

The Utility of the ABLLS in Comprehensive Assessment and Educational Planning: A Comparison of Instruments

Peter Doehring¹, Donna Gigliotti², Heather Calkins², & Donna Cain²
Delaware Autism Program ¹ *Statewide Programs* ² *The Brennen School*

The ABLLS has become an increasing popular choice amongst educators of children with autism. In this poster, we seek to understand this popularity by assessing the strengths and limitations of the ABLLS, relative to other commonly available instruments for assessment and/or curriculum planning. We evaluate the capacity of the ABLLS to comprehensively assess the full range of skills addressed in the IEPs of school-aged children, by systematically comparing the range of domains and distribution of items to that obtained via other assessment procedures. We summarize information supporting the potential of the ABLLS to describe distinctive profiles within and across domains that may predict future performance. To this end, we also include examples of statistical analyses of the integrity of specific subscales. Finally, we assess the utility of the ABLLS in IEP development and tracking, including teacher ratings of the potential of sample items from selected instruments for IEP purposes. These results suggest that the ABLLS' popularity may stem from the fact that it is uniquely suited to certain critical aspects of educational assessment and progress monitoring, although some limitations of the ABLLS underscore the need to fold it into a broader educational assessment.

INTRODUCTION

The challenge of assessment

Assessment serves many important purposes. It can:

- Establish an individual's functioning level relative to peers
- Set a baseline against which overall progress is measured
- Identify areas of relative strength and weakness, and
- Aid in the development of specific educational goals.

Educators working with students with autism face, however, many challenges to assessment

- These students are particularly difficult to assess because of their unique skill /behavior profile, and the tremendous variability across individuals.
- Many of available assessment instruments are not appropriate or easily adapted for use with these students.
- Educators do not always know how to evaluate the appropriateness of a given instrument for the various purposes listed above

These challenges have become more prominent in recent years with the increased accountability for students' continued progress, and the need for empirical research demonstrating the impact of interventions.

Goals of this review

The Assessment of Basic Language and Learning Skills (ABLLS) (Partington & Sundberg, 1998) – and the subsequent revision, the ABLLS-R (Partington, 2006a) – have become increasingly popular tools for assessment and IEP development for students with autism. At the largest site of the Delaware Autism Program (the Brennen School), we have begun to systematically implement the ABLLS (and now the ABLLS-R). To evaluate the success of this implementation prior to making recommendations beyond this site, we must establish guidelines to more clearly define the place of the ABLLS within the context of a broader assessment. As a point of reference, we compare it to the Vineland Adaptive Behavior Scales, or Vineland II (Sparrow et al., 2005). Though the primary stated purpose of the Vineland differs from the ABLLS, the comparison nonetheless may help to assess the potential of an instrument with respect to key questions outlined in the sections. Note that information from the ABLLS-R is taken from the test manual or protocol (Partington, 2006a; 2006b), while information from the Vineland is taken from the manual (Sparrow et al., 2005) unless otherwise indicated.

REVIEW

How easy is it to administer and interpret?

The ABLLS-R Protocol is a comprehensive and highly detailed administration and scoring guide intended for assessment and ongoing skill tracking over 3 subsequent assessments for a particular student. Annual re-assessment is recommended.

Mode of administration: Though some of the items on the ABLLS-R are based on answers provided by a teacher/caregiver, most are administered to the student, and therefore tap the student's skills directly. The relative lack of clear guidelines regarding the role and use of reinforcement during the assessment may, however, lead to variability in student performance. The Vineland consists entirely of questions regarding student performance posed to a knowledgeable adult, and accurate assessment is entirely dependent on the respondent's knowledge and reliability.

Format of questions: For each of the 544 items (Task Name) in 25 domains, the ABLLS-R guide includes a more detailed description of the item (Task Objective and Question), examples, and criteria. Some items also reference lists of exemplars provided in the back of the guide (e.g., Associations, Simple Instructions, Body Parts, etc). While these all make the test cumbersome to administer, it may help increase reliability. Each of the 432 Vineland items includes only a question, and may be less reliable, especially for defining an IEP objective and baseline for an item.

Range of scores: Except for Gross/Fine Motor items which are scored Yes or No (0 or 1), all ABLLS-R items provide opportunities to capture variations in the response or task. Of the remaining items, about one-half (259) are scored on a 3 point scale (0, 1, or 2), and others are scored on a 5 point scale (0 to 4). A partial score may result from completing fewer components (e.g., items in a list or task analysis), completing them without prompting, etc. This variation allows the emerging skills so critical to educational planning to be more accurately tracked.

Almost all of the Vineland items are scored as Usually, Sometimes or Partially, or Never. This may allow some emerging skills to be tracked, though with less precision.

Starting and stopping points: No recommended starting or stopping points in a given subscale of the ABLLS-R are provided. As a result, examiners risk stopping too early (and missing acquired skills) or stopping too late (and frustrating the student with repeated failures, as well as lengthening the administration). In contrast, the Vineland includes recommended starting points based on chronological or estimated mental age, as well as clear criteria for establishing basal and ceiling scores.

Scoring summary: ABLLS-R performance is summarized in a 3 page grid included in the Protocol. This provides a handy visual reference for considering progress within each domain across multiple assessment points but, because the ABLLS-R is not converted into standardized scores, is difficult to accurately interpret (see below). The Vineland provides a detailed scoring summary of standardized scores with confidence intervals, and the statistical significance of differences between domains & subdomains.

Supporting materials: Neither the Vineland nor the ABLLS-R comes with a kit of materials, even though such a kit is essential to reliable administration, at least of the latter.

Administration time: Even for a more seasoned user, the ABLLS-R takes a long time to administer (3 to 10 hours, depending on the age and functioning level of the child). The Vineland can usually be administered in less than an hour.

What skills and behaviors does it address?

We summarize the distribution of items in Figures 1 and 2. Consistent with its stated emphasis on verbal skills, more items focus on language in the ABLLS-R versus the Vineland. The ABLLS-R also discriminates among items related to Expressive Language, largely based on the functions of language identified by Skinner.

- The ABLLS-R also addresses a range of skills specifically intended to assess learner readiness in various classroom settings (e.g., Cooperation & Reinforcers, Group Instruction, Classroom Routines, and Generalized Responding).
- The ABLLS-R does not address: coping skills; peer relations; maladaptive behaviors, or; certain self-help skills typically viewed as critical to educational programs for more challenged individuals (e.g., community and domestic skills).
- Except for certain basic pre-academic and academic skills (e.g., reading, writing, spelling, math, and visual performance) neither assessment provides

Fig. 1: Communication Items by Domain (*Subdomain*)

	# items (% of total)		# items (% of total)
Vineland	432	ABLLS-R	544
Overall	99 (23%)	Overall	245 (45%)
<i>Expressive</i>	54		(164)
		Spont.	
		Vocalizations	19
		Requests	29
		Labeling	47
		Intraverbals	49
		Syntax and Grammar	20
		Receptive Language	57
<i>Receptive</i>	20		(34)
<i>Written</i>	25	Reading ¹	17
		Writing ¹	10
		Spelling ¹	7

great detail regarding academic skills, although the authors of both instruments clearly acknowledge that this was never their intent.

A more detailed examination of certain comparable subscales (e.g., Receptive language, Social Interaction, and Personal). to informally assess the range of skills addressed by each instrument, suggest that there may be important differences – e.g., some items across instruments are almost identical while others are very different. Unfortunately, a more detailed comparison is beyond the scope of the present paper.

With what students can it be used?

There are no specific guidelines specifying the population with which the ABLLS-R can be used. In contrast, the Vineland was explicitly developed to be applicable into adulthood. We also scanned some of the ABLLS-R scales which correspond closely with the Vineland (e.g., Receptive language, Social Interaction, and Personal). In each case, the top item on the corresponding ABLLS-R scale fell at least several items below the top item on the Vineland. This suggests that the range of the Vineland clearly extend beyond the

Fig. 2: Other Domains (*Subdomains*)

		# items (% of total)		# items (% of total)
	Vineland	432	ABLLS-R	544
Daily Living	Overall	109 (25%)	Overall	42 (8%)
	<i>Community</i>	44		(42)
	<i>Domestic</i>	24		
	<i>Personal</i>	41		
				Dressing ¹ 15 Eating ¹ 10 Grooming ¹ 7 Toileting ¹ 10
Maladapt. Behavior	Overall	49 (11%)	Overall	
	<i>Critical items</i>	14		
	<i>Externalizing</i>	9		
	<i>Internalizing</i>	11		
	<i>Other</i>	15		
Motor	Overall	76 (18%)	Overall	58 (11%)
	<i>Fine</i>	36	Fine ¹	28
	<i>Gross</i>	40	Gross ¹	30
Social	Overall	99 (23%)	Overall	96 (18%)
	<i>Coping</i>	30		
	<i>Interpersonal</i>	38	Social Interaction	34
	<i>Play & Leisure</i>	31	Motor Imitation	27
Other			Vocal Imitation	20
			Play and Leisure	15
			Overall	103 (19%)
			Math ¹	29
			Visual Performance	27
			Cooperation & Reinforcers	19
			Group Instruction	12
			Classroom Routines	10
			Generalized Responding	6

¹These are not considered part of the Basic Learner Skills

ABLLS-R, and that the usefulness of the ABLLS-R for older and/or higher functioning students is questionable.

Does it identify profiles of strength & weakness?

The first 15 scales of the ABLLS-R are part of the Early Learner Assessment. These are viewed as more critical skills that help prepare the student for learning other, more traditional academic (e.g., reading, math, writing, spelling) or self-care skills. The author recommends that one-half to two thirds of the objectives be selected from this section, and presents an example of one Early and one Advanced Learner Profile, together with accompanying IEP objectives.

- The inclusion and prioritization of learner readiness skills in the ABLLS-R (e.g., Cooperation & Reinforcers, Group Instruction, Classroom Routines, and Generalized Responding) clearly increases its utility for early learners
- While certain skills (e.g., communication) are clearly more important to teach before other skills (e.g., academic), we believe that such profiles will be much more useful when broken down further. For example, we believe that a subset of learner readiness and communication skills can be identified within the Early Learner profile that are absolutely critical

Much research has sought to establish different patterns of skill profiles in the Vineland – e.g., to differentiate between populations individuals with and without autism, or to understand the relation between adaptive and intellectual functioning within populations of individuals with autism. We do not know of studies in which specific Vineland profiles were used to predict individual outcomes within a population of students with autism.

Has research been conducted to validate its structure?

Despite an exhaustive search, we have found no published research that addresses questions of validity or reliability. To convey the importance of such research, we describe some of the many tests conducted to test the psychometric qualities of the Vineland.

Test-Retest reliability is used to confirm that similar performance is obtained via a subsequent administration, providing evidence that the results are reasonably stable. We therefore lack empirical evidence that results obtained via the ABLLS-R are stable.

Inter-rater reliability is used to confirm that two different raters draw the same conclusions. This provides evidence that the scoring criteria are clear and that examiner scores are not too subjective. We therefore lack empirical evidence that results obtained via the ABLLS-R are not overly affected by the judgment of the examiner.

Internal consistency is used to verify that all items within a given subscale are reasonably correlated with the total score. This provides evidence that individual items are related to the overall construct represented by the domain. We therefore lack empirical evidence that all of the items in a given domain (e.g., receptive language) contribute reasonably strongly to a single overall score.

Factor analysis is used to confirm that the pattern of relationships between items corresponds to the overall structure of the domains. For example, it would confirm that; (a) items related to requesting are strongly correlated with each other, (b) items related to a labeling are correlated with each other, and (c) the correlations within (a) and within (b) are clearly stronger than between (a) and (b). We therefore lack empirical evidence that items within a given domain of the ABLLS-R (e.g., requesting) are more strongly correlated with each other than they are with another domain (e.g., labeling).

External validity may be evident when performance on the measure of interest is strongly related to another measure of the same type of behavior.

Scalogram analysis is used to verify that items in a sequence actually emerge in the predicted order (e.g., that in a hypothesized sequence A-B-C, Skill C is never evident unless Skill A and Skill B are also evident). In fact, we are not aware of any scalogram analyses conducted using the Vineland, and such analyses are rarely used even in research with a clear developmental focus. We therefore lack empirical evidence that the individual skills emerge reliably in the same order as listed in the ABLLS-R.

Can it be useful in setting benchmarks for tracking overall progress or progress within a domain?

One of the most commonly used purposes of testing is to generate an overall score that can be used to assess level of functioning or overall progress. Though not the stated intent of the author, such a goal has been expressed by others. In addition, the visual presentation of the scores within individual subscales may lead educators to draw comparable conclusions - e.g., that a student who completes 10 items in a given scale the first time around and now can complete 20 has made twice as much progress as a student who now only completes 15 items. It becomes even more problematic when trying to draw relative comparisons across domains.

Many questions center on the nature of the underlying scale:

- The total score in any individual domain in the ABLLS-R is likely to represent an ordinal scale (e.g., someone who passes 6 requesting items demonstrates superior performance to someone passing 3).
- We have no evidence that the underlying scale is interval (e.g., the difference between a 4 and a 5 is the same as the difference between a 3 and a 4) let alone a ratio scale.

The nature of the scale can significantly curtail the range and the power of statistical analysis that can subsequently be conducted on any summed scores – e.g., analyses comparing overall score or domain scores over time to track progress, to compare students with each other, to establish profiles, or to examine relationships with other measures. Until additional analyses establish the reliability and validity of summed ABLLS-R scores, or clarify the nature of the underlying scale, we believe that interpretations based on summed or overall scores may be fatally flawed.

In contrast, the raw scores obtained on the Vineland are converted into standardized scores, addressing many of the concerns noted above. As a result, the Vineland has been used to assess: educational outcomes for broad programs and specific interventions; impact of medications, and; changes over time independent of a specific program of intervention.

Is it useful in generating possible IEP objectives?

On order to assess how readily performance on individual items on the ABLLS-R and Vineland could be translated into IEP objectives, we asked 8 teachers and specialists to rate sample items on two dimensions:

Figure : Ratings

	Domain	ABLLS-R	Vineland	Overall
Rating	Receptive Language			
Easy?	1=Very Easy/Important	2.38	2.78	2.58
Important?	4=Not Easy/Important	1.97	2.16	2.06
Overall		2.18	2.47	
	Self Care			
Easy?	1=Very Easy/Important	1.76	2.02	1.90
Important?	4=Not Easy/Important	1.35	1.62	1.50
Overall		1.56	1.82	

(a) how important and relevant is this item as an IEP objective, and (b) how easy it would be to develop an objective based on this item. Each rating was on a scale from 1 to 4 (very important/easy to not important/easy). Raters were all experienced in teaching children with autism. We selected items from two scales that covered similar domains in the two instruments: receptive language (C in the ABLLS-R), and Self Care (Personal Skills in the Vineland, and U, V, W, and X in the ABLLS-R). We randomly selected twenty items from the ABLLS-R and Vineland with respect to each domain, and then randomly mixed them together to create a 40 item questionnaire for Receptive Language and 40 items for Self-Care.

Results indicate that, for both domains assessed, the ABLLS-R items were considered to tap relatively more important skills, and were easier to translate into IEP objectives.

:

Most difficult/least important Language items	Scale	Rating
Follows instructions or directions heard 5 minutes before	Vineland	3.7 (E)
Listens to a story for at least 30 minutes	Vineland	3.7 (E)
Listens to an informational talk for at least 30 minutes	Vineland	3.7 (E) 3.3 (I)
Follows instructions in “if-then” form (for example, “If you want to play outside, then put your things)	Vineland	3.6 (E)
Acquires new selection skills without intensive training	ABLLS-R	3.3 (E)
Listens to an informational talk for at least 15 minutes	Vineland	3.3 (E)
Understands sayings that are not meant to be taken word for word (for example Button your lip, Hit the road etc)	Vineland	3.3 (E)
Selects items by following another’s gaze	ABLLS-R	3.3 (E)
Most difficult/least important Self-Care items		
Takes medicine as directed (that is, follows directions on label).	Vineland	3.8 (I) 3.8 (E)
Keeps track of medications (nonprescription and prescription) and refills them as needed)	Vineland	3.5 (E) 3.3 (I)
Seeks medical help in an emergency (for example, recognizes symptoms of serious illness or injury, such as shortness of breath, chest pain, uncontrolled bleeding, etc.)	Vineland	3.3 (I) 3.2 (E)
Turns faucets on and adjusts temperature by adding hot or cold water	Vineland	3.3 (E)
Student will be able to independently blow his nose as needed.	ABLLS-R	3.2 (E)

CONCLUSIONS AND RECOMMENDATIONS

Strengths of the ABLLS-R

- Individual items are easy to translate into meaningful IEP objectives
- The fact that most items are directly administered to the student, and the detail provided in test protocol itself (e.g., example, criteria, etc), are likely to enhance its reliability and precision
- The range of clearly specified scores for individual items helps in the development of specific IEP objectives and establish baselines

- It addresses a wide range of skills, including Learning readiness skills, and the distinction various language functions

Weaknesses

Major Weaknesses

- There are no data to suggest that scores can be added together in a meaningful way, especially for the purpose of statistical analyses seeking to track progress overall or within a domain, or examine relationships between domains. The visual scoring grid, though handy, may mislead scorers seeking to interpret progress
- There are no published data regarding reliability and/or validity – e.g., to confirm that score are stable across time and testers, that individual items are clearly and distinctly related to the domains in which they have been placed, that they emerge in the order suggested, and that the results obtained correspond in any way with those obtained using other comparable measures

Other weaknesses.

- The test is very long, lacks baseline/ceiling guidelines, and is likely to be cumbersome to administer except for a more experienced user
- The lack of a test kit, or clear guidelines regarding the use of reinforcement may lead to important variability in test administration
- It does not address domestic, skills, community skills, coping skills, peer relations, or problem behaviors
- It is unclear with which students the ABLLS-R may be used
- The distinction between early and advanced learner profiles is broadly useful but limited in projecting progress individually.

In general, many of the strengths and weakness identified for the ABLLS-R and the Vineland are not surprising given their original stated purpose. For example, the ABLLS-R was not originally developed to yield an overall summed score, but to distinguish between various verbal behaviors. Some of the limitations identified here arise simply from attempts to extend its usefulness beyond the purposes stated by the author. In general, it is likely to be difficult or impossible for a single instrument to answer all or the questions posed here.

Recommendations

We conclude that the ABLLS-R is an important tool for educational planning purposes because it addresses a broad range of critical skills in a manner that translates readily to the development of relevant IEP objectives.

- We must recognize that administration of the ABLLS-R is time-consuming, and must ensure that supports (e.g., adequate training, a test kit, etc) are in place before we can begin
- While it can be employed for younger and/or more challenged students, its usefulness for other students remains to be determined. For example, we should not plan on using the ABLLS-R with students who can be meaningfully assessed using standardized statewide tests used with typical students.
- We cannot interpret and analyze summed scores for purposes of evaluating progress until more information is available regarding the validity and reliability of the instrument, as well as the underlying scale. In the interim, we must supplement the ABLLS-R with other instruments that can track progress.

- It is clear that the ABLLS-R addresses many - but not all – skills. The extent to which it may be the primary basis for a school-wide curriculum will depend on more detailed analysis of the comprehensiveness of items within the specified domains.

REFERENCES

- Partington, J. W. (2006a). *The Assessment of Basic Language and Learning Skills - Revised (The ABLLS-R Guide): Scoring Instructions and IEP Development Guide*. (Version 3.0 ed.) Pleasant Hill, CA: Behavior Analysts Inc.(
- Partington, J. W. (2006b). *The Assessment of Basic Language and Learning Skills - Revised (The ABLLS-R Guide): The ABLLS-R Protocol*. (Version 3.0 ed.) Pleasant Hill, CA: Behavior Analysts Inc.(
- Partington, J. W. & Sundberg, M. L. (1998). *The Assessment of Basic Language and Learning Skills*. Pleasant Hill, CA: Behavior Analysts, Inc.(
- Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *The Vineland Adaptive Behavior Scales: Second Edition*. Minneapolis, MI: American Guidance Service.(