




An Initial Evaluation of an Assessment Method for the PEAK Relational System Direct Training Module

James W. Moore¹  · Hayden B. Rizer¹ · Lana K. Warren¹ · M. Garrett Yeager¹ · Robyn R. Brewer¹ · Laura K. Barker¹ · BreAnna Newborne¹ · Kayla L. Russo¹ · Madeline Potter¹

Published online: 18 October 2019

© Association for Behavior Analysis International 2019

Abstract

When considering the best available treatment, it is crucial that assessments yield valid and reliable measures to produce effective treatment options. Currently, this research is limited, giving behavior analysis a platform to evaluate the psychometric properties and content validity of assessment tools used across settings. One major issue that practicing behavior analysts face is conducting the most comprehensive assessment within the time constraints put in place by insurance companies. Utilizing a method of assessment that includes indirect and direct descriptive methods and experimental manipulations could aid in cutting down assessment time, especially if those methods have known correspondence with each other. The purpose of the present study was to assess the components of the *Promoting the Emergence of Advanced Knowledge Relational Training System: Direct Training Module* (PEAK-DT) for children with autism. More specifically, this study evaluated the correspondence between the PEAK indirect assessment (PEAK-IA) and PEAK preassessment (PEAK-PA) for the Direct Training Module. Comparisons were also made to determine which method offers the best predictive validity of actual performance on the PEAK-DT module. Results indicate that PEAK-IA completed by parents and PEAK-PA share moderate correspondence, with the PEAK-PA offering the strongest predictive validity of direct testing outcomes. Implications for behavior-analytic practice, as well as directions for future research, are discussed.

Keywords Applied behavior analysis · Autism · PEAK Relational Training System · Assessment

The Centers for Disease Control and Prevention (2008) has reported a continued rise in the prevalence of autism spectrum disorder (ASD), with current estimates around 1 in 59 live births. As a result, demand for insurance coverage of evidence-based treatments for ASD continues to increase from families across the nation (Roane, Fisher, & Carr, 2016). Applied behavior-analytic (ABA; Baer, Wolf, & Risley, 1968) treatments are no stranger to the field of autism, and it is important for service providers and certified professionals to utilize valid, reliable, and effective methods. Behavior analysts must continually explore the issue of evidence-based practice when working with children with ASD across a number of domains, including issues related

to best practice assessment methods (Gould, Dixon, Najdowski, Smith, & Tarbox, 2011; Smith, 2013; Smith & Iadarola, 2015). This is of particular importance when a majority of autism insurance statutes explicitly mandate coverage of treatments that are based on the principles of ABA (Roane et al., 2016). The National Institutes of Health (2017) reiterates this importance of professionals applying the best available evidence for decision-making and designing programs and interventions. Providing effective, evidence-based treatment plans requires a closer review of assessment instruments to ensure that treatments are empirically sound.

For practitioners, an additional concern regarding the efficiency of assessment procedures must also garner consideration. In January 2019, the Current Procedural Terminology Codes that govern billing for ABA were updated. Among some of the major changes include moving the current Category III codes, which were temporary, to Category I, or permanent codes. In terms of language and communication assessment, two major codes apply: Behavior Identification Assessment (97151) and Behavior

✉ James W. Moore
james.moore@mycanopy.org

¹ Canopy Children's Solutions, 1465 Lakeland Drive, Jackson, MS 39216, USA

Identification Supporting Assessment (97152). These codes are billed in 15-min units and must apply to all facets of assessment, including language and communication. As of now, there are no clear guidelines regarding billing restrictions practitioners will face, but obviously the use of effective and efficient assessment methods is a major concern.

With the increasing trend toward earlier identification and diagnosis of ASD (Daniels & Mandell, 2014), it is crucial to investigate ABA assessments that yield valid and reliable measures, leading to the most effective treatment options possible. Currently, this type of research is in its infant stages; therefore, behavior analysis would benefit from work that evaluates both the psychometric properties and the content validity of assessment tools commonly used across settings. A recent review of available published ABA assessment packages found that less than 50% of the programs on the market have data supporting their psychometric properties or treatment efficacy (Ackley et al., *in press* 2019). A lack of published data on the effectiveness, validity, reliability, and efficiency of assessment instruments may compromise a behavior analyst's ability to satisfy the requirement of clients having the right to effective treatment (Behavior Analyst Certification Board, 2014a, Section 2.09), with specific emphasis on assessing behavior (BACB, 2014a, Section 3.0). As the prevalence of ASD continues to rise, so will the demand for effective treatments that target language and learning skills. The types of assessment tools used in behavior analysis warrant further discussion.

The *Promoting the Emergence of Advanced Knowledge Relational Training System* (PEAK; Dixon, 2014a, b), stands as one of the exceptions to the lack of research on assessment outcomes, with PEAK accounting for over 40% of the total studies published across all in-print ABA assessment packages. The PEAK Relational Training System is an evidence-based autism assessment and treatment curriculum divided into four modules: Direct Training (DT; Dixon 2014a), Generalization (G; Dixon, 2014b), Equivalence (E; Dixon, 2015), and Transformation (T; Dixon, 2016). PEAK-DT is exclusively based on Skinner's (1957) verbal behavior model, and one of the learning processes within this module is contingency-based (or directly trained) learning. As a curriculum, PEAK-DT includes 184 unique targets across such operant categories as imitation, echoics, tacting, manding, and intraverbals, to name a few. From an assessment standpoint, PEAK-DT offers multiple methods of assessment to determine the client's current operant level across these categories. The behavior analyst uses these multiple sources of data, including indirect and direct methods, to determine if each of the 184 targets are within the client's behavioral repertoire or not. Dixon, Belisle, Whiting, and Rowsey (2014) reported normative sample scoring on PEAK-DT, with total scores ranging from 0 to 184.

As stated earlier, behavior analysts also face an efficiency challenge, often required in a managed health care world of

conducting the most comprehensive assessment possible under the thin time constraints often placed upon them by insurance companies. According to the *Applied Behavior Analysis Treatment of Autism Spectrum Disorder: Practice Guidelines for Healthcare Funders and Managers* initial assessments for comprehensive treatment could take 20 hr (BACB 2014b). Although directly testing all items within an assessment may yield the absolutely most valid results, the feasibility of such an approach for a practicing behavior analyst is low. For example, just one module of PEAK, such as the DT module, contains 184 unique behaviors. Overall, the four modules of PEAK constitute over 700 individual skills.

Best practice in behavioral assessment suggests collecting data from multiple sources, multiple informants, and using multiple methods (BACB, 2014b). Research over the last 30 years on functional behavior assessment (FBA) might offer a unique solution to the barriers faced by practicing behavior analysts in a managed-care reality. As described by Cooper, Heron, and Heward (2007), three distinct FBA methods are available and can be used to inform overall decisions regarding behavioral function: indirect descriptive assessment, direct descriptive assessment, and experimental manipulations. A similar approach could help cut down assessment time if research suggests good correspondence and predictive validity between methods.

Indirect assessment instruments have been presented as the least intrusive and possibly most cost-effective methods in behavioral assessment because data are gathered apart from the actual behavior of interest. Indirect assessments include interviews, rating scales, checklists, self-report measures, and questionnaires and can be completed in person, via the Internet, or sent home and returned via mail (Floyd, Phaneuf, & Wilczynski, 2005). However, research has questioned the predictive validity of such findings (Thompson & Iwata, 2007). In the present day, managed-care world, behavior analysts must find the most time-efficient yet effective assessment tools to implement during a behavior identification assessment. Research from the area of school psychology has endorsed the use of a data-driven decision-making model that incorporates multiple sources of data (Doggett, Edwards, Moore, Tingstrom, & Wilczynski, 2001; Dufrene, Doggett, Henington, & Watson, 2007; Dufrene et al., 2012; LeGray, Dufrene, Sterling-Turner, Olmi, & Bellone, 2010). Moreover, best practice standards suggest the use of multiple methods, multiple modalities, and multiple informants. ABA is well known for its use of direct observation; however, frontline contingencies are working against this, with only 1 to 2 hr available for billing assessment time. As a result, indirect assessment instruments supplementing direct instruments should and need to be evaluated.

A unique feature of the PEAK Relational Training System assessment involves three sources of data gathering. Each module includes an indirect assessment, which is a yes, no, don't know questionnaire completed by parents or teachers; a preassessment, presented in a quick flip-book format; and then

direct testing of questionable items from the overall module. Recently, the PEAK preassessment (PEAK-PA) has been freely offered as a quick method to determine initial verbal operant levels for children with ASD. It is time efficient and more innovative and practical for a therapist to implement, providing clear instruction and guidance to the practitioner throughout the assessment process. How these two methods may work together in a multimethod assessment is currently unclear. If the combined use of PEAK-PA and PEAK indirect assessment (IA) could offer strong predictive validity of overall module performance, this could offer a significantly efficient model of assessment as opposed to directly testing all 184 items from a PEAK module. However, the PEAK-PA only contains 64 selected items from the overall module, whereas the PEAK-IA has parents or teachers rate learner performance across all module items. Parents often provide input to clinicians upon intake; therefore, it is important to consider the value of their report or how well it corresponds with a preassessment or direct assessment. If good correspondence exists between the PEAK-IA and PEAK-PA, this could offer a powerful tool for behavior analysis, both from an effective and efficiency standpoint.

Behavioral assessments like PEAK could be implemented in a similar manner as data-driven FBAs and conceptualized as a funnel that proceeds from general information (e.g., indirect descriptive data) to specific individualized data (e.g., direct descriptive data). The PEAK-IA and PEAK-PA could work in tandem to help reduce the overall number of targets the practitioner must directly test, but only if these methods produce reliable and valid outcomes. Indirect and direct descriptive data can then be evaluated and used to focus on the number of programs that need thorough direct testing, yielding even more specificity on a client's current behavioral repertoire. Having assessment instruments that are more time efficient and demonstrate higher predictive validity can benefit the overall time spent on assessment while also ensuring that accurate results are being reported. The first step, however, should involve an evaluation of these multiple sources of data. Research on all assessment instruments should be ongoing for behavior analysts to review. In the current article, we evaluated a multimethod module of assessment, using the PEAK-IA and PEAK-PA to determine the correspondence between the PEAK-IA and PEAK-PA for the DT module on the 64 items that correspond to each method.

Method

Participants, Materials, and Setting

Sixteen individuals ranging from 2 to 8 years of age were selected from an early intervention clinic for children with ASD in the southern United States. All individuals had a

previous diagnosis of ASD from a multidisciplinary ASD clinic or from a major hospital in the southern United States. The preassessment sessions were conducted in a therapy room measuring 6.4 meters \times 6.7 meters located within the facility. The preassessment sessions ranged from 15 to 45 min, with an average time of 17 min, 33 s. The indirect assessments were administered to the parents of the individuals that participated in the study to be completed at home and returned to the day treatment facility. In the case of two potential informants for the indirect (e.g., father and mother), these dyads were asked to complete a single indirect assessment together, thus producing one indirect assessment for each participant. Specific demographic information related to each participant and caregiver is outlined in Table 1.

Materials included the PEAK-IA from the PEAK-DT, the PEAK-PA, and various stimulus materials needed for PEAK-DT direct testing. The PEAK-DT indirect assessment is an indirect descriptive assessment method, including questions on all 184 individual skill programs from the overall module. The skill is described, and then informants mark "yes," "no," or "?" Written instructions were sent home to guide parents through the completion of the PEAK-IA, and then a copy of the indirect assessment was sent home with each family and returned prior to the administration of direct PEAK-DT testing. The instructions stated, "Read each question and apply it to your child. If you are *absolutely* sure he/she can do the skill listed, mark 'Yes.' If you are *absolutely* sure he/she cannot do the skill listed, mark 'No.' If you are not sure, mark '?'." At the time of this study, the PEAK-PA was a free download that was printed from the PEAK website (<https://www.peak2aba.com/>) and arranged in a three-ring binder to make a flip-book. One side of the flip-book presented stimulus materials, and the other presented instructions to the assessor, similar to common intelligence tests. The PA contains 64 testing items that correspond with items from the overall module. Table 2 outlines the correspondence between the PA items and the overall PEAK-DT module. Stimulus materials varied based on the items tested for each participant but included many of the items described in the PEAK-DT module.

Response Definitions, Measurement, and Interobserver Agreement

For assessment sessions, data were collected on correct independent responses to the 64 overlapping items from the PEAK-DT module, as measured by PEAK-IA and PEAK-PA. On the PEAK-IA, a correct independent response was defined as the informant marking "yes" to a specific question related to a skill item. If the parent marked "no" or "?" this was considered an incorrect response, meaning the skill was not currently in the child's repertoire. For the PA and direct assessment, a correct independent response was defined as the participant independently selecting or emitting the correct

Table 1 Participant demographic information

Age	Gender	Duration of prior ABA treatment	Past assessments	Hours of therapy per week (prior)	Duration of PEAK assessment
2	Male	None	None	0	2 hr 15 min
8	Male	3 years 7 months	ABLIS, VB-MAPP	15	2 hr
4	Female	5 months	VB-MAPP	20	2 hr 30 min
7	Male	2 years 8 months	ABLIS, VB-MAPP	15	3 hr
7	Male	1 year 8 months	VB-MAPP	14	3 hr
6	Male	1 year 4 months	VB-MAPP	6	2 hr 45 min
5	Male	1 year 4 months	VB-MAPP	6	2 hr
4	Female	10 months	VB-MAPP	8	2 hr
8	Male	2 years 1 month	VB-MAPP	6	3 hr 15 min
8	Male	1 year 5 months	VB-MAPP	6	3 hr
6	Male	1 year 6 months	VB-MAPP	9	3 hr
7	Male	2 years 1 month	VB-MAPP	N/A	2 hr 30 min
8	Male	2 years 1 month	VB-MAPP	2.5	1 hr
5	Female	1 year 11 months	VB-MAPP	20	1 hr 45 min
9	Female	2 years 0 months	VB-MAPP	4	2 hr
8	Male	2 years 2 months	VB-MAPP	6.5	3 hr 30 min

ABLIS = Assessment of basic language and learning skills; *VB-MAPP* - Verbal behavior milestones assessment and placement program

pictorial, written, or verbal response following the presentation of the discriminative stimulus associated with the trial.

For the PEAK-PA, assessors collected paper-and-pencil data on each trial related to correct and incorrect responding during the 64-item assessment. Four factors were assessed in the DT

module preassessment. These factors were foundational learning skills (FLS; e.g., basic instruction following, modeled responding, elementary verbal operants, and attention to the environment); perceptual learning skills (PLS; e.g., matching, finding objects from an array, naming items, and answering *wh-*

Table 2 Correspondence between PEAK-IA and PEAK-PA items

Foundational learning skills		Perceptual learning skills		Verbal communication skills		Verbal, reasoning, memory, and mathematics skills	
PEAK-IA	PEAK-PA	PEAK-IA	PEAK-PA	PEAK-IA	PEAK-PA	PEAK-IA	PEAK-PA
1A	FLS-1	7D	PCS-1	7H	VCS-1	11M	VMS-1
3C	FLS-2	8A	PCS-2	8H	VCS-2	11D	VMS-2
3D	FLS-3	8B	PCS-3	9H	VCS-3	14A	VMS-3
4A	FLS-4	8C	PCS-4	9P	VCS-4	14C	VMS-4
4B	FLS-5	8D	PCS-5	10A	VCS-5	14E	VMS-5
4C	FLS-6	8F	PCS-6	10C	VCS-6	14F	VMS-6
4D	FLS-7	8I	PCS-7	10J	VCS-7	14H	VMS-7
4F	FLS-8	8K	PCS-8	10Q	VCS-8	14I	VMS-8
5A	FLS-9	8N	PCS-9	11A	VCS-9	14L	VMS-9
5C	FLS-10	9A	PCS-10	11G	VCS-10	14N	VMS-10
6A	FLS-11	9B	PCS-11	11L	VCS-11	14O	VMS-11
6B	FLS-12	9C	PCS-12	12B	VCS-12	14P	VMS-12
6G	FLS-13	9E	PCS-13	12L	VCS-13	14Q	VMS-13
6I	FLS-14	9F	PCS-14	12S	VCS-14	14R	VMS-14
7A	FLS-15	9K	PCS-15	13K	VCS-15	14V	VMS-15
7C	FLS-16	9L	PCS-16	14G	VCS-16	14W	VMS-16

questions); verbal comprehension skills (VCS; e.g., multiple-step instruction following, multiword vocal utterances, concept formations, large intraverbal focus, and social exchanges); and verbal reasoning, memory, and mathematical skills (VMS; e.g., basic logic processes, complex language, and mathematical computation skills). Sixteen items in each factor directly corresponded with items from the DT module (see Table 1). Each preassessment began with the examiner providing an instruction for each item to the participant. The preassessment assessor did not have access to the results of the indirect assessment until after completion of the current study.

Once the PEAK-PA and PEAK-IA were both completed, a Board Certified Behavior Analyst (BCBA™) and Board Certified Behavior Analyst–Doctoral level (BCBA-D™) both compared the results and assessed agreement between the two measures. Table 3 outlines the overall agreement across each participant. An agreement was defined as the PEAK-IA producing a “yes” and the child responding correctly to the same item on the PEAK-PA, or the PEAK-IA producing a “no” and the child not responding correctly to the same item on the PEAK-PA. A disagreement was defined as the PEAK-IA producing a “no” or “?” with the child responding correctly to the same item on the PEAK-PA, or the PEAK-IA producing a “yes” with the child not scoring correctly to the corresponding item on the PEAK-PA. All disagreements were assigned to direct testing, which included the same response definitions as the PA; however, for the item to be considered “yes,” the participant must correctly respond to three presentations of the response item.

Interobserver agreement (IOA) data for correct and incorrect responses were collected during assessment sessions. Trial-by-trial IOA was calculated by comparing observer scores across every trial, with an agreement considered when both observers marked the trial correct or incorrect. Disagreements were considered when one observer scored the trial correct, whereas the other scored the trial incorrect. Agreements were then divided by total trials and then multiplied by 100%. IOA averaged 100% across all sessions.

General Procedures

Indirect Assessment As stated earlier, each family completed one indirect assessment at home and returned to the clinic. The second author, a BCBA-D™, met with each family prior to sending the form home and gave each caregiver the same instructions for completing the indirect assessment. Caregivers were told, “Pay close attention to the information under the ‘Description’ column. Where you see the word ‘participant,’ replace that with your child’s name. For example, for the first item, you would read, ‘After being given a prompt, [child’s name] will make eye contact for two seconds.’ If you are absolutely sure he/she can do this, mark ‘Y.’ If you are absolutely sure he/she cannot do this, mark ‘N.’ If you’re not sure, please don’t try to figure it out, but simply mark ‘?’”

Indirect assessments were completed and returned directly to the second author, who then recorded the caregiver responses that corresponded to the 64 items from the PEAK-PA into a database not accessible by therapists conducting PEAK-PA or direct testing sessions.

Preassessment For all PEAK-PA sessions, the participant and the Registered Behavior Technician (RBT) sat across from each other at a table. The RBT was blind to the results of the indirect assessment. In all cases, the PEAK-PA for PEAK-DT was completed in one session, with sessions ranging from 15 to 45 min. The RBT used a flip-book as described previously as a visual aid for conducting the sessions and collected paper-and-pencil data on participant responses. Each participant sat at a table across from the RBT. The flip-book contained instructions for the administrator to run the session, which were read verbatim, as well as the required two-dimensional visual stimuli that corresponded with each question item for the participant. All sessions included the 64-item assessment. For copyright reasons, we have deleted this appendix.

Direct Testing These sessions were run similarly to the PEAK-PA sessions, except rather than using two-dimensional printed representations of stimuli, these sessions included the use of three-dimensional physical objects. As common to directly assessing the items from the PEAK-DT module, the direct testing sessions included three trials of each item, rather than one (as conducted with PEAK-PA). PEAK-DT direct testing sessions ranged from 15 to 75 min, and the assessor prepared all stimulus materials prior to the session. For the current study, direct assessments involved items that yielded disagreements of any type between the PEAK-IA and PEAK-PA outcomes, as described previously. Agreements between PEAK-IA and PEAK-PA were not tested. A testing script was generated by a BCBA™ or BCBA-D™ to standardize administration across participants. For copyright reasons, we have deleted this appendix.

Data Analysis

We used a variety of statistical and visual analytic methods to (a) assess the relationship between the PEAK-IA and the PEAK-PA; (b) determine if either predictably generated greater estimates of performance; and (c) when disagreements did occur, determine which method was more predictive of true performance. To assess the relationship between both assessment methods, we calculated Pearson’s correlation coefficients that provide a measure of the linear correlation between two variables (PEAK-IA and PEAK-PA). Correlational coefficients for the scores produced in both assessments, as well as within each factor, were obtained along with a measure of statistical significance. In addition, we determined the frequency of disagreements between the two assessments as a percentage. To determine if either assessment generated

Table 3 Pearson's correlational analysis on scores for PEAK-IA and PEAK-PA across factors

		PEAK-IA	FLS PEAK-IA	PCS PEAK-IA	VCS PEAK-IA	VMS PEAK-IA	PEAK-PA	FLS PEAK-PA	PCS PEAK-PA	VCS PEAK-PA	VMS PEAK-PA
PEAK-IA	Pearson correlation	1	.883**	.931**	.916**	.817**	.889**	.804**	.889**	.858**	.673**
	Sig. (2-tailed)		.000	.000	.000	.000	.000	.000	.000	.000	.004
	N	16	16	16	16	16	16	16	16	16	16
FLS PEAK-IA	Pearson correlation	.883**	1	.902**	.666**	.561*	.802**	.836**	.878**	.685**	.494
	Sig. (2-tailed)	.000		.000	.005	.024	.000	.000	.000	.003	.052
	N	16	16	16	16	16	16	16	16	16	16
PCS PEAK-IA	Pearson correlation	.931**	.902**	1	.741**	.620*	.849**	.781**	.912**	.766**	.614*
	Sig. (2-tailed)	.000	.000		.001	.010	.000	.000	.000	.001	.011
	N	16	16	16	16	16	16	16	16	16	16
VCS PEAK-IA	Pearson correlation	.916**	.666**	.741**	1	.831**	.765**	.643**	.703**	.804**	.635**
	Sig. (2-tailed)	.000	.005	.001		.000	.001	.007	.002	.000	.008
	N	16	16	16	16	16	16	16	16	16	16
VMS PEAK-IA	Pearson correlation	.817**	.561*	.620*	.831**	1	.765**	.615*	.668**	.829**	.675**
	Sig. (2-tailed)	.000	.024	.010	.000		.001	.011	.005	.000	.004
	N	16	16	16	16	16	16	16	16	16	16
PEAK-PA	Pearson correlation	.889**	.802**	.849**	.765**	.765**	1	.912**	.947**	.937**	.845**
	Sig. (2-tailed)	.000	.000	.000	.001	.001		.000	.000	.000	.000
	N	16	16	16	16	16	16	16	16	16	16
FLS PEAK-PA	Pearson correlation	.804**	.836**	.781**	.643**	.615*	.912**	1	.846**	.792**	.674**
	Sig. (2-tailed)	.000	.000	.000	.007	.011	.000		.000	.000	.004
	N	16	16	16	16	16	16	16	16	16	16
PCS PEAK-PA	Pearson correlation	.889**	.878**	.912**	.703**	.668**	.947**	.846**	1	.878**	.705**
	Sig. (2-tailed)	.000	.000	.000	.002	.005	.000	.000		.000	.002
	N	16	16	16	16	16	16	16	16	16	16
VCS PEAK-PA	Pearson correlation	.858**	.685**	.766**	.804**	.829**	.937**	.792**	.878**	1	.730**
	Sig. (2-tailed)	.000	.003	.001	.000	.000	.000	.000	.000		.001
	N	16	16	16	16	16	16	16	16	16	16
VMS PEAK-PA	Pearson correlation	.673**	.494	.614*	.635**	.675**	.845**	.674**	.705**	.730**	1
	Sig. (2-tailed)	.004	.052	.011	.008	.004	.000	.004	.002	.001	
	N	16	16	16	16	16	16	16	16	16	16

**Correlation is significant at the 0.01 level (2-tailed)

*Correlation is significant at the 0.05 level (2-tailed)

significantly greater estimates of performance, we conducted a paired-samples *t*-test. This test provided an analysis of the likelihood that any differences were due to chance alone. We also calculated Cohen's *d* as an estimate of effect size. This test provides an estimate of the effect that using either assessment method could have on test performance. Finally, to determine which method was more predictive of true performance, we conducted a chi-squared test. If both methods were equally predictive of true performance, we would expect that an equal number of directly observed outcomes would correspond with both assessments. The chi-squared test provides an estimate of the significance in the difference between the observed distribution and that which would be expected by chance alone.

Results

Relationship Between PEAK-IA and PEAK-PA Items

Table 2 summarizes all correlations conducted to evaluate the relationship between the corresponding 64 items from PEAK-IA and PEAK-PA. A Pearson's correlational analysis found a strong, significant correlation, $r = .89$, $p < 0.001$, between scores on the two assessments. In other words, scores on one assessment appear to predict scores on the other assessment. There was also a high correlation across assessment types for FLS, PCS, and VCS, with a moderate correlation found for VMS. Strong discriminant validity was also found for FLS, PCS, and VCS but not VMS. Figure 1 shows the relationship for scores across each measure. Linear regression provides a strong fit for the data. The equation included in Fig. 1 suggests

Fig. 1. Linear regression analysis comparing PEAK-IA (*y*-axis) and PEAK-PA (*x*-axis) scores

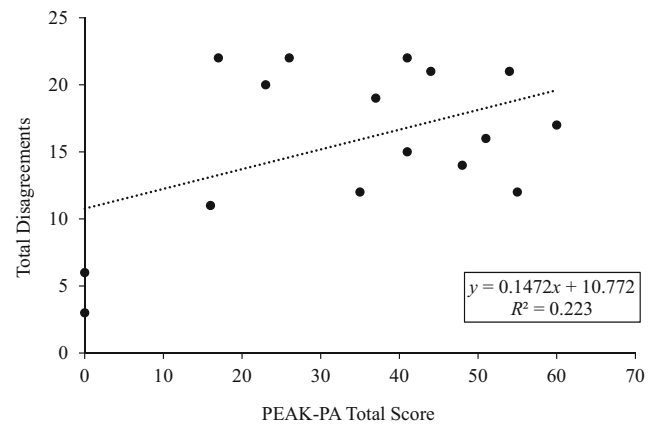
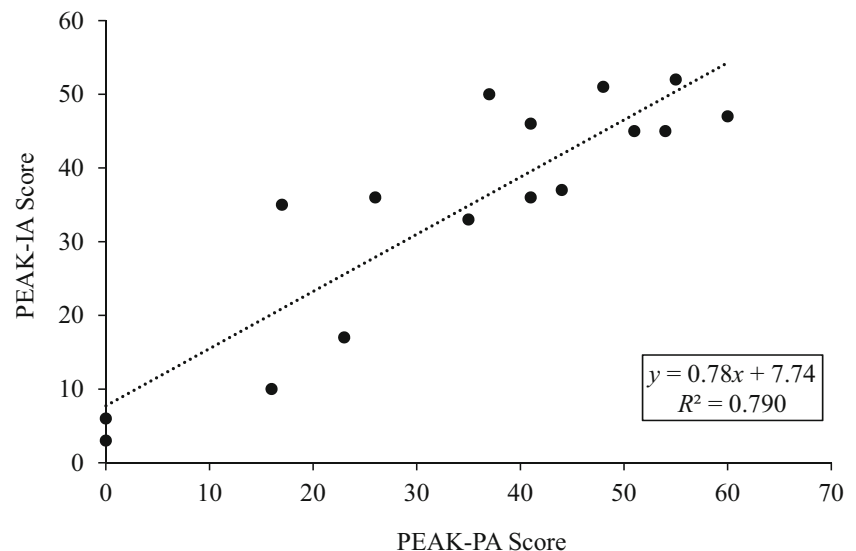


Fig. 2. Relationship between disagreements across the PEAK-PA and PEAK-IA, depicted as a function of PEAK-IA scores

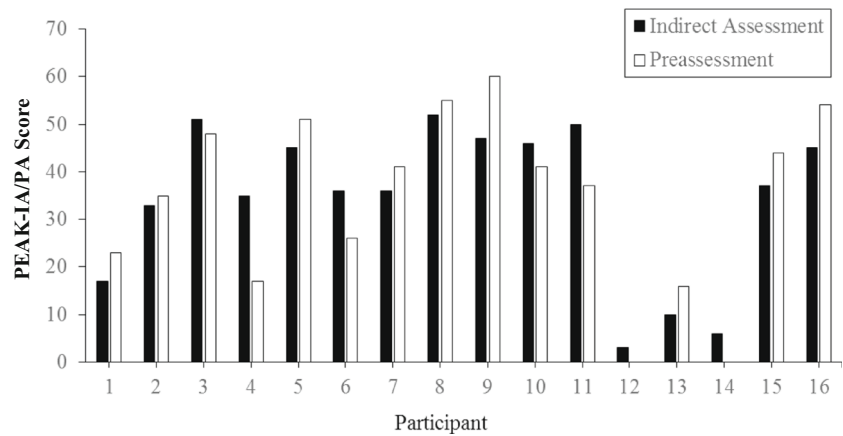
that indirect scores could theoretically estimate preassessment scores and vice versa.

Disagreements between the PEAK-IA and PEAK-PA, however, ranged from 5% to 34% of the assessment. On average, 24.7% of the items showed disagreement. Figure 2 shows the relationship between the score on PEAK-IA and disagreements between PEAK-IA and PEAK-PA. A modest correlation was found between PEAK-IA scores and the number of disagreements. In other words, the number of disagreements appeared positively related to the overall PEAK score.

Estimated Performance Between Assessment Methods

Figure 3 depicts the score across both PEAK-IA and PEAK-PA assessments for all participants. A paired-samples *t*-test

Fig. 3. Scores for each participant for the PEAK-IA (filled bars) and PEAK-PA (unfilled bars)



failed to support a significant difference in participant scores, $t(16) = 0.03, p = 0.98$. Seven participants had higher scores on PEAK-IA. Nine had higher scores on PEAK-PA. This corresponded with the t -test, as there does not seem to be a consistent difference in the directionality of the scores. The results of the effect size analysis also suggested a minimal effect for either testing method over the other, $d = .004$, further supporting that both methods appear likely to yield a similar overall estimate of true performance.

Analysis of Disagreements

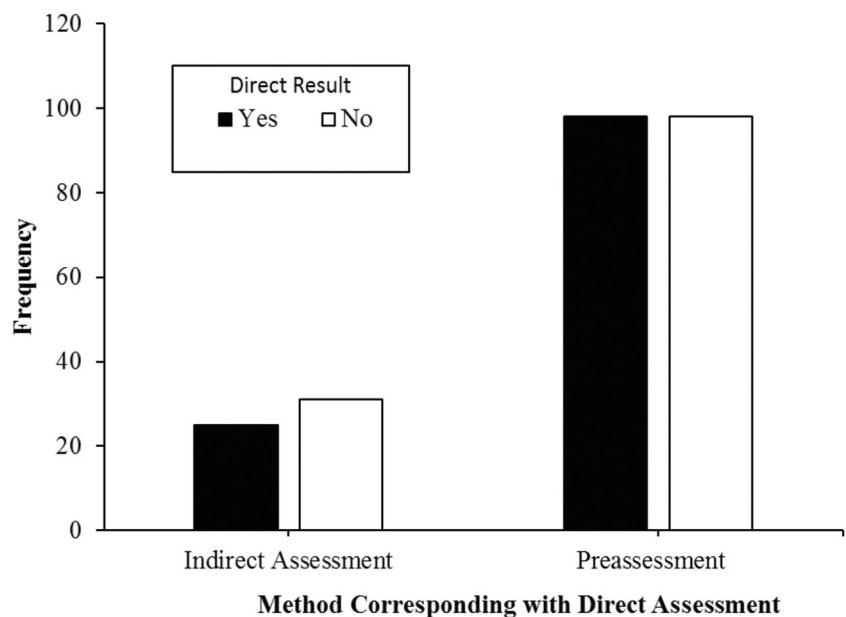
Across all 16 participants, there were 253 disagreements, or 24.7% of the items. A chi-squared analysis found that the distribution of disagreements significantly deviates from what would be expected, $\chi^2 = 78.6, p < .001$. As such, the current analysis found that PEAK-PA was significantly more predictive than the PEAK-IA. Figure 4 depicts the number of disagreements verified by direct testing across the PEAK-IA and

PEAK-PA. Visual analysis of these data also shows that, on average, disagreements between the assessment methods tended to favor PEAK-PA findings when verified by direct testing.

Discussion

In the current study, we compared the PEAK-IA and PEAK-PA to determine the potential correspondence between the two assessment methods, as well as evaluating predictive validity when compared to direct testing from the PEAK-DT module. Results suggest that for the PEAK-DT module, there is good correspondence between the corresponding 64 items included in the PEAK-IA and PEAK-PA assessments, with the PEAK-PA offering the strongest predictive validity on these items. When a behavior analyst conducts an initial PEAK assessment and finds strong agreement between PEAK-IA and PEAK-PA, these data suggest that PEAK-IA items beyond the 64

Fig. 4. Frequency of items corresponding with PEAK-DT direct testing for the PEAK-IA (left side) and PEAK-PA (right side). “Yes” disagreements are depicted by filled bars. “No” disagreements are depicted by unfilled bars



corresponding items can be viewed with greater confidence, thus reducing the need for more intensive testing. These data offer a potentially effective and efficient manner in which to conduct a comprehensive behavior identification assessment that fits with many current insurance billing parameters.

The importance of the correspondence between the PEAK-IA and PEAK-PA should be emphasized. Some FBA literature has questioned the utility of indirect assessment methods, though others have suggested that emphasis should be placed on informant methods (Hanley, 2012). The current results suggest that the PEAK-IA and PEAK-PA could be used in tandem to help streamline assessment and accurately place a client within the overall PEAK-DT module. One might question why use the PEAK-IA at all if the PEAK-PA has stronger predictive validity. With regard to PEAK-PA, caution should be taken when it is used to determine a client's overall module placement due to the fact that the PEAK-PA only contains a small sample of targets from the module. For example, VCS includes 100 unique targets, but only 16 are represented in PEAK-PA. The PEAK-IA, however, requires informants to respond to questions on every item in the module.

As such, one approach a practitioner could employ involves comparing the correspondence between obtained PEAK-IA and PEAK-PA scores. Given the strong predictive validity of the PEAK-PA, strong correspondence between parent ratings on PEAK-IA and obtained scores from PEAK-PA could lend to greater confidence in determining module placement based on PEAK-IA outcomes. The use of the multiple sources of data across PEAK-IA and PEAK-PA also would allow practitioners the ability to scrutinize performance across operant classes to determine what, if any, items require direct testing. For example, the PEAK-DT module contains seven total programs from the imitation operant class. From these seven, all except two (2B and 4E) are included on the PEAK-PA. If a client demonstrates the five imitation skills included on the PEAK-PA, coupled with a strong correspondence between all PEAK-IA and PEAK-PA items, the behavior analyst could confidently count 2B and 4E as mastered, if rated a "yes" on PEAK-IA, and then not include 2B and 4E in direct testing. This model of assessment using PEAK warrants further research, as it might offer an effective and efficient method to conduct meaningful behavioral assessments in a managed-care practice.

Though promising, the current study is limited by a number of factors. First, the sample size employed in this comparison was small when compared to similar lines of research. Given that the sample came from the same clinic, randomized sampling was not implemented, and there were no control group comparisons made. Similar to other research on PEAK, however, the results of this study suggest that the sample size was sufficient to produce reliable results, and statistical significance was demonstrated (Dixon et al., 2014a). Future research may conduct a power analysis at the onset of the study to

determine the appropriate sample size to avoid a Type II statistical error (i.e., failing to detect a true difference when one exists), especially given we did not find a significant difference between the scores produced by either method. Despite this, however, and in order to drive replication and broader acceptance of behavior analysis by the overall scientific community, future research should include larger sample sizes with more traditional group statistical analysis methods. Another limitation is that a large portion of the sample scored 32 or higher on both assessments, which limits the generality of these findings to children with autism who obtain lower scores on the assessments. Perhaps even more importantly, however, these results are limited by a lack of treatment data showing that assessment outcomes lead to meaningful changes in behavior when exposed to behavior-analytic methods based on assessment data gathered in a manner as described previously.

The current study only assessed parent ratings on the PEAK-IA. Future research should evaluate the correspondence between PEAK-IA and PEAK-PA when different raters, such as teachers or staff, complete the PEAK-IA. Additionally, similar studies are needed to determine the correspondence and predictive validity of the PEAK-IA and PEAK-PA across the other three modules (PEAK-G, PEAK-E, PEAK-T). Additionally, more investigation is needed relating to the number of disagreements between PEAK-IA and PEAK-PA. In this study, higher PEAK scores tended to involve more disagreements. Also, the weakest correspondence between methods was found on the fourth, and most complex, factor in the PEAK-DT module (i.e., VMS). It is currently unknown what contributes to this lower correspondence between higher PEAK scores and scores on the VMS factor.

In all, this study offers a promising initial inquiry into the use of multiple sources of data to determine a client's placement within the PEAK-DT module. Behavior analysts should constantly seek out the most effective and efficient ways to assess complex behavioral repertoires in hopes of producing improved treatment outcomes (Dixon et al., 2017). Practicing behavior analysts could help contribute through a willingness to collect data in a similar manner as described here and disseminating results whenever possible. This study also offers further support for the PEAK Relational Training System as a viable and practical tool for behavior analysts across a number of settings.

Author Note Portions of the study came from the master's project of the first author for partial completion of the requirements for the master's of science degree.

Compliance with Ethical Standards

Conflict of Interest No author benefitted financially from participating in this research.

Ethical Approval Procedures were in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All procedures were in accordance with the ethical standards of the Behavior Analyst Certification Board.

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

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 Behavior Education*. <https://doi.org/10.1007/s10864-01909327-8>.
- 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. <https://doi.org/10.1901/jaba.1968.1-91>.
- Behavior Analyst Certification Board. (2014a). *Professional and ethical compliance code for behavior analysts*. Retrieved from <https://bacb.com/wp-content/uploads/170706-compliance-code-english.pdf>
- Behavior Analyst Certification Board. (2014b). *Applied behavior analysis treatment of autism spectrum disorder: Practice guidelines for healthcare funders and managers*. Retrieved from https://www.bacb.com/wp-content/uploads/2017/09/ABA_Guidelines_for_ASD.pdf
- Centers for Disease Control and Prevention. (2008). Prevalence of autism spectrum disorders—Autism and Developmental Disabilities Monitoring Network, 14 sites, United States, 2008. *MMWR Surveillance Summary*, *61*, 1–19.
- Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice-Hall.
- Daniels, A. M., & Mandell, D. S. (2014). Explaining differences in age at autism spectrum disorder diagnosis: A critical review. *Autism*, *18*(5), 583–597. <https://doi.org/10.1177/1362361313480277>.
- Dixon, M. R. (2014a). *The PEAK relational training system: Direct training module*. Carbondale, IL: Shawnee Scientific Press.
- Dixon, M. R. (2014b). *The PEAK relational training system: Generalization module*. Carbondale, IL: Shawnee Scientific Press.
- Dixon, M. R. (2015). *The PEAK relational training system: Equivalence module*. Carbondale, IL: Shawnee Scientific Press.
- Dixon, M. R. (2016). *The PEAK relational training system: Transformation module*. Carbondale, IL: Shawnee Scientific Press.
- 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*(2), 493–521. <https://doi.org/10.1007/s40614-017-0119-4>.
- Dixon, M. R., Belisle, J., Stanley, C., Rowsey, K., Daar, J. H., & Szekely, 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., 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. <https://doi.org/10.1016/j.rasd.2014.07.020>.
- Dixon, M. R., Carman, J., Tyler, P. A., Whiting, S. W., Enoch, M. R., & Daar, J. H. (2014). 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., Stanley, C., 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., 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/10.1016/j.rasd.2014.05.005>.
- Doggett, R. A., Edwards, R. P., Moore, J. W., Tingstrom, D. H., & Wilczynski, S. M. (2001). An approach to functional assessment in general education classroom settings. *School Psychology Review*, *30*, 313–328.
- Dufrene, B. A., Doggett, R. A., Henington, C., & Watson, T. S. (2007). Functional assessment and intervention for disruptive classroom behaviors in preschool and Head Start classrooms. *Journal of Behavioral Education*, *16*(4), 368–388.
- Dufrene, B. A., Parker, K., Menousek, K., Zhou, Q., Harpole, L. L., & Olmi, D. J. (2012). Direct behavioral consultation in Head Start to increase teacher use of praise and effective instruction delivery. *Journal of Educational and Psychological Consultation*, *22*(3), 159–186. <https://doi.org/10.1080/10474412.2011.10474412>.
- Floyd, R. G., Phaneuf, R. L., & Wilczynski, S. M. (2005). Behavioral assessment: A review of research. *School Psychology Review*, *34*, 58–73.
- 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*(3), 990–1002. <https://doi.org/10.1016/j.rasd.2011.01.012>.
- Hanley, G. P. (2012). Functional assessment of problem behavior: Dispelling myths, overcoming implementation obstacles, and developing new lore. *Behavior Analysis in Practice*, *5*, 54–72. <https://doi.org/10.1007/FBF03391818>.
- LeGray, M. W., Dufrene, B. A., Sterling-Turner, H., Olmi, D. J., & Bellone, K. (2010). A comparison of function-based differential reinforcement interventions for children engaging in disruptive classroom behavior. *Journal of Behavioral Education*, *19*(3), 185–204. <https://doi.org/10.1007/s10864-010-9109-2>.
- National Institutes of Health. (2017). *What does it mean for something to be evidence-based?* Retrieved from <https://prevention.nih.gov/resources-for-researchers/dissemination-and-implementation-resources/evidence-based-programs-practices>
- Roane, H. S., Fisher, W. W., & Carr, J. E. (2016). Applied behavior analysis as treatment for autism spectrum disorder. *Journal of Pediatrics*, *175*, 27–32. <https://doi.org/10.1016/j.jpeds.2016.04.023>.
- 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*, 15–23. <https://doi.org/10.1007/s10882-014-9398-9>.
- Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice-Hall.

- Smith, T. (2013). What is evidence-based behavior analysis? *The Behavior Analyst*, *36*, 7–33. <https://doi.org/10.1007/bf03392290>.
- Smith, T., & Iadarola, S. (2015). Evidence base update for autism spectrum disorder. *Journal of Clinical Child & Adolescent Psychology*, *44*, 897–922. <https://doi.org/10.1080/15374416.2015.1077448>.
- Sundberg, M. L. (2008). *VB-MAPP: Verbal behavior milestones assessment and placement program*. Concord, CA: AVB Press.
- Thompson, R. H., & Iwata, B. A. (2007). A comparison of outcomes from descriptive and functional analyses of problem behavior. *Journal of Applied Behavior Analysis*, *40*, 333–338. <https://doi.org/10.1901/jaba.2007.56-06>.

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