Autism Behavior Inventory: A Novel Tool for Assessing Core and Associated Symptoms of Autism Spectrum Disorder

Abi Bangerter, MA,[1] Seth Ness, MD, PhD,[1] Michael G. Aman, PhD,[2] Anna J. Esbensen, PhD,[3] Matthew S. Goodwin, PhD,[4] Geraldine Dawson, PhD,[5] Robert Hendren, DO,[6] Bennett Leventhal, MD,[7] Anzalee Khan, PhD,[8] Mark Opler, PhD,[9] Adrianne Harris, PhD,[5] and Gahan Pandina, PhD[1]

Abstract

Objective: Autism Behavior Inventory (ABI) is a new measure for assessing changes in core and associated symptoms of autism spectrum disorder (ASD) in participants (ages: 3 years-adulthood) diagnosed with ASD. It is a web-based tool with five domains (two ASD core domains: social communication, restrictive and repetitive behaviors; three associated domains: mental health, self-regulation, and challenging behavior). This study describes design, development, and initial psychometric properties of the ABI.

Methods: ABI items were generated following review of existing measures and inputs from expert clinicians. Initial ABI scale contained 161 items that were reduced to fit a factor analytic model, retaining items of adequate reliability. Two versions of the scale, ABI-full (ABI-F; 93 items) and ABI-short version (ABI-S; 36 items), were developed and evaluated for psychometric properties, including validity comparisons with commonly used measures. Both scales were administered to parents and healthcare professionals (HCPs) involved with study participants.

Results: Test–retest reliability (intraclass correlation coefficient [ICC] = 0.79) for parent ratings on ABI was robust and compared favorably to existing scales. Test–retest correlations for HCP ratings were generally lower versus parent ratings. ABI core domains and comparison measures strongly correlated ($r \geq 0.70$), demonstrating good concurrent validity.

Conclusions: Overall, ABI demonstrates promise as a tool for measuring change in core symptoms of autism in ASD clinical studies, with further validation required.

Keywords: autism spectrum disorder, rating scale, software, assessment, outcome, measures

Introduction

INTERVENTIONS FOR ADDRESSING core symptoms of autism spectrum disorder (ASD) are limited, in part, due to lack of valid and reliable objective endpoints for measuring clinically and statistically significant changes in core symptoms (Ghosh et al. 2013). Two core symptoms of ASD diagnosis include social communication deficits and restricted and repetitive patterns of behavior. Apart from these, other symptoms (obsessive/compulsive attributes, aggression, self-injury, unmanageable mood swings, hyperactivity and concentration issues, anxiety, and sleep disorders) (Anagnostou et al. 2015) coexist and impact overall functioning, quality of life, and treatment outcomes for ASD (Matson et al. 2013).

[1]Janssen Research & Development, LLC, Titusville, New Jersey.
[2]The Nisonger Center University Center for Excellence in Developmental Disabilities (UCEDD), Ohio State University, Columbus, Ohio.
[3]Division of Developmental and Behavioral Pediatrics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio.
[4]312E Robinson Hall, Department of Health Sciences, Bouvé College of Health Sciences, Northeastern University, Boston, Massachusetts.
[5]Duke Center for Autism and Brain Development, Duke University, Durham, North Carolina.
[6]Department of Psychiatry, University of California, San Francisco, California.
[7]UCSF Benioff Children's Hospital, San Francisco, California.
[8]Nathan Kline Institute, Orangeburg, New York.
[9]ProPhase, LLC, NYU School of Medicine, Columbia University Medical Center, New York, New York.

To date, different tools have been used to measure core and co-occurring symptoms, outcomes, and progress in ASD. The varying relevance of existing rating scales and limited evidence of their measurement properties hinder interpretation of treatment outcomes (McConachie et al. 2015). A number of scales used, such as the Autism Diagnostic Interview-Revised, are designed primarily for diagnosing ASD and are not ideally suited to assess treatment outcomes (Lord et al. 1994; Le Couteur et al. 2003). Other scales, such as the Aberrant Behavior Checklist (ABC) (Aman et al. 2004) and the Child Behavior Checklist (Achenbach and Rescorla 2001), measure overlapping constructs and have been used to measure change in ASD interventions, but were not specifically developed for this purpose.

Compounding evidence and recent advances describing outcome measures in the field of ASD emphasize the need for more targeted scale development in this area (Lecavalier et al. 2014; Aman et al. 2015; Anagnostou et al. 2015; McConachie et al. 2015; Scahill et al. 2015). Autism Speaks assembled a panel of experts to conduct a systematic review of available measures in Social Communication (Anagnostou et al. 2015), Repetitive Behaviors (Scahill et al. 2015), and Anxiety (Lecavalier et al. 2014). Besides a few measures whose relevance was conditional, the panel found no measures to be fully appropriate for measuring ASD outcomes. A recent editorial further outlined difficulties with existing rating scales (Aman et al. 2015). First, the items are often too complex and may use overly clinical symptom terminology, making it difficult for parents to understand. Second, several items refer to behaviors that can only be observed in a particular context or only during a particular developmental stage, thus limiting age appropriateness. Involving the respondent in the development process is an important aspect (U.S. Department of Health and Human Services 2009) and can ameliorate some of the difficulties described.

Another limitation of existing scales relates to anchor choice. If fewer anchor points are selected, this creates a coarser scale with limited options for respondents to rate, providing only a few points to capture severity and responsiveness to change. The ability of rating scales to capture change, particularly change early in intervention, may be critical for evaluating the success of intervention effectiveness.

Establishing sensitivity to change in core ASD symptoms has been particularly difficult, due to the limited number of effective interventions in the field (Anagnostou et al. 2015). The Vineland Adaptive Behavior Scale (Sparrow and Balla 2005)-Revised—suitable for birth to 90 years—demonstrated evidence of sensitivity to change although it was inconsistent and time-consuming (Anagnostou et al. 2015). The Social Responsiveness Scale, second edition (SRS-2), which measures both social impairment and repetitive behaviors, has been used as an outcome measure in ASD interventions (Singh et al. 2014). However, the specified recall period of 6 months with the SRS-2 limits its use in measuring change over shorter periods of time. The Autism Impact Measure (Kanne et al. 2014) has been developed to measure short-term change in ASD core symptoms but, to the best of our knowledge, is yet to be validated in an intervention study. Other instruments frequently used as clinical trials endpoints (McCracken et al. 2002; Research Units on Pediatric Psychopharmacology Autism Network 2005; Akhondzadeh et al. 2008; King et al. 2009; Marcus et al. 2009, 2011; Scahill et al. 2012; Singh et al. 2014) include the Repetitive Behavior Scale–Revised (RBS-R) (Lam and Aman 2007), Child and Adolescent Symptom Inventory-4 Anxiety Scale (CASI-ANX) (Sukhodolsky et al. 2008), and ABC (Aman et al. 2004). However, these scales assess either a single behavioral domain or do not cover both social communication and repetitive behaviors comprehensively. Therefore, they may be suboptimal for measuring the results of interventions on core ASD symptoms.

Described herein is the initial development and pilot testing of a novel measure, the Autism Behavior Inventory (ABI), which was evaluated in its full form (ABI-F) and short form (ABI-S). These measures are web-based scales completed by parents for reporting the behaviors of individuals with ASD. Our primary objective was to develop a scale for measuring change in core and associated symptoms in both children and adults with ASD. The intention was that the scale will be freely available for use in research and clinical practice. This pilot investigation of the ABI-F and ABI-S includes comparisons with established and currently available scales for measuring autism symptoms, and the evaluation of associated scale characteristics, validity, and other psychometric properties.
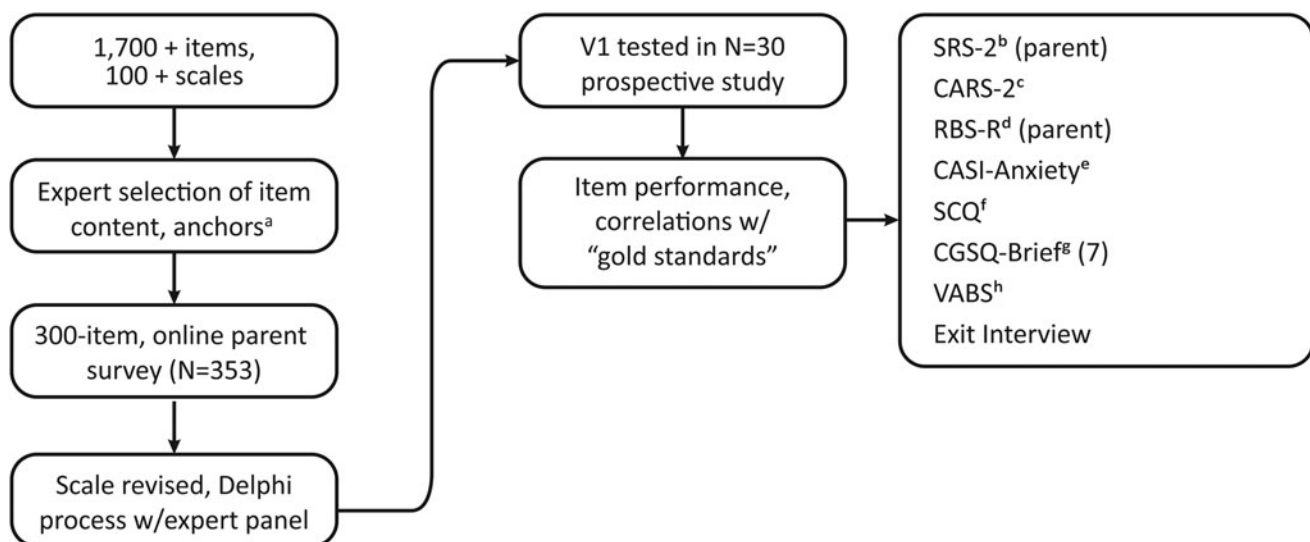
Anchor refers to the type of domain employed for rating and the Likert scale on which items were rated; [b]Social Responsiveness Scale, Version 2; [c]Childhood Rating Scale- Version 2; [d]Repetitive Behavior Scale – Revised; [e]Child Adolescent Symptom Inventory Anxiety; [f]Social Communication Questionnaire; [g]Care Giver Strain Questionnaire – Brief; [h]Vineland Adaptive Behavior Scales

FIG. 1. Scale development process.

Methods

Development of the ABI consisted of phase 0—item selection and scale design—and phases 1 and 2—item testing analyzed for psychometric performance and further item and scale optimization (Fig. 1).

Phase 0: scale design

Drafting of items to be measured was based on literature reviews and expert opinion. Clinicians, public health experts, and expert practicing clinicians provided input to conceptualize ASD assessment tool development. This group generated items, refined item wording, evaluated completeness of item coverage across ASD domains, and performed initial assessment of clarity and readability.

After items were selected, they were assigned to item groups by expert clinicians: these groups of items formed the domains and subdomains of the ABI. A total of five domains were selected: two core domains of Social Communication and Restrictive and Repetitive Behaviors, and co-occurring domains of Mental Health, Self-Regulation, and Challenging Behavior. Each domain was divided into subdomains and items of each subdomain were linked to two of four possible anchors (quality, context, frequency, and intensity). For measures of subdomains, clinical experts reviewed how well individual items were associated with each other, how items were conceptually associated with each domain, how domains were conceptually associated with each other, and the overall construct of the scale.

Likert scale anchors were customized to each behavior item as appropriate. Valence of items was adjusted for scoring so that in all cases 0 = absence of symptoms, 6 = maximum symptoms, for example, "gets upset over small changes in routine" Frequency: never, would score 0; whereas "uses facial expressions that are appropriate to the situation" Frequency: never, would score 6.

Using a web-based format, items were automatically personalized using the child's name and gender-specific pronoun in each of the behavior descriptions. It also enabled filtering of items as appropriate for age, and verbal ability (submitted for publication). Anchors that were not appropriate were not displayed (e.g., if a parent selected "none" for frequency of a behavior, they were not prompted to rate the intensity of the behavior).

Phase 1: item reduction

After providing consent, eligible parents of children aged 3 years or older, who had a diagnosis of ASD, completed a version of the ABI containing 161 items. Factor and usability analyses were performed on these data to derive optimal behavioral items that met a variety of psychometric criteria.

Phase 2: reliability and validity

To determine reliability and validity, the ABI was subjected to preliminary psychometric analyses in a pilot study. Twenty-five parents and 20 healthcare professionals (HCPs) completed the ABI as part of a noninterventional validation study of the JAKE™ system (submitted for publication). The scores on the ABI were also compared with other scales measuring similar constructs specific to ASD, or that have been used in ASD populations previously.

The ABI-F was administered at baseline and endpoint visits in an 8-week monitoring period. It was completed independently by the same parent and site HCP (three sites) at each visit.

TABLE 1. DEMOGRAPHICS AND BASELINE CHARACTERISTICS

	Total (n = 353)
Phase 1	
Participant age, n (%)	
3–10 years	198 (56)
11–21 years	155 (44)
Gender, n (%)	
Male	275 (78)
Female	78 (22)
Participant's level of language	
No language	18 (5)
Uses signs or picture exchange for single words	17 (5)
Uses single words/2–3 utterances	49 (14)
Puts signs or picture exchange together to make simple sentences	9 (3)
Uses simple sentences	100 (28)
Speaks in full sentences	159 (45)
Parent age, n (%)	
18–24 years	21
25–34 years	125 (35)
35–44 years	118 (33)
55–64 years	66 (19)
65–74 years	23 (65)
Parent education	
Less than high school	5 (1)
High school graduate	52 (15)
Some college credit, no degree	118 (33)
Bachelor's degree	123 (35)
Master's degree	37 (10)
Professional degree	14 (4)
Doctorate degree	4 (1)
Phase 2	
Age (years)	
Mean (SD)	10 (5.3)
Age, n (%)	
3–5 years	3 (13.0)
6–9 years	8 (34.8)
10–12 years	7 (30.4)
13–17 years	3 (13.0)
>18 years	2 (8.7)
Gender, n (%)	
Male	21 (91.30)
Female	2 (8.70)
Race, n (%)	
White	22 (95.7)
Multiple	1 (4.3)
CARS total, mean (SD)	47.7 (8.8)

CARS, Child Autism Rating Scale; SD, standard deviation.

Results

The demographics and baseline characteristics are summarized in Table 1.

Phase 1

Confirmatory factor analysis. The items were subjected to confirmatory factor analysis (CFA) with principal axis factoring. The rotation method was varimax with Kaiser Normalization. The Kaiser–Meyer–Olkin measure of the sampling adequacy ratio of total items was 0.932, indicating appropriateness for factor structure.

The CFA with five factors accounted for 63.74% of the variance. All items loaded on their respective domains, with several items cross loading. Cross loading items were reviewed further as a part of the item-reduction process (Supplementary Table S1; Supplementary Data are available online at www.liebertpub.com/cap).

Item reduction.

Item reduction processes were carried out using the item characteristics performance criteria described in Table 2. Items that did not meet criteria set out for item characteristics were flagged for deletion or review by clinician experts. Items with a low item-total correlation (i.e., low correlation with respective subscale) and low item discrimination index (IDI) (<0.30) were considered for elimination. Differential item functioning (DIF) (Rogers 2005; Karami 2012) was examined for age and verbal ability (verbal vs. nonverbal). If moderate-to-severe DIF was noted, the item was flagged for deletion or review.

After factor analysis and item reduction, two versions of the ABI scale were developed: ABI-F 93 items and ABI-S 36 items.

Consultation and feedback from parents on scale development.

Parents were asked further questions regarding the items and response choices, and given the opportunity to provide qualitative feedback. In all, 86% of respondents rated items on the survey as extremely representative of their child's behaviors. Eight percent of respondents suggested additional items for inclusion. Specifically, 25% suggested food-sensitivity issues, which resulted in the generation of one new food-sensitivity item. Seven percent of suggestions related to sleep, leading to one new sleep item. Other suggestions related to developmental outcomes—toileting and physical development or child-specific items, which may be important and related to level of functioning but were considered outside the scope of the ABI.

Parents were asked about the suitability of response options. A total of 93% respondents selected "choices always made sense" option. Overall, 94% of parents responded as having "no difficulty at all" to a question about difficulty understanding or responding to items. Items for which parents indicated "they did not understand what was being asked" were removed or reworded with specific examples added.

Parents also responded to an open-ended question regarding improvements to the scale. Fifty-one percent of respondents stated that they had no suggestions for changes, 19% gave a positive response such as "This was one of the easiest I've seen." or "I was surprised at the inclusiveness and wide range of detailed issues that go on with a child dealing with such a diagnosis, well put together." There were very few responses that were not positive. Two parents found the survey lengthy and others (22%) added suggestions for improvement to scale content, presentation, or response structure. Responses relating to suitability of the scale for verbal and nonverbal participants and the inclusion of specific examples for items and scale length were taken into account when finalizing the pilot scale. In addition, based on parent feedback, we added a completion "progress bar" and examples for some of the questions that they reported as more difficult to understand.

Phase 2

Test–retest reliability.

Test–retest reliability ($n = 43$) was assessed using the Wilcoxon signed-rank test to compare ABI score changes at the beginning of study (first) and the end of the study (last) for parents and HCPs. This accounted for 23 parents and 20 HCPs; data from 2 parents were not available at both time points.

Most ABI correlations were statistically significant (Table 3). Some subdomains failed to show significant correlations either for parents (elopement only: $r = 0.23$, $p = 0.277$) or site HCPs (e.g., hypersensitivity $r = 0.103$, $p = 0.78$, tantrums $r = 0.351$, $p = 0.199$). Test–retest reliability was higher for parents than for site HCPs, but all estimates fell within the 95% confidence interval (CI). Test–retest correlations were often >0.90 for parents and >0.80 for site HCPs (Table 3). The median correlation was 0.86 for parents and 0.71 for HCPs, and for 48% of site HCPs, the lower limit of the 95% CIs was >0.70.

Internal consistency reliability.

Cronbach's alpha was used to measure internal consistency for the subdomain scores within

TABLE 2. REASONS FOR CHANGING ITEMS DURING SURVEY TOOL DEVELOPMENT

Item characteristic	Reasons for amendments or deletion
Clarity and implications	Reported as not relevant by Subject Matter Experts or caregivers completing the survey
	Generated an unacceptably large amount (>65%) of missing data points
	Generates many questions or requests for clarification
	Caregivers interpret items and responses in a way that is inconsistent with the conceptual framework
Range of response (ceiling or floor effects)	A high percent of respondents at the floor or ceiling
	None of the response choices applies to them
	Distribution of item responses is highly skewed (skewness, kurtosis)
Variability	All or most (>80%) of caregivers give the same answer (i.e., no variance)
	Differences among subjects are not detected when important differences are known
Interitem correlation	Item is highly correlated with other items in the same subdomain (this can represent redundant concepts being measured)
Item discrimination	Item is highly correlated with measures of concepts other than the one it is intended to measure
	Item does not show variability in relation to some known population characteristics (i.e., severity level, classification of condition, or other known characteristic)
Differential item functioning	Item shows variable for different groups (e.g., females compared to males, verbal vs. nonverbal, younger subjects vs. older subjects)
Item information function	If the amount of information an item conveys is small, it means that the ability (behavior) cannot be estimated with precision and the estimates will be widely scattered about the true ability (behavior).

TABLE 3. SUMMARY OF TEST–RETEST CORRELATIONS FOR ALL
SUBSCALES FOR FIRST-LAST AND BASELINE-ENDPOINT VISITS

	n	*Min*	*Q1*	*Median*	*Q3*	*Max*
First vs. last, ABI-F scores (parents and HCPs)						
Correlations	43	0.10	0.67	0.79	0.88	0.96
Differences	44	−0.62	−0.32	−0.15	0.02	1.22
Correlation *p*-values	43	<0.001	<0.001	<0.001	0.001	0.778
Difference *p*-values	43	<0.001	0.013	0.219	0.488	0.965
Baseline vs. endpoint, standard scales						
Correlations	21	0.618	0.703	0.767	0.83	0.92
Differences	21	−2.857	−0.714	−0.286	−0.071	1.571
Correlation *p*-values	21	<0.0001	<0.0001	<0.0001	<0.0001	0.0005
Difference *p*-values	21	0.0012	0.3965	0.4987	0.7756	0.9404

ABI, Autism Behavior Inventory; HCPs, healthcare professionals; n, number of domains/scales.

their respective domain scores for the ABI and the ABI-S. Acceptable alpha values between 0.70 and 0.90 across raters and visits were recorded for restrictive and repetitive behaviors and self-regulation. However, due to limited sample size, most values for challenging behavior, mental health, and social communication were below that range but close to the required threshold (Table 4). Alpha for the social communication scale on the ABI-S was poor (0.15), indicating that internal consistency for the subset of social communication items in the short scale was unacceptable.

Validity. To evaluate concurrent, convergent, and divergent validity of the scale, the ABI domain and subdomain scores were compared with existing criterion assessments: SRS (parent), Social

TABLE 4. CRONBACH'S ALPHA FOR AUTISM BEHAVIOR
INVENTORY SUBDOMAINS WITHIN DOMAINS

	Cronbach's alpha	
	Full	*Short*
Domain: social communication	0.58	0.15
Subdomains:		
Joint attention, nonverbal communication, reciprocity, verbal communication		
Domain: restrictive repetitive behaviors	0.80	0.60
Subdomains:		
Hypersensitivity, resistance to change, restricted interests, stereotypical behaviors		
Domain: mental health	0.63	NC
Subdomains:		
Irritability, anxiety		
Domain: self-regulation	0.70	NC
Subdomains:		
Hyperactivity, impulsivity		
Domain: challenging behavior	0.68	0.46
Subdomains:		
Aggression, antisocial behavior, elopement, temper tantrums		

NC, not calculated.

TABLE 5. CORRELATION OF TOTAL DOMAIN SCORE
ON THE AUTISM BEHAVIOR INVENTORY
WITH TOTAL SCORE ON STANDARD SCALES

	SRS	*SCQ*	*RBS-R*	*CASI*	*CGSQ*
Social communication	0.76**	0.79**	0.71**	0.24	0.22
Restrictive and repetitive behavior	0.83**	0.80**	0.89**	0.32	0.47
Mental health	0.31	−0.06	0.03	0.73**	0.22
Self-regulation	0.22	−0.17	0.02	0.61*	0.12
Challenging behavior	0.37	0.24	0.35	0.27	0.14

** $p < 0.001$; * $p < 0.01$.
CASI, Child and Adolescent Symptom Inventory, Anxiety Subscale; CGSQ, Caregiver Strain Questionnaire; RBS-R, Repetitive Behavior Scale–Revised; SCQ, Social Communication Questionnaire; SRS, Social Responsiveness Scale.

Communication Questionnaire (SCQ), RBS-R, CASI-5-ANX, Caregiver Strain Questionnaire (CGSQ–short form 7).

Correlations between the ABI-F core domains and standard rating scale total scores for caregiver ratings are summarized in Table 5. Pearson's correlations between the ABI domains and corresponding scales were strongly positive ($r \geq 0.70$) and highly significant ($p < 0.001$), showing concurrent validity between the ABI and scales designed to measure similar constructs. In addition, divergent validity was shown between the ABI domains and scales measuring different constructs. For example, there was no significant correlation between ABI mental health domain and the RBS-R.

Item reduction analysis. The assessments were based on only a modest sample size compared with phase 1 ($n = 37$, clinician and parent data); thus, the IDI was used instead of the Item Information Function. Items with values with IDI <0.003 (considered as poor items) were eliminated if they did not affect the alpha or construct validity of the assessment. Moreover, a clinical expert panel (nine experts) underwent three rounds of consultation and feedback based on the Delphi process (Dalkey 1972) to reach consensus on whether to retain or eliminate items for the ABI-S. Repetitive items (IDI >0.8 and confirmed by independent clinician review) were merged into single items or the best-performing item (based on alpha) was kept. The panel considered items for inclusion in the ABI-S, which were of clinical importance, most likely to be sensitive to change and more likely to be assessed through observation and interview by HCPs.

In addition, item reduction analysis carried out during this phase led to further scale refinement resulting in the ABI-F (73 items) and the ABI-S (35 items). All items comprised a subset of the ABI-F (Table 6).

Discussion

There exists a paucity of validated outcome measures to assess severity and change in core and associated behaviors in ASD as also elucidated in several pivotal reviews and an editorial (Lecavalier et al. 2014; Aman et al. 2015; Anagnostou et al. 2015; McConachie et al. 2015; Scahill et al. 2015). The ABI was developed through an iterative process, involving expert consultation, statistical validation, and parent feedback, resulting in the current scale, which will be available free for professional use, including for further validation studies and use in evaluation of interventions.

Correlations between ABI core domains and commonly used existing scales measuring similar constructs demonstrated good

TABLE 6. FINAL LIST OF ITEMS FOR THE AUTISM BEHAVIOR INVENTORY SHORT FORM AND LONG FORM

DOMAIN subdomain	Attribute	Anchor type[a]	Short form
Social communication			
Reciprocity	\<name\> responds to familiar things	Q/C	
Reciprocity	\<name\> resists affection from familiar people	F/I	
Reciprocity	\<name\> shows appropriate affection toward familiar people	Q/C	
Reciprocity	\<name\> shows inappropriate affection toward unfamiliar people	F/I	
Reciprocity	\<name\> comments on other people's emotions	F/C	
Reciprocity	\<name\> looks when he/she is called or praised	F/C	x
Reciprocity	\<name\> looks where another person is looking and pointing	Q/C	x
Reciprocity	\<name\> shows an interest in what other people are doing	Q/C	x
Reciprocity	\<name\> responds to attempts to initiate social interaction	Q/C	x
Reciprocity	\<name\> imitates an action sometime after he/she originally observed it	F/C	
Reciprocity	\<name\> gives things to others to get help	Q/C	
Reciprocity	\<name\> shows pleasure in shared interactions	F/C	
Reciprocity	\<name\> engages in make believe play with another person	Q/C	x
Reciprocity	\<name\> has difficulty interacting with peers	F/I	
Reciprocity	\<name\> is able to take turns in conversation	Q/C	x
Nonverbal communication	\<name\> directs facial expressions toward other people to communicate feelings	Q/C	
Nonverbal communication	\<name\> uses facial expressions that are appropriate to the situation	F/C	x
Nonverbal communication	\<name\> waves ''Hello'' and ''Goodbye''	Q/C	
Nonverbal communication	\<name\> uses common gestures	Q/C	x
Nonverbal communication	\<name\> uses gestures to request something	Q/C	x
Nonverbal communication	\<name\> uses gestures to comment on what is happening	Q/C	
Nonverbal communication	\<name\> requests items by pointing	F/C	x
Nonverbal communication	\<name\> combines gestures with vocalizations to enhance communication	Q/C	x
Verbal communication	\<name\> says socially inappropriate things	F/I	
Verbal communication	\<name\> uses tone of voice appropriately to emphasize content of speech	Q/C	
Verbal communication	\<name\> offers information about his/her own thoughts or feelings	Q/C	
Verbal communication	\<name\> attends to parts of a sentence and misinterprets the whole	F/I	
Restrictive repetitive behaviors			
Resistance to change	\<name\> gets upset over small changes in routine	F/I	x
Resistance to change	\<name\> has difficulty being flexible	F/I	x
Resistance to change	\<name\> resists trying out new things	F/I	
Resistance to change	\<name\> insists on doing things the same way each time	F/I	x
Restricted interests	\<name\> is fixated on certain topics or activities and unable to move on	F/I	x
Restricted interests	\<name\> has an unusually narrow range of interests	F/I	
Stereotypical behaviors	\<name\> repeats/echoes what others say	F/I	
Stereotypical behaviors	\<name\> insists on saying words or phrases over and over	F/I	x
Stereotypical behaviors	\<name\> has mannerisms or odd ways of moving her/his hands or fingers	F/I	x
Stereotypical behaviors	\<name\> makes repetitive movements	F/I	x
Stereotypical behaviors	\<name\> uses objects repetitively	F/I	x
Stereotypical behaviors	\<name\> attempts to harm himself/herself	F/I	x
Hypersensitivity	\<name\> over-reacts to common smells	F/I	
Hypersensitivity	\<name\> over-reacts to noise or sounds	F/I	
Hypersensitivity	\<name\> over-reacts to touch or being held	F/I	
Hypersensitivity	\<name\> has sensitivities to certain food textures	F/I	
Mental health			
Irritability	\<name\> cries over minor annoyances and hurts	F/I	
Irritability	\<name\> is irritable and whiny	F/I	x
Anxiety	\<name\> worries about things	F/I	
Anxiety	\<name\> complains about physical problems, without a known medical reason	F/I	
Anxiety	\<name\> is tense or anxious	F/I	x
Anxiety	\<name\> looks worried or concerned	F/I	
Anxiety	\<name\> clings to adults or is too dependent on them	F/I	
Anxiety	\<name\> gets upset when separated from a parent	F/I	x
Anxiety	\<name\> is anxious in social situations	F/I	x
Depression	\<name\> appears sad	F/I	
Phobia	\<name\> is fearful of specific objects or situations	F/I	
Self-regulation			
Impulsivity	\<name\> has difficulties waiting his/her turn	F/I	x
Impulsivity	\<name\> acts without thinking	F/I	x
Impulsivity	\<name\> acts recklessly	F/I	

TABLE 6. (CONTINUED)

DOMAIN subdomain	Attribute	Anchor type[a]	Short form
Hyperactivity	\<name> switches quickly from one topic or activity to another	F/I	x
Hyperactivity	\<name> has difficulties playing or engaging in leisure activities quietly	F/I	x
Hyperactivity	\<name> fidgets	F/I	
Hyperactivity	\<name> has difficulty remaining seated	F/I	
Hyperactivity	\<name> is excessively active	F/I	
Sleep	\<name> has sleep problems	F/I	x
Challenging behavior			
Aggression	\<name> is verbally aggressive toward other children or adults	F/I	x
Aggression	\<name> is physically aggressive toward other children or adults	F/I	x
Aggression	\<name> reacts with aggression when he/she is upset or stressed	F/I	x
Aggression	\<name> throws things inappropriately	F/I	
Aggression	\<name> is mean to animals	F/I	
Aggression	\<name> hits or kicks	F/I	
Elopement	\<name> runs away	F/I	x
Antisocial behavior	\<name> takes or grabs things that belong to others	F/I	
Temper tantrums	\<name> screams, yells, and cries	F/I	
Temper tantrums	\<name> has temper outbursts or tantrums	F/I	x

[a]Anchor combinations: frequency (F), intensity (I), quality (Q), context (C). A pdf version of the ABI is available at https://www.janssenmd.com/
ABI, Autism Behavior Inventory; x, items included in the ABI-S.

construct validity. In addition, as hypothesized, there were no, or smaller, correlations between some of the ABI domains and the scales measuring different constructs, demonstrating good divergent validity of the ABI. Thus, the ABI measures a range of behaviors that are sometimes co-occurring with, but not part of, the core symptomology.

Test–retest reliability was shown to be good, and the core ASD domains rated by parents had comparable or better test–retest correlations than standard scales rated at the same time points. HCP-rated ABI scores showed lower correlations between visits compared with parents, indicating lower test–retest reliability. This is not surprising, given that parents and clinicians had different roles and observed behaviors in very different contexts (Achenbach et al. 1987). HCPs report on a small sample of behavior observed only during the clinic visit of the patient, while parents draw on broader observations in a range of contexts over vastly longer periods of time spent with the patient. Behavior in the clinics may have been somewhat "rarefied" in the sense that children with ASD tend to respond negatively to change in routine; early visits to the clinics may have been accompanied by emotional lability not observed later on.

Ongoing studies are currently evaluating the ability of HCPs to use the ABI, specifically the ABI-S, from the perspectives of burden, validity, and reliability. Given the unacceptable alpha levels, particularly for the social communication domain of the ABI-S, a revised set of items was developed with the involvement of a Delphi panel. The items were suggested based on likelihood that they were observable in a clinic setting or most likely to be sensitive to change, in addition to consideration of how the items performed statistically. This enabled retention of items that we felt were of clinical importance, but with less statistical validation in this small sample. Around 50% of the items met the criteria for inclusion based on statistical performance and clinician panel opinion. Other items, such as "has temper outbursts or tantrums when he doesn't get his own way" and "makes repetitive movements," would not have been included in the ABI-S based on statistical performance, but were included based on strong consensus from the panel. In addition to careful consideration of the items to be included in the

ABI-S, an observation period and semistructured interview have been added to the HCP ABI process for future studies, to enhance the ability of the clinician to respond to the items.

Parents and caregivers may complete the ABI, or the ABI-S. The creation of a scale that is both easy to use and measures the behaviors that are of interest and importance to parents has been a central part of the development process. Both phase 1 and 2 incorporated feedback on content and usability, in accordance with FDA guidelines (U.S. Department of Health and Human Services 2009). Acceptability by parents was good, in terms of understanding of items and anchors, and ability to access and navigate the scale. A small number of items were removed due to difficulties with parents understanding or being able to report on a particular behavior for their child. A large majority of parents in the first phase found the scale to be comprehensive, but the addition of items such as sleep and food sensitivity ensured that behaviors reported as key concerns for parents were included in the ABI. We took other feedback into account, such as the use of progress bar, and incorporated several features to personalize the scale to enhance parents' experiences when completing it.

The ABI differs from other scales in that it was developed for online/tablet administration and validated in this context. This online administration format reduces burden on participants, and may improve study adherence and reduce attrition (Gwaltney et al. 2008). It offers convenience to families to complete the assessments from home on devices parents are comfortable and familiar with, which minimizes the need to travel for a study visit. The system can remind parents when a rating is due and confirm when the family completes the scale. This helps to ensure that scales are completed within expected time frames, and prevents backdating of scale completion without the researcher's knowledge. Reduction of nonresponse to items is inherent in the scale design, as missing items are highlighted and required before a respondent is able to move on. A further advantage is the possibility of instantaneous data transfer and scoring.

The online presentation of the ABI allows adaptability of the scale. The ability to customize the ABI may improve the experience

for parents by reducing burden of completing questions that are not relevant. For example, the addition of filters to exclude items not appropriate for certain ages or developmental levels means that parents are not required to repeatedly complete items at ceiling or baseline. Another benefit is that the software can provide participants (and separately, researchers) with a display of their individual outcomes relative to previous ratings or relative to appropriate normative groups to make interpretation of change relatively easy. Thus, several potential benefits make online completion of ABI the preferred format. However, paper versions of the scale have also been created that allow for more traditional paper-and-pen format.

A further advantage of the ABI is the inclusion of a number of subscales within one tool. This also has potential to reduce parent burden, while enabling researchers to get an indication of change in both core and most commonly associated symptoms. For example, the strong correlation between the CASI-ANX and the mental health domain (0.73) may mean that the ABI could be completed without the need for a scale that measures core symptoms, and an additional scale to measure anxiety, if these correlations can be replicated in larger studies.

The current study has several limitations. ASD diagnoses of participants were based on a parent report in the phase 1 item generation study and through medical record in the phase 2 item reliability study. Whereas in phase 2, clinically experienced research staff validated the ASD diagnosis based on direct observation using the CARS-2, this tool is not considered the gold standard. The original sample for item generation was large ($n = 353$) and the age range and breadth of impairment were substantial. However, the phase 2 sample for item reliability did not include individuals with a wide range of IQ and language ability, and included only a small proportion of adults. Due to concern over parent burden, measures of externalizing behaviors, such as the ABC, were not used as a comparison, and therefore, we were unable to ascertain the concurrent validity of domains such as self-regulation and challenging behavior. In addition, although we established strong test–retest reliability over an 8-week interval observational study, further research is needed to establish the sensitivity to change of the ABI. As with other scales in development, such as the AIM (Kanne et al. 2014), the ABI sensitivity must await future studies of various interventions. In an ongoing study, we are identifying sensitivity to change with other concomitant biosensor information, and future intervention studies using the ABI as a primary outcome measure are planned. These will also involve additional standard scales for comparison and inclusion of participants with a broader range of ability and also of age (in particular, more adults).

Conclusions

This study provides initial support for the ABI and is a critical first step in indicating robust reliability and validity of the web-based tool for assessing ASD symptoms. Overall, the results indicate that ABI holds promise as a measure of clinically relevant aspects in ASD symptomatology. Studies are ongoing with a larger sample to evaluate the validation of change in the context of intervention, and to improve clinician ratings through revised ABI-S, direct observation, and standardized interview schedules.

Clinical Significance

The ABI scale has performed well, showing high internal consistency and moderate-to-high associations with standardized paper-and-pencil measures. Daily measures of behavior also showed statistically significant correlations with the ABI and other scales. Daily tracking of symptoms by parents could be a useful way of identifying early change in response to the intervention and is a critical first step in indicating robust reliability and validity of the web-based tool for assessing ASD symptoms.

Authors' Contributions

Abi Bangerter, Seth Ness, Gahan Pandina, Mark Opler, Michael Aman, Anna Esbensen, Matthew Goodwin, Geri Dawson, Adrianne Harris, Bennett Leventhal, and Robert Hendren were involved in study design, data collection, analysis, and interpretation. Anzalee Khan was responsible for the statistical analyses. All authors were involved in interpretation of the results. All authors had full access to all the data in the study and take responsibility for integrity of the data and the accuracy of the data analysis.

Disclosures

This study was funded by Janssen Research & Development, LLC. Seth Ness and Gahan Pandina are employees of Janssen Research & Development, LLC, and hold company stocks/stock options. Anna Esbensen has consulted with Roche on outcome measures in Down syndrome. Robert Hendren received reimbursement for consultation from Janssen Research & Development, LLC. Michael Aman has received research contracts, consulted with, served on advisory boards, or done investigator training for AMO Pharma, CogState, Inc.; CogState Clinical Trials, Ltd.; Coronado Biosciences; Forest Research; Hoffman-La Roche; Johnson and Johnson; Lumos Pharma, MedAvante, Inc.; ProPhase LLC; and Supernus Pharmaceuticals. Geraldine Dawson is on the Scientific Advisory Boards of Janssen Research and Development and Akili, Inc., a consultant to Roche, has received grant funding from Janssen Research & Development, LLC, and PerkinElmer, and receives royalties from Guildford Press and Oxford University Press.

References

Achenbach TM, McConaughy SH, Howell CT: Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. Psychol Bull 101:213–232, 1987.

Achenbach TM, Rescorla LA: Manual for the ASEBA School-Age Forms & Profiles. Burlington, Research Centre for Children, Youth and Families, University of Vermont, 2001.

Akhondzadeh S, Tajdar H, Mohammadi MR: A double-blind placebo controlled trial of piracetam added to risperidone in patients with autistic disorder. Child Psychiatry Hum Dev 39:237–245, 2008.

Aman MG, Arnold LE and Hollway JA: Assessing change in core autism symptoms: Challenges for pharmacological studies. J Child Adolesc Psychopharmacol 25:282–285, 2015.

Aman MG, Novotny S, Samango-Sprouse C: Outcome measures for clinical drug trials in autism. CNS Spectr 9:36–47, 2004.

Anagnostou E, Jones N, Huerta M: Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. Autism 19:622–636, 2004.

Dalkey N: The Delphi method: An experimental study of group opinion. Available at www. rand. org/pubs/research_memoranda/2005 RM5888. pdf (last accessed on December 8, 2010), 1972.

Ghosh A, Michalon A, Lindemann L: Drug discovery for autism spectrum disorder: Challenges and opportunities. Nat Rev Drug Discov 12:777–790, 2013.

Gwaltney CJ, Shields AL, Shiffman S: Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-analytic review. Value Health 11:322–333, 2008.

Kanne SM, Mazurek MO, Sikora D, Bellando J, Branum-Martin L, Handen B, Katz T, Freedman B, Powell MP, Warren Z: The autism impact measure (AIM): Initial development of a new tool for treatment outcome measurement. J Autism Dev Disord 44:168–179, 2014.

Karami H: An introduction to differential item functioning. Int J Educ Psychol Assess 11:59–76, 2012.

King BH, Hollander E, Sikich L: Lack of efficacy of citalopram in children with autism spectrum disorders and high levels of repetitive behavior: Citalopram ineffective in children with autism. Arch Gen Psychiatry 66:583–590, 2009.

Lam KS, Aman MG: The Repetitive Behavior Scale-Revised: Independent validation in individuals with autism spectrum disorders. J Autism Dev Disord 37:855–866, 2007.

Le Couteur A, Lord C, Rutter M: Autism Diagnostic Interview-Revised (ADI-R). Los Angeles (California), Autism Genetic Resource Exchange, 2003. Available at www.agre.org/program/aboutadi.cfm (last accessed on March 5, 2009)

Lecavalier L, Wood JJ, Halladay AK: Measuring anxiety as a treatment endpoint in youth with autism spectrum disorder. J Autism Dev Disord 44:1128–1143, 2014.

Lord C, Rutter M, Le Couteur A: Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 24:659–685, 1994.

Marcus RN, Owen R, Kamen L: A placebo-controlled, fixed-dose study of aripiprazole in children and adolescents with irritability associated with autistic disorder. J Am Acad Child Adolesc Psychiatry 48:1110–1119, 2009.

Marcus RN, Owen R, Manos G: Aripiprazole in the treatment of irritability in pediatric patients (aged 6–17 years) with autistic disorder: Results from a 52-week, open-label study. J Child Adolesc Psychopharmacol 21:229–236, 2011.

Matson JL, Dempsey T, LoVullo SV: The moderating effects of intellectual development on core symptoms of autism and PDD-NOS in toddlers and infants. Res Dev Disabil 34:573–578, 2013.

McConachie H, Parr JR, Glod M: Systematic review of tools to measure outcomes for young children with autism spectrum disorder. Health Technol Assess 19:1–506, 2015.

McCracken JT, McGough J, Shah B: Risperidone in children with autism and serious behavioral problems. N Engl J Med 347:314–321, 2002.

Research Units on Pediatric Psychopharmacology Autism Network: Randomized, controlled, crossover trial of methylphenidate in pervasive developmental disorders with hyperactivity. Arch Gen Psychiatry 62:1266–1274, 2005.

Rogers HJ: Differential Item Functioning. Encyclopedia of Statistics in Behavioral Science, Chichester: John Wiley & Sons, Ltd, 2005.

Scahill L, Aman MG, Lecavalier L: Measuring repetitive behaviors as a treatment endpoint in youth with autism spectrum disorder. Autism 19:38–52, 2015.

Scahill L, McDougle CJ, Aman MG: Effects of risperidone and parent training on adaptive functioning in children with pervasive developmental disorders and serious behavioral problems. J Am Acad Child Adolesc Psychiatry 51:136–146, 2012.

Singh K, Connors SL, Macklin EA: Sulforaphane treatment of autism spectrum disorder (ASD). Proc Natl Acad Sci U S A 111:15550–15555, 2014.

Sparrow SSCD, Balla DA: Vineland-II Adaptive Behavior Scales: Survey Forms Manual. Circle Pines (Minnesota), American Guidance Service, 2005.

Sukhodolsky DG, Scahill L, Gadow KD: Parent-rated anxiety symptoms in children with pervasive developmental disorders: Frequency and association with core autism symptoms and cognitive functioning. J Abnorm Child Psychol 36:117–128, 2008.

U.S. Department of Health and Human Services: Guidance for industry. Patient-reported outcome measures: Use in medical product development to support labeling claims. 2009. Available at www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf (last accessed on January 14, 2017).

Address correspondence to:
Abi Bangerter, MA
Janssen Research & Development, LLC
Titusville, NJ 08534

E-mail: abangert@its.jnj.com;
autismbehaviorinventory@its.jnj.com