PEAK Pre-Assessments: Preliminary Evidence Establishing Internal Consistency and Construct Validity

Brandon K. May [1] · Lisa Flake [1]

Abstract

Promoting the Emergence of Advanced Knowledge (PEAK) is a behavior-analytic tool that assesses and teaches language and cognitive abilities. PEAK preassessment total scores showed statistically significant correlations with measures of intelligence ($r = .703$, $p = .023$) and adaptive behavior ($r = .618$, $p = .018$), whereas no significant correlations were found between PEAK and age, autism diagnostic instruments, or aggression scales in a sample ($N = 18$) receiving behavior-analytic assessment in an autism clinic. Statistically significant correlations were found between all modules within the PEAK system ($p \leq .001$). Results provide preliminary evidence of the construct validity and internal consistency of the PEAK preassessments.

Keywords PEAK · Construct validity · Autism · Language assessment

In the United States, autism spectrum disorder (ASD) prevalence rates are estimated at 1 in 59, representing a 150% increase from 2000 to 2018 (Baio et al., 2018). The average lifetime cost of treating an individual with ASD is between $1.4 million and $2.4 million, with the largest contributors being special education, residential care, personal/parental productivity loss, and medical expenses (Buescher, Cidav, Knapp, & Mandell, 2014). Behavior-analytic interventions are the most frequently used treatment for individuals with ASD (Hess, Morrier, Heflin, & Ivey, 2008). Although generally accepted as an evidence-based treatment for ASD, traditional assessments and curricula in applied behavior analysis (ABA) have not been validated scientifically (Malkin, Dixon, Speelman, & Luke, 2017).

Construct validity is a statistical analysis of convergent and discriminant validity, designed to study the link between a new measurement tool and a well-established measurement tool. *Convergent validity* refers to the correlation between two assessments that are hypothesized to measure the same construct. *Discriminant validity* refers to the lack of statistical correlation between two assessments that are hypothesized to be unrelated. Although a variety of language-based assessments and curricula are available to behavior analysts, only

recently has the construct validity of these assessments been studied (Malkin et al., 2017). Behavior analysts must continue to study these connections in order to strengthen the evidence base of behavior-analytic assessment and treatment and to broaden their acceptance across scientific disciplines.

The Promoting the Emergence of Advanced Knowledge (PEAK) Relational Training System is one example of a behavior-analytic protocol that combines assessment, a curriculum, and data tracking. PEAK has demonstrated internal consistency (e.g., Rowsey, Belisle, & Dixon, 2015) and convergent validity with measures of adaptive behavior (Malkin et al., 2017), language (e.g., McKeel, Rowsey, Dixon, & Daar, 2015), and intelligence (Dixon, Whiting, Rowsey, & Belisle, 2014). Although research has validated the internal structure of the PEAK Direct Training (DT) and Generalization (G) modules, this is the first study to evaluate the internal consistency across the PEAK system. This article also provides preliminary evidence of construct validity between the PEAK preassessments and standardized assessments, providing the first evaluation of their clinical utility.

Method

Participants and Setting

Participants were evaluated in an outpatient autism clinic that offers diagnostic assessment, psychiatric care, genetic

✉ Brandon K. May
maybk@wustl.edu

1   Department of Psychiatry, Washington University in St. Louis, 4444 Forest Park Ave. Suite 2600, St. Louis, MO 63110, USA

screening, neuropsychological assessment, and behavior analysis; 18 participants (6 females, 12 males) were referred to the ABA clinic by their psychiatrist. The age range of the participants was 3–18 years ($M = 9$, $SD = 4.75$). Of the 18 participants, 16 were diagnosed with ASD, whereas the remaining 2 participants had a primary diagnosis of attention deficit hyperactivity disorder (ADHD). Comorbid diagnoses were indicated for eight participants. Five participants were diagnosed with a comorbid ADHD diagnosis (two of which also had an additional mood disorder diagnosis) and three participants with a comorbid anxiety disorder. IQ scores ranged from 75 to 134 ($M = 104$, $SD = 20.3$). PEAK assessments were completed in a 10 ft × 10 ft (3 m × 3 m) office with a couch, two chairs, and a desk. All assessments were completed by Board Certified Behavior Analysts and/or a Board Certified Assistant Behavior Analyst.

Procedure

Participants were assessed using the psychological instruments outlined in the following sections. IQ measures were obtained through record review. As part of initial assessment in the ABA clinic, the four PEAK preassessments were administered to each participant. Each preassessment was administered using the corresponding flip-book with standard stimuli and scripts. Responses were scored using the accompanying preassessment scoring sheets. For the PEAK-DT and PEAK-G preassessments, every item was presented and scored (Dixon, 2014a, b). For the PEAK Equivalence (E) and PEAK Transformation (T) preassessments, a section was terminated contingent on three consecutive incorrect responses (Dixon, 2015, 2016). Questions were repeated once if no answer was provided or if the participant requested, but scripts were not rephrased or repeated following an incorrect response. Responses were scored as correct if they matched the answers provided in the answer bank of the PEAK scoring sheet or at the assessor's discretion. No feedback was provided during the assessment. Noncontingent reinforcement was provided as needed throughout the assessment. Sessions ranged from 10 to 45 min.

Materials

PEAK Relational Training System PEAK is composed of four separate modules, each addressing 184 skills of increasing difficulty. The four modules include PEAK-DT, PEAK-G, PEAK-E, and PEAK-T. PEAK-DT and PEAK-G focus on foundational language, academics, and social skills, whereas the PEAK-E and PEAK-T modules evaluate relational abilities. The PEAK-T preassessment includes both receptive (PEAK-TR) and expressive (PEAK-TE) domains. Each module contains a separate full assessment and preassessment. Preassessments are abridged versions of each module's full

assessment and are administered using manualized flip-books. Each flip-book contains standard visual stimuli for the assessor to present and uniform scripts for the assessor to read.

Adaptive Behavior Assessment System-3 (ABAS-3) The ABAS-3 is a caregiver report used to assess adaptive skills, assist in diagnostic evaluations, aid in treatment planning, and document progress (Harrison & Oakland, 2015). The ABAS-3 assesses 11 skill areas—grouped into three domains, Conceptual (Conc), Social (Soc), and Practical (Prac)—and provides a total score called the general adaptive composite (GAC). Standard scores are normally distributed with an average score of 100 and a standard deviation of 10, with higher scores indicating more adaptive skills.

Childhood Autism Rating Scale–II (CARS-II) The CARS-II is one of the most widely used and empirically validated autism diagnostic tools (Schopler & Van Bourgondien, 2010). The CARS-II assesses 15 functional areas associated with ASD symptoms. Each area is scored on a Likert-type scale to indicate severity of impairment. For individuals 12 and younger, scores of 15–29.5 indicate minimal symptoms, scores of 30–36.5 indicate mild to moderate symptoms, and 37 and higher indicate severe symptoms of ASD. For individuals 13 and older, scores of 15–27.5 indicate minimal symptoms, scores of 28–34.5 indicate mild to moderate symptoms, and scores of 35 and higher indicate severe symptoms.

Social Responsiveness Scale (SRS) The SRS is a 65-question caregiver report that identifies social impairment associated with ASD and quantifies its severity (Constantino, 2005). A total standard score of 76 or higher indicates significant ASD symptoms. Standard scores of 66–75 indicate moderate clinical symptoms. Standard scores of 60–65 are in the mild range, and standard scores of 59 and below are considered within the typical range.

IQ IQ tests are standardized assessments that measure cognitive ability. IQ scores are normally distributed with an average score of 100 and a standard deviation of 15, with higher scores indicating higher levels of intelligence. Because IQ measures were obtained via record review, scores were only available for a subset of participants. A variety of intelligence measures were used, including the Weschler Intelligence Scale for Children, Fourth Edition ($N = 1$); the Weschler Intelligence Scale for Children, Fifth Edition ($N = 1$); the Reynolds Intellectual Assessment Scale ($N = 1$); the Comprehensive Test of Nonverbal Intelligence, Second Edition ($N = 1$); the Wechsler Nonverbal Scale of Ability ($N = 1$); the Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition ($N = 3$); and the Kaufman Assessment Battery for Children, Second Edition ($N = 2$).

Modified Overt Aggression Scale (MOAS) The MOAS is a caregiver rating scale that measures four categories of aggressive behavior: verbal aggression, property destruction, self-injurious behavior, and physical aggression (Yudofsky, Silver, Jackson, Endicott, & Williams, 1986). Each category contains five statements, rating the intensity of each target behavior from 0 to 5. Items on each scale are summed to yield a maximum score of 10 per scale.

Data Analysis

Statistical analyses were performed using Microsoft Excel 2011 for Mac. Pearson's correlations and significance values were calculated for the total PEAK score and each score on the standardized battery of assessments and age. Pearson's correlations and significance values were also calculated between individual preassessments in the PEAK system.

Results

Table 1 displays the relationship between total PEAK score and the battery of psychological measures. Correlational coefficients and significance values are included for reference. Total PEAK score was strongly and significantly correlated with IQ ($r = .703$, $p = .023$) and moderately and significantly correlated with the ABAS-3 GAC ($r = .618, p = .018$), Conc ($r = .67, p = .018$), and Prac ($r = .531, p = .041$) scores. Figure 1 provides a scatterplot depicting the correlation between total PEAK score, IQ, and ABAS-3 GAC score. Statistically significant correlations were not found between PEAK and age ($r = .351$, $p = .140$), the ABAS-3 Soc ($r = .386$, $p = .155$), the CARS-II ($r = .390$, $p = .372$), the SRS ($r = .248$, $p = .210$), or the MOAS ($r = .003$, $p = .992$) scores.

Table 2 displays the relationship between PEAK preassessment modules. A strong and statistically significant correlation was found between all preassessment modules within the PEAK system ($p \leq .001$).

Discussion

The structure of the PEAK preassessments offers two distinct advantages over other behavior-analytic assessments and the full PEAK assessments. First, by providing standardized stimuli and scripts, the assessor provides more consistent administration within and between clients. Second, the preassessments represent efficient methods of skill assessment that are directly tied to the attached curriculum. The efficiency of the preassessments is achieved not just by abridging the full assessments. The addition of the stimuli flip-books allows the assessor to complete a full assessment of language/cognitive

abilities with just a set of binders or a tablet rather than an entire kit of materials.

Results from this study replicate and extend previous research on the internal structure and construct validity of PEAK. Dixon et al. (2014) demonstrated significant correlations between the full PEAK-DT assessment and IQ. Results from the current study demonstrate similar results using the four PEAK preassessments. Malkin et al. (2017) showed statistically significant correlations between the PEAK-DT and the Vineland Adaptive Behavior Scales, Second Edition. The current study extended this result by demonstrating statistically significant correlations between PEAK preassessment total score and a different scale of adaptive behavior, the ABAS-3, on its total GAC score, as well as two of three subscales. The results listed previously demonstrate preliminary evidence of convergent validity of the PEAK preassessments.

PEAK is designed as a developmental assessment of language and cognitive ability. For individuals that develop typically, language/cognitive assessments correlate to age (e.g., Rowsey, Belisle, & Dixon, 2015). However, for individuals diagnosed with ASD, for which communication is a core deficit, research has demonstrated that age does not correlate to language/cognitive ability. The current study replicated the finding that PEAK score is not correlated with age in the ASD population. This investigation also provides initial evidence that PEAK is not correlated with diagnostic measures of ASD in participants diagnosed with autism. Finally, PEAK was not associated with a standard aggression scale. Results demonstrating no correlation between PEAK preassessment total score and age, ASD diagnostic instruments, or aggression scales provide preliminary evidence for discriminant validity. However, results should be interpreted with caution, as a non-ASD control group was not included in this sample.

The results outlined previously provide preliminary evidence that the PEAK preassessments demonstrate construct validity and may be a valid instrument to measure developmental level regardless of diagnosis or age. Although this article provides additional evidence of the clinical utility of the PEAK preassessments, results should be interpreted with caution. A large and representative sample is critical for statistical analysis in group-design experimental analyses. The small sample presented in this article may lead to data that are not representative of the larger population, thus limiting external validity. Connecting behavior-analytic assessment tools with standardized psychological measures used across other scientific disciplines would provide a method for behavior analysis to extend its reach into other clinical domains. Future research should replicate these correlational analyses with much larger and targeted samples.

Evidence (e.g., Cassidy, Roche, & Hayes, 2011) has demonstrated that relational training may improve scores on standardized intelligence measures. Given the links between PEAK and IQ measures (Dixon et al., 2014) and the evidence

Table 1 The Relationship Between Total PEAK Score and the Battery of Psychological Measures, Along With Basic Descriptive Data

	PEAK Preassessment Total Scores (N = 18)		Number of Scores per Instrument (N)	Mean Age (years)	Males	Females
	r value	p value				
Age	.351	.140	18	9.4	12	6
ABAS-3 GAC	.618	.018	13	8.2	10	3
ABAS-3 Conc	.67	.018	13	8.2	10	3
ABAS-3 Soc	.386	.155	14	8.4	10	4
ABAS-3 Prac	.531	.041	14	8.4	10	4
CARS-II	.390	.372	12	9.5	7	5
SRS	.248	.210	15	9	9	6
MOAS	.003	.992	13	9.3	8	5
IQ	.703	.023	10	9.8	8	2

that relational training may improve performance on cognitive assessments, future research should evaluate whether training using the PEAK curriculum would yield improvements on IQ tests. Limited research is available regarding improvements in adaptive behavior scales following behavior-analytic training. Because multiple studies have now demonstrated statistically significant relationships between PEAK and adaptive behavior measures, it would follow that treatment using the PEAK curriculum could lead to improvements on these standardized developmental measures. This would be an important next step in demonstrating generalized outcomes for the PEAK curriculum and behavior analysis in general.

A within-system analysis of PEAK preassessments demonstrated strong and statistically significant correlations across all modules. These results extend research demonstrating the internal consistency of PEAK. Strong correlations have been documented within the PEAK modules, but this is the first study to review the psychometric properties between PEAK modules. Each module is designed to measure unique skills ranging from foundational to complex relational. Research on the emergence patterns of derived relations has not been well studied. Statistical analyses comparing foundational skills to

the development of relational abilities may have a significant impact on knowledge of the development of relational skills.

The correlations between PEAK modules in this sample suggest that future research should investigate whether the preassessments can be shortened even further and yield meaningful results. No research is currently available regarding the best practice of linking assessment to the curriculum in the PEAK preassessments. Identifying the most effective and efficient manner to link PEAK preassessment results to the curriculum should be an area of future study. Future research should also compare the PEAK preassessments to the PEAK full assessments to evaluate whether results from the preassessments are truly representative of total PEAK score using the full assessments, which to date have been studied more extensively.

The greatest limitation of this study is the relatively small sample size. In addition, the range of IQ scores and age are large, which may affect results. Again, future research should replicate these correlational analyses with much larger and targeted samples.

Implications for Practice

- The PEAK preassessments are manualized, abridged versions of the four PEAK assessment modules designed to
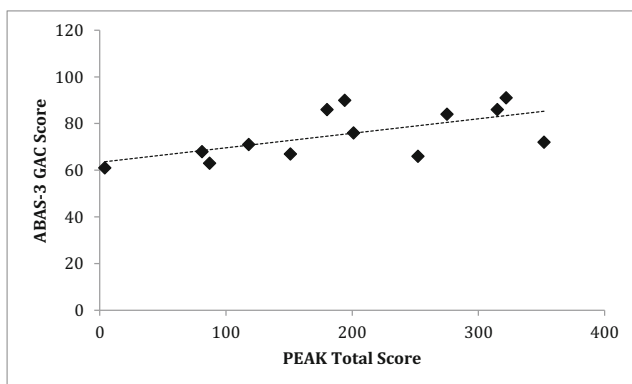
Fig. 1 Scatterplot depicting correlations between each participant's total score on the PEAK preassessments, IQ, and the ABAS-3 GAC

Table 2 The Correlational Coefficient (r) and Statistical Significance (p) Between PEAK Preassessment Modules

	G		TE		TR		E	
	r-value	p-value	r-value	p-value	r-value	p-value	r-value	p-value
DT	.899	<.001	.722	<.001	.761	<.001	.773	<.001
G	1	NA	.909	<.001	.875	<.001	.898	<.001
TE	.909	<.001	1	NA	.909	<.001	.948	<.001
TR	.875	<.001	.908	<.001	1	NA	.900	<.001
E	.898	<.001	.948	<.001	.900	<.001	1	NA

provide efficient evaluation of skill level across language and cognitive constructs.

- This initial evaluation of the PEAK preassessments establishes preliminary evidence of internal consistency across the four modules.

- The PEAK-DT assessment has been correlated with measures of adaptive behavior and intelligence; however, this investigation is the first to demonstrate preliminary data toward construct validity between standardized psychological measures and preassessment total scores across the full PEAK system.

- Establishing the construct validity of the PEAK preassessments may provide practicing behavior analysts with confidence in its clinical utility and lead to wider acceptance of behavior-analytic assessment across scientific orientations.

Compliance with Ethical Standards

Conflict of Interest All authors declare they have no conflicts of interest.

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

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

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

Baio, J., Wiggins, L., Christensen, D., Maenner, M. J., Daniels, J., Warren, Z., & Dowling, N. F. (2018). Prevalence of autism spectrum disorder among children aged 8 years: Autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *Morbidity and Mortality Weekly Report, 67*(6), 1–23. https://doi.org/10.15585/mmwr.ss6706a1.

Buescher, A. V., Cidav, Z., Knapp, M., & Mandell, D. S. (2014). Costs of autism spectrum disorders in the United Kingdom and the United States. *Journal of American Medical Association Pediatrics, 168*(8), 721–728. https://doi.org/10.1001/jamapediatrics.2014.210.

Cassidy, S., Roche, B., & Hayes, S. C. (2011). A relational frame training intervention to raise intelligence quotients: A pilot study. *The Psychological Record, 61*(2), 173–198. https://doi.org/10.1007/BF03395755.

Constantino, J. N. (2005). *The social responsiveness scale.* Retrieved from https://www.wpspublish.com/store/p/2993/srs-social-responsiveness-scale

Dixon, M. R. (2014a). *PEAK relational training system: Direct training module.* Carbondale, IL: Shawnee Scientific Press.

Dixon, M. R. (2014b). *PEAK relational training system: Generalization module.* Carbondale, IL: Shawnee Scientific Press.

Dixon, M. R. (2015). *PEAK relational training system: Equivalence module.* Carbondale, IL: Shawnee Scientific Press.

Dixon, M. R. (2016). *PEAK relational training system: Transformation module.* Carbondale, IL: Shawnee Scientific Press.

Dixon, M. R., Whiting, S. W., Rowsey, K., & Belisle, J. (2014). Assessing the relationship between intelligence and the PEAK relational training system. *Research in Autism Spectrum Disorders, 8*(9), 1208–1213. https://doi.org/10.1007/s10864-015-9219-y.

Harrison, P., & Oakland, T. (2015). *The adaptive behavior assessment system* (3rd ed.). Retrieved from https://www.pearsonclinical.com/psychology/products/100001262/adaptive-behavior-assessment-system-third-edition-abas-3.html.

Hess, K. L., Morrier, M. J., Heflin, L. J., & Ivey, M. L. (2008). Autism treatment survey: Services received by children with autism spectrum disorders in public school classrooms. *Journal of Autism and Developmental Disorders, 38*(5), 961–971. https://doi.org/10.1007/s10803-007-0470-5.

Malkin, A., Dixon, M. R., Speelman, R. C., & Luke, N. (2017). Evaluating the relationship 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*(2), 341–351.

McKeel, A., Rowsey, K., Dixon, M. R., & Daar, J. H. (2015). Correlation between PEAK relational training system and one-word picture vocabulary tests. *Research in Autism Spectrum Disorders, 12*, 34–39.

Rowsey, K. E., Belisle, J., & Dixon, M. R. (2015). Principal component analysis of the PEAK relational training system. *Journal of Developmental and Physical Disabilities, 27*(1), 15–23. https://doi.org/10.1007/s10882-014-9398-9.

Schopler, E., & Van Bourgondien, M. E. (2010). *The childhood autism rating scale* (2nd ed.). Retrieved from https://www.wpspublish.com/store/p/2696/cars-2-childhood-autism-rating-scale-second-edition

Yudofsky, S. C., Silver, J. M., Jackson, W., Endicott, J., & Williams, D. (1986). The overt aggression scale for the objective rating of verbal and physical aggression. *American Journal of Psychiatry, 143*, 35–39.