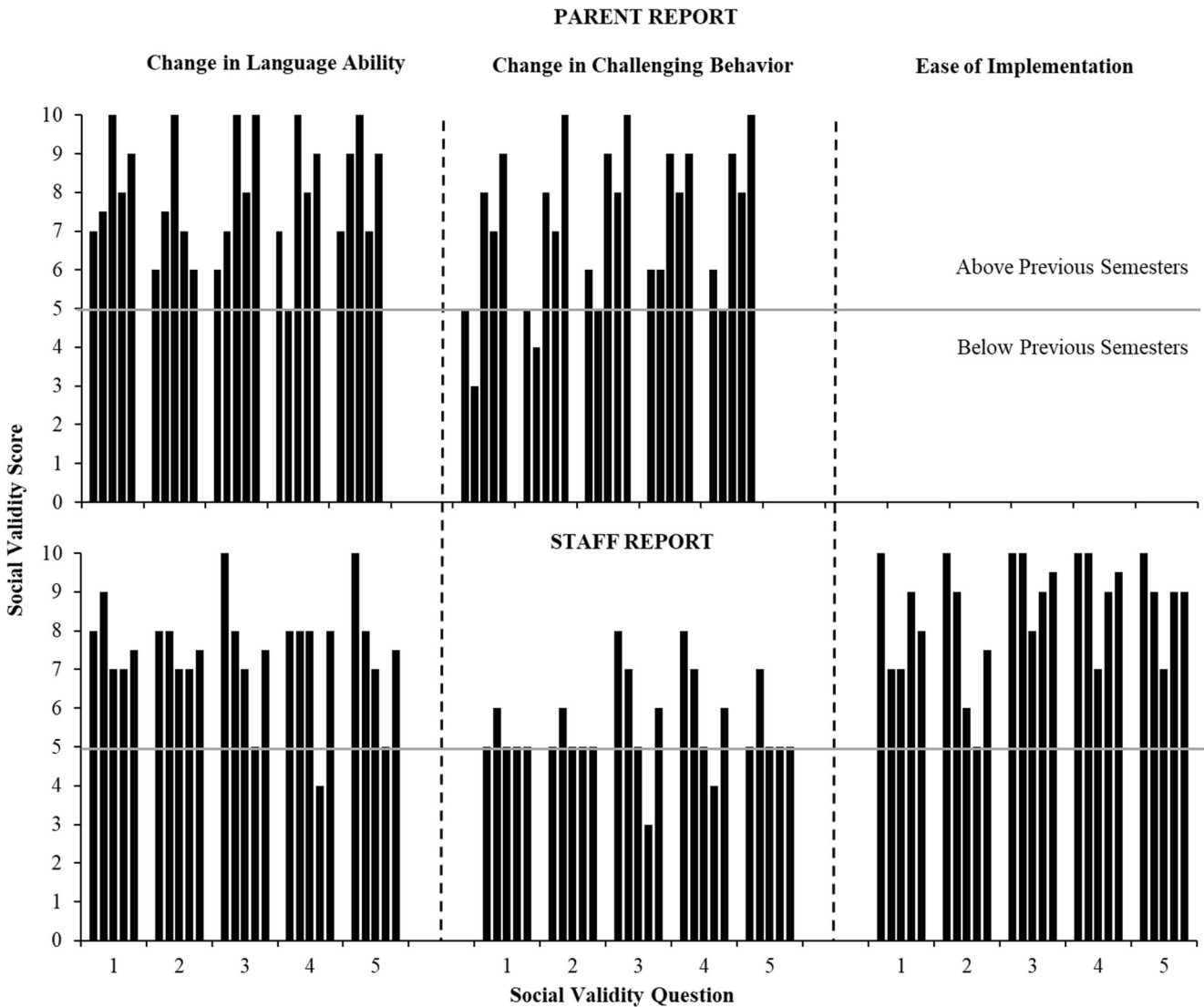
Fig. 3 Social validity scores across each of the five participants (Individual Bars) as reported by parents (Top) and by staff (Bottom). *Note.* Social validity questions are provided in the Supplemental Materials

results may further support the use of this social validity tool developed for this study as being sensitive enough to capture observed changes in language and challenging behavior as a result of PEAK implementation.

Results of the study should be interpreted as preliminary and as an extension of prior work due to several limitations. First, the results were replicated across five participants, and there was not a control group for pre-post PCA comparison. Although results reported by McKeel et al. (2015) and by Dixon et al. (2019) may allow some certainty that individual programs may not have been mastered without intervention, we do not know the degree to which PCA scores change due to testing effects or maturation. For this reason, this study should be viewed as an applied replication of prior work. Second, we did not calculate interobserver agreement on the scores obtained during discrete-trial training. This may become more common as we attempt to increasingly evaluate broader treatment outcomes, moving away from some lower level analyses that may be less common in applied settings. However, future researchers may embed interobserver agreement probes throughout the intervention to improve confidence in the obtained results.

Beyond addressing these limitations, future research may also use measures like the PCA that capture broader skill repertoires to determine factors that are most predictive of treatment success. Whereas the research conducted by Lovaas and colleagues (e.g., Lovaas & Smith, 1987) under intensive intervention conditions produced robust changes in intellectual functioning, results from this study as an applied replication of results reported by Dixon, Paliliunas et al. (2019) suggest that impactful gains may be achieved given fewer training hours and when embedded in a therapeutic or special education setting as was done in the present study.

Declarations

Conflict of interest The authors declare that they have no 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.

References

Ackley, M., Subramanian, J. W., Moore, J. W., Litten, S., Lundy, M. P., & Bishop, S. K. (2019). A review of language development protocols for individuals with autism. *Journal of Behavioral Education, 28*(3), 362–388. https://doi.org/10.1007/s10864-019-09327-8

Belisle, J., Clark, L., Welch, K., & McDonald, N. R. (2021). Synthesizing the Multiple-Probe Experimental Design With the PEAK Relational Training System in Applied Settings. *Behavior Analysis in Practice, 14*(1), 181–192. https://doi.org/10.1007/s40617-020-00520-0

Dixon, M. R., Belisle, J., McKeel, A., Whiting, S., Speelman, R., Daar, J. H., & Rowsey, K. (2017). An internal and critical review of the PEAK Relational Training System for children with autism and related intellectual disabilities: 2014–2017. *The Behavior Analyst, 40*, 493–521. https://doi.org/10.1007/s40614-017-0119-4

Dixon, M. R. (2016). PEAK Transformation Module. Shawnee Scientific Press.

Dixon, M. R., Belisle, J., & Stanley, C. R. (2018). Derived relational responding and intelligence: Assessing the relationship between the PEAK-E pre-assessment and IQ with individuals with autism and related disabilities. *The Psychological Record, 68*(4), 419–430.

Dixon, M. R., Belisle, J., Whiting, S. W., & Rowsey, K. E. (2014). Normative sample of the PEAK relational training system: Direct training module and subsequent comparisons to individuals with autism. *Research in Autism Spectrum Disorders, 8*(11), 1597–1606.

Dixon, M. R., Paliliunas, D., Barron, B. F., Schmick, A. M., & Stanley, C. R. (2019). Randomized controlled trial evaluation of ABA content on IQ gains in children with autism. *Journal of Behavioral Education. Advance online publication.* https://doi.org/10.1007/s10864-019-09344-7

Granpeesheh, D., Tarbox, J., & Dixon, D. R. (2009). Applied behavior analytic interventions for children with autism: A description and review of treatment research. *Annals of Clinical Psychiatry, 21*, 162–173.

Lovaas, O. I., & Smith, T. (1989). A comprehensive behavioral theory of autistic children: Paradigm for research and treatment. *Journal of Behavior Therapy and Experimental Psychiatry, 20*, 17–29. https://doi.org/10.1016/0005-7916(89)90004-9

McKeel, A. N., Dixon, M. R., Daar, J. H., Rowsey, K. E., & Szekely, D. (2015). 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

Smith, T., & Lovaas, I. O. (1998). Intensive and early behavioral intervention with autism: The UCLA Young Autism Project. *Infants & Young Children, 10*, 67–78.