

Validity of the Aberrant Behavior Checklist in Children with Autism Spectrum Disorder

Aaron J. Kaat · Luc Lecavalier · Michael G. Aman

Published online: 29 October 2013
© Springer Science+Business Media New York 2013

Abstract The Aberrant Behavior Checklist (ABC) is a widely used measure in autism spectrum disorder (ASD) treatment studies. We conducted confirmatory and exploratory factor analyses of the ABC in 1,893 children evaluated as part of the Autism Treatment Network. The root mean square error of approximation was .086 for the standard item assignment, and in exploratory factor analysis, the large majority of items continued to load on the originally assigned factors. Correlations between the ABC subscales and multiple external variables including the Child Behavior Checklist and demographic variables supported the convergent and divergent validity of the ABC as a measure of behavior problems in ASD. Finally, we examined the impact of participant characteristics on subscale scores and present normative data.

Keywords Autism spectrum disorder · Aberrant Behavior Checklist · Factor analysis · Irritability · Rating scales · Assessment

Introduction

The Aberrant Behavior Checklist (ABC; Aman et al. 1985) was developed for assessing treatment effects in people with

developmental disabilities. Since its introduction, the ABC has been used in over 325 studies, and it has been translated into more than 30 languages other than English (Aman 2012a). It was developed in samples that were primarily composed of adolescents and adults in residential facilities, although modest numbers of children were included, and some samples were drawn from group homes as well. The ABC was derived by factor analysis, yielding five subscales: Irritability (15 items), Lethargy/Social Withdrawal (16 items), Stereotypic Behavior (7 items), Hyperactivity/Non-compliance (16 items), and Inappropriate Speech (4 items). Although scores on some of the subscales (especially Irritability and Hyperactivity/Noncompliance) are moderately correlated, use of the total score lacks construct validity and is strongly discouraged in the ABC manual (Aman and Singh 1986) and elsewhere (Aman 2012b).

Later, the ABC was revised to eliminate references to residential terminology (e.g., “ward” was replaced) and more explicitly to allow use in children (e.g., referencing activities at school). This led to two versions, one called the ABC-Residential and the other ABC-Community (Aman and Singh 1986; Marshburn and Aman 1992). Most of the research among children and adolescents has had parents rate the ABC. However, teachers have completed the ABC for some psychometric evaluations (Freund and Reiss 1991; Marshburn and Aman 1992), and drug studies [e.g., Research Units on Pediatric Psychopharmacology (RUPP) Autism Network 2005a]. Other raters who know the child (or adult) well may also complete the ABC. For example, in the original developmental study, direct care staff members in residential facilities often completed the ABC (Aman et al. 1985).

Use of the ABC has significantly grown over the years, especially among children and adolescents with intellectual disability (ID) and/or with autism spectrum disorder (ASD; Aman 2012a). The Irritability subscale was the primary

A. J. Kaat
Department of Psychology, Nisonger Center, The Ohio State University, Columbus, OH 43210, USA

L. Lecavalier · M. G. Aman (✉)
Department of Psychology, Nisonger Center, The Ohio State University, 1581 Dodd Drive Room 175C, Columbus, OH 43210, USA
e-mail: aman.1@osu.edu
URL: <http://www.psychmed.osu.edu>

outcome measure in pivotal large multi-site studies in ASD of risperidone (RUPP 2002; RUPP 2005b) and of aripiprazole (Marcus et al. 2009; Owen et al. 2009), which led to FDA clinical indications. The ABC has also been used to measure treatment response to psychosocial interventions (e.g. Aman et al. 2009; Tse et al. 2007), to evaluate phenotypic classifications (e.g. Oliver et al. 2008; Walley and Donaldson 2005), and as part of large-scale ASD databases [e.g., the Autism Treatment Network (ATN) and Simons Simplex Collection].

In recent years, there has been continued interest in the ABC's factor structure and psychometric properties in children and adults. Most studies have found substantial concordance with the original factor structure, but exploratory analyses have yielded various structures ranging from four factors (e.g., Brinkley et al. 2007; Brown et al. 2002) to six (e.g., Sansone et al. 2012). Several observations have been noted across studies. First, various items from the Irritability and Hyperactivity/Noncompliance factors occasionally crossed to the alternate factor. Second, Inappropriate Speech did not always emerge as a factor, resulting in four factors on two instances (e.g., Brown et al. 2002; Marshburn and Aman 1992). Third, the three self-injurious behavior (SIB) items emerged as a separate factor in one instance (e.g., Brinkley et al. 2007), or performed better as an item parcel rather than as separate items in another instance (e.g. Sansone et al. 2012). Finally, in one study involving males with Fragile-X syndrome, a condition characterized by social anxiety, social avoidance, shyness, and gaze aversion (Budimirovic and Kaufmann 2011), Sansone et al. (2012) found that Lethargy and Social Withdrawal split into separate factors. This finding has not been demonstrated in other clinical samples.

Although there have been several evaluations of the ABC among children and adolescents, only Brinkley et al. (2007) had a sample entirely of children and adolescents with ASD ($n = 275$ between the ages of 3 and 21 years with an average age of 10.6 years). Most studies of children and adolescents have been in samples largely selected for ID.

As use of the ABC has increased in assessing individuals with ASD, further psychometric evaluation of the ABC in this population was necessary. The purpose of this study was to explore the factor structure of the ABC in a large sample of children with ASD ($n = 1,893$) who were not selected for clinically-significant behaviors, such as high Irritability scores or having a specific clinical condition, such as Fragile-X. Presumably, this allows for more representativeness and more stable factor solutions. Such a large sample allowed (a) comparison of various models, (b) cross-validation using split-samples, and (c) examination of associations with subject characteristics such as age, level of functioning, and severity of autistic symptoms on factor structure. We expected that the original factor

structure would accurately describe youth with ASD, regardless of age, level of functioning, and severity of autism symptoms.

In addition to examining the factor structure of the ABC, secondary aims included assessment of convergent and divergent validity with measures of ASD severity and behavioral and emotional problems. ABC scores were examined in relation to the *Autism Diagnostic Observation Schedule* (ADOS; Gotham et al. 2009; Lord et al. 2000) and Child Behavior Checklist (CBCL) scores, including the Pervasive Developmental Problems (PDP) subscale for children younger than 6 years old (Achenbach and Rescorla 2000, 2001). We expected positive correlations between ABC and CBCL subscales that assess externalizing symptoms and between Lethargy/Social Withdrawal and CBCL PDP and Withdrawn Behavior subscales, which is consistent with previous research on the ABC among preschoolers with developmental disabilities, including ASD (Karabekiroglu and Aman 2009).

Finally, such a large national sample from the USA and Canada allowed examination of the impact of subject characteristics, such as age, IQ, and adaptive behavior, on subscale scores. We hypothesized that significant relationships with age would be found, which is consistent with previous longitudinal studies (e.g. Anderson et al. 2011), especially reductions in hyperactivity with increasing age. We also predicted that Stereotypic Behavior would be increased in children with lower IQ scores, which is consistent with previous research on repetitive behaviors (e.g. Gabriels et al. 2005). Correlations with adaptive behavior were exploratory.

Methods

Participants

This study used archival data from the ABC collected as part of the ATN enrollment. The ATN is a network of 17 children's hospitals across the USA and Canada which provide diagnostic and treatment services to children with ASD. As of November 2012, there were 3,132 parent-completed ABCs in the ATN database. For those with multiple ABC ratings ($n = 33$), only the first-completed one was used. In order to create a more homogenous sample while still allowing for the heterogeneity associated with ASD, we only included individuals between the ages of 2 and 18 years who met criteria for autism or ASD on Modules 1, 2, or 3 of the ADOS (Lord et al. 2000) utilizing the revised algorithms (Gotham et al. 2007). This resulted in a full sample of 1,893 individuals across 14 ATN sites. The full sample was randomly split into two subsamples stratified by ATN site (to ensure proportionality across

Table 1 Demographics characteristics of the Autism Treatment Network sample

	Full sample	Calibration sample	Validation sample	Sample difference
Sample size	1,893	1,130	763	
Age in years at time of ABC completion <i>M</i> (<i>SD</i>)	6.5 (3.6)	6.4 (3.5)	6.7 (3.6)	$t(1,891) = -1.83, p = .07$
Percent male	83.7 %	83.6 %	83.7 %	$\chi^2(1) = .01, p = .95$
Percent with IQ ≤ 70	47.0 %	46.6 %	47.4 %	$\chi^2(1) = .12, p = .73$
Percent Caucasian	85.8 %	86.5 %	84.8 %	$\chi^2(1) = 1.03, p = .31$
ADOS classification: percent with autism	88.1 %	88.1 %	87.9 %	$\chi^2(1) = .02, p = .90$
ADOS classification: percent with other ASD	11.9 %	11.9 %	12.1 %	

sites) for the purpose of validation. The first sample included 60 % of the participants ($n = 1,130$), with whom more exploratory analyses [exploratory factor analysis (EFA) and analysis of competing ABC structures] were conducted. The second sample included 40 % of participants ($n = 763$) and was used as a Validation sample to cross-validate the optimal structure (i.e., the one chosen from the earlier structures). Table 1 presents demographic information for the full sample and each subsample.

Multiple measures of developmental and intellectual functioning were used. The ATN maintains a clinical database on all participants who are diagnosed with an ASD. For IQ assessment, clinicians chose the best measure for each child. Because scores on the various abbreviated and full IQ measures are not comparable, IQ was dichotomized at 70 for most analyses, but there was a sufficiently large subsample with scores on the full Stanford Binet-5 (SB5; Roid, 2003) to include it in various validity analyses.

Measures

Aberrant Behavior Checklist (ABC; Aman et al. 1985)

The ABC is a 58-item behavior rating scale used to measure behavior problems across five subscales. Items are rated on a 4-point Likert scale (ranging from 0 [not at all a problem] to 3 [The problem is severe in degree]), with higher scores indicating more severe problems. Parents generally completed the ABC upon enrollment in the ATN if it was not included during the diagnostic assessment.

Autism Diagnostic Observation Schedule

The ADOS is a semi-structured clinician-child interaction involving several presses for specific social behaviors. It is composed of four modules based on the person's language abilities and age. Ratings on the ADOS resulted in classifications of nonspectrum, autism spectrum, or autism. Algorithms for Modules 1-3 were revised to improve sensitivity and specificity (Gotham et al. 2007). This enabled us to derive an ASD-calibrated severity score

(Gotham et al. 2009), called the Comparison Score with the ADOS-Second Edition (ADOS-CS). The ADOS was completed by trained clinicians or individuals working under their supervision as part of the diagnostic assessment.

Vineland Adaptive Behavior Scales-II

The Vineland Adaptive Behavior Scales-II (VABS-II; Sparrow et al. 2005) was completed by parents in interview or rating scale format. It included four domains (Communication, Daily Living Skills, Socialization, and Motor Skills) and 11 subdomains. The VABS-II also resulted in an Adaptive Behavior Composite (VABS-C) based on the domain scores (excluding Motor Skills for older children). The VABS-II scores measured adaptive behavior relative to developmental expectations and same-aged peers. As such, older children have more expectations to reach an average score than younger children. The ATN does not require uniform administration of the VABS-II. As such, it may have been completed in interview format with a clinician or it may have been completed as a rating scale at a time convenient for the family during the diagnostic process.

Stanford Binet-5 (SB-5)

The SB5 (Roid 2003) is an individually administered IQ test. It is composed of five cognitive factors, each measured nonverbally and verbally. In addition to factor scores, the SB5 results in Nonverbal, Verbal, and Full Scale IQ composites. It also compares children relative to same-aged peers. The SB5 was completed by clinicians or assessors working under their supervision as part of the diagnostic assessment.

Child Behavior Checklist (CBCL; Achenbach and Rescorla 2000, 2001)

The CBCL is a parent-completed rating scale of challenging behavior. It measures both internalizing and externalizing

symptoms. The CBCL was empirically-derived via factor analysis, but it also includes DSM-oriented subscales. There are two versions of the CBCL depending on a child's age (Preschool version; 1.5–5 years or School-age version; 6–18 years). Items and subscale composition differ across CBCL versions. Parents generally completed the CBCL upon enrollment in the ATN if it was not included during the diagnostic assessment and then annually thereafter. The completion closest in date to completing the ABC was chosen for analysis in this study.

Analytic Plan

Exploratory factor analysis is usually conducted when there is insufficient empirical or theoretical support for the factor structure of a scale. Although the ABC has previous factor structures that were known, an EFA was conducted because few studies have examined only children and adolescents, and only one previous study was confined to ASD. EFA also allows examination of the optimal dimensionality of the ABC. The EFA was conducted in the Calibration sample using ordinary least squares estimation with oblique Crawford-Ferguson Quartimax rotation on the polychoric correlation matrix, as implemented in the Comprehensive Exploratory Factor Analysis program (Browne et al. 2008). Choice of dimensionality was guided by the eigenvalues >1.0 rule, examination of the scree plot, and clinical meaningfulness (c.f., Norris and Lecavalier 2010).

Factor structures previously proposed in the literature (Brinkley et al. 2007; Brown et al. 2002; Sansone et al. 2012) were submitted to categorical confirmatory factor analysis (CFA) in the Calibration sample. The expectation–maximization (EM) algorithm was used to impute missing values to maximize sample size. The EM algorithm is an iterative model-based approach used to predict missing values assuming underlying multivariate normality. Analyses were conducted using diagonally-weighted least squares (DWLS) estimation on the polychoric correlation matrix and sample-estimated asymptotic covariance matrix, as implemented in Lisrel 8.8 (Jöreskog and Sörbom 2007).

The optimal model (theoretically and by measures of fit) from the EFA and CFAs was reanalyzed in the Validation sample, where a CFA was conducted using similar procedures. The internal consistency of the optimal model was also examined in both samples via Cronbach's alpha.

Convergent and divergent validity of the optimal model was then examined in the full sample. These analyses were conducted under listwise deletion for the ABC (i.e., no imputed values were used) and pairwise deletion for external correlates (i.e., an individual was not required to have scores on all the VABS-II, SB5, and CBCL variables to be included in pairwise comparisons). Pearson

correlations were used for all analyses except correlations with the ADOS-CS. The ADOS-CS is an ordinal rating and, as such Spearman correlations were used. Given the large sample, size and number of analyses, only correlations significant at the $p < .01$ level were reported. We also reported strength of association, following Cohen's (1992) classification: negligible ($r < .10$), small ($.10 \leq r < .30$), medium ($.30 \leq r < .50$), and large ($r \geq .50$).

Assessing Model Fit

A combination approach to evaluating models was used, as no single fit statistic is adequate for all analyses. Thus, several model fit statistics for the CFA analyses are reported. First, the Satorra-Bentler Chi square ($SB-\chi^2$) was used to measure absolute fit (Satorra and Bentler 1994). The $SB-\chi^2$ is a mean-adjusted statistic that corrects for kurtosis in the data and is robust against non-normal data. A significant $SB-\chi^2$ suggests that the model does not fit due to residual covariation in the data. The $SB-\chi^2$ per degrees of freedom ratio is also reported, where lower values indicate better fit. The Root Mean Square Error of Approximation (RMSEA) measures model fit while adjusting for the complexity of the model. Browne and Cudeck (1992) suggested that RMSEA values <.05 indicate good fit, between .05 and .08 indicate reasonable fit, between .08 and .10 marginal fit, and values greater than .10 unacceptable fit. The Standardized Root Mean Square Residual (SMR) was also examined. The SMR measures discrepancy between the observed and predicted correlations, where values less than .08 are considered good fit (Hu and Bentler 1999). Measures of fit reflect residual (i.e., unmodeled) covariance in the data and thus are unrelated to significance or magnitude of factor loadings.

Results

Exploratory Factor Analysis

In the Calibration sample, 11 eigenvalues were >1.0. However, examination of the scree plot supported a 5-factor solution. Examination of a 4-factor solution appeared similar to the original structure without an Inappropriate Speech factor. The 5-factor solution was remarkably similar to the original factor structure. Items were considered to have moved factors if the highest factor loading was on a new factor even if there remained a substantial cross-loading on the original factor. Compared to the original structure, two items (#18 and #24) moved from Hyperactivity to Irritability; one item (#21) moved from Hyperactivity to Inappropriate Speech; one item (#43) moved from Lethargy/Social Withdrawal to

Inappropriate Speech; one item (#46) moved from Inappropriate Speech to Lethargy/Social Withdrawal; and one item (#51) moved from Hyperactivity to Lethargy/Social Withdrawal. Thus 52 of 58 items (90 %) continued to load primarily on their originally assigned factors (Aman et al. 1985). Several items had significant cross-loadings. The 6-factor solution was largely equivalent to the 5-factor solution with a 3-item SIB factor strongly correlated with the remaining Irritability items and little relationship with other factors. Table 2 shows factor loadings for the 5-factor solution. Items are grouped according to EFA subscale assignment, which largely coincides with the original assignment, and ranked by factor loading magnitude. Factor loadings for the other EFAs are available from the corresponding author and also appear on our website (www.psychmed.osu.edu).

Note that the two far right columns of Table 2 contain the factor loading for the CFA of the original 5-factor solution for the Calibration and Validation samples, respectively. In the CFAs, the factor loadings and internal consistency ratings were quite similar across samples. The internal consistency was good to excellent for every factor (in the Calibration and Validation samples, respectively, Irritability $\alpha = .92, .92$; Lethargy/Social Withdrawal $\alpha = .88, .89$; Stereotypic Behavior $\alpha = .87, .85$; Hyperactivity/Noncompliance $\alpha = .94, .93$), excluding Inappropriate Speech, which had acceptable levels of internal consistency ($\alpha = .77, .77$).

Comparison with Previous Models

The 4-factor structures proposed by Brown et al. (2002) and Brinkley et al. (2007), the 5-factor structures proposed by Aman et al. (1985) and Brinkley et al. (2007), and the 6-factor solution proposed by Sansone et al. (2012) were fit to the Calibration sample. Table 3 provides measures of fit for these models.

All models were statistically significant and thus rejected by the SB- χ^2 test. The 4-factor model proposed by Brown et al. (2002) had the poorest fit, whereas the others had broadly equivalent marginal fit. The historical basis and widespread use of the original factor structure and results of other factor analytic studies (Aman 2012a; Aman and Singh 1986) led us to prefer it as the optimal solution both from an historical and pragmatic perspective. Table 2 includes the factor loadings for this CFA model.

Insofar as all previous models exhibited marginal fit, we examined whether the original factor structure fit better in certain subsamples. The Calibration sample was split separately by age at 6 years, by IQ at 70, and by ADOS-CS at an approximate median split of 7. Factorial invariance was not supported by any split, so the original model was fit to each subsample separately. Table 3 presents the fit

statistics for these analyses. Note that SB- χ^2/df is not reported, since it is sample-size dependent and, as such, cannot be compared across subsamples. Marginal fit was observed in all subsamples, with minimal differences between groups.

The original factor structure, as the preferred optimal solution, was also fit in the Validation sample. A one-factor Total Score model was also fit in the Validation sample. Table 2 provides factor loadings for this analysis and Table 3 provides model fit statistics for the Validation sample analyses also. As with the Calibration sample, the original structure exhibited marginal fit. The Total Score model (use of a single total score alone) exhibited *poor fit to the data* and as such should not be used. Once again SB- χ^2 could not be compared across Calibration and Validation samples.

Possible total score representations were also fit in the Validation sample. A one-factor and second order factor analysis solutions were conducted, and there was no empirical basis for a single total score.

Computation of Subscale Scores

In the full sample, summed Likert scores were created for the five ABC domains contained in the ABC scoring system. Henceforth, these domains are referred to as subscales. Individuals with missing items were excluded from this analysis. This left a total sample size of 1,796. The bottom of Table 2 contains the subscale correlations. As previously reported (Aman et al. 1985) and consistent with the factor analyses, the correlations were highest between the Irritability and Hyperactivity/Noncompliance subscales and lowest between Inappropriate Speech and other subscales, although the inter-subscale correlations were larger on average for this sample.

External Correlations and Effects of Participant Characteristics

The subscale scores were correlated with external variables such as age, IQ, adaptive behavior, ADOS-CS, and CBCL subscales. Not all individuals had scores on all external variables. Table 4 shows the results of these correlations. Given the large sample size and number of analyses, only correlations significant at the $p < .01$ level were reported.

Age, sex, and IQ were largely unrelated to the ABC subscale scores. Increased age was associated with decreased Irritability and Hyperactivity, whereas lower IQ scores were associated with increased Stereotypic Behavior. The reduction with age in Hyperactivity is consistent with expectations. However these correlations reflected small effects. Decreased adaptive behavior was related to increased ABC subscale scores, excluding Inappropriate

Table 2 Factor loadings for the 5-factor EFA and original structure CFAs

Abbreviated item	5-Factor EFA in calibration sample (n = 1,130)					Original structure CFAs		
	Irritability	Lethargy/social withdrawal	Stereotypic behavior	Hyperactivity/noncompliance	Inappropriate speech	Original factor	Calibration (n = 1,130)	Validation (n = 763)
50. Hurts self	.83	-.05	.15	-.02	-.21	I	.91	.84
2. Injures self	.82	-.04	.19	-.06	-.21	I	.88	.83
10. Tantrums	.81	-.02	-.05	.04	.10	I	.83	.84
52. Self-violence	.80	.00	.13	-.04	-.19	I	.87	.84
4. Aggressive	.75	-.06	-.12	.02	.05	I	.64	.68
57. Temper tantrums	.73	.06	-.04	.18	.01	I	.88	.85
36. Mood changes	.65	.14	.00	-.01	.23	I	.79	.81
41. Cries and screams	.65	.09	.10	.02	.13	I	.82	.84
18. Disobedient ^a	.62^a	.10	-.12	.34	.05	H/N	.86	.85
47. Stamps feet	.62	.01	.09	.15	.09	I	.79	.78
14. Irritable	.57	.05	-.09	.03	.30	I	.70	.70
19. Yells	.55	.01	.07	.17	.23	I	.84	.85
29. Demands	.54	.11	-.04	.24	.14	I	.82	.80
8. Screams	.52	-.03	.14	.15	.16	I	.79	.82
24. Uncooperative ^a	.51^a	.25	-.15	.31	.11	H/N	.83	.81
34. Cries	.48	.07	-.02	-.10	.36	I	.59	.63
25. Depressed mood	.35	.32	-.08	-.19	.35	I	.49	.49
42. Prefers alone	-.07	.82	.07	-.01	-.04	L/SW	.80	.81
16. Withdrawn	-.02	.79	.09	-.02	-.04	L/SW	.82	.85
30. Isolates self	.02	.78	.06	-.01	-.05	L/SW	.83	.89
5. Seeks isolation	.02	.71	.09	-.04	-.02	L/SW	.77	.78
58. Few social reactions	.02	.69	.05	.18	-.13	L/SW	.72	.73
23. Watches others	.06	.59	.11	-.22	.19	L/SW	.55	.60
20. Lacks emotional response	.09	.56	.14	-.02	.16	L/SW	.71	.71
37. Unresponsive	.01	.56	.09	.33	-.11	L/SW	.77	.73
51. Pays no attention ^a	.01	.53^a	-.02	.44	-.09	H/N	.64	.68
53. Inactive	.10	.53	.06	-.35	.18	L/SW	.42	.54
3. Sluggish	-.06	.50	-.02	-.37	.38	L/SW	.25	.26
40. Difficult to get through	.15	.48	.02	.42	-.06	L/SW	.90	.90
12. Preoccupied	-.05	.47	.24	.08	.09	L/SW	.64	.72
32. Stays in one position	.14	.47	.23	-.28	.18	L/SW	.55	.58
55. Responds negatively to affection	.34	.41	.01	-.03	.04	L/SW	.68	.70
26. Resists contact	.31	.40	.08	-.09	.12	L/SW	.67	.71
35. Repetitive hand movements	.01	.03	.90	.01	-.02	S	.87	.88
11. Repetitive movements	-.06	.07	.84	.05	.01	S	.83	.87
6. Body movements	-.08	.04	.82	.08	.04	S	.83	.80
45. Waves extremities	.07	.02	.76	.03	-.03	S	.77	.75
49. Rocks body	.22	.09	.58	.00	.01	S	.79	.75
27. Rolls head	.31	.07	.52	-.08	.06	S	.77	.68
46. Repeats words ^a	.05	.01	.38^a	.09	.37	IS	.85	.85
17. Odd behavior	.14	.26	.37	.11	.20	S	.87	.90
54. Excessively active	.11	-.11	.15	.81	-.02	H/N	.81	.82

Table 2 continued

Abbreviated item	5-Factor EFA in calibration sample (n = 1,130)					Original structure CFAs		
	Irritability	Lethargy/social withdrawal	Stereotypic behavior	Hyperactivity/noncompliance	Inappropriate speech	Original factor	Calibration (n = 1,130)	Validation (n = 763)
15. Restless	.04	−.02	.15	.75	.11	H/N	.81	.79
39. Not sit still	.02	.11	.05	.74	.03	H/N	.77	.76
1. Active	.11	−.13	.13	.73	.10	H/N	.78	.75
38. Not stay in seat	.13	.12	.01	.72	−.04	H/N	.79	.76
48. Runs or jumps	.20	−.05	.25	.60	−.06	H/N	.77	.79
28. Not pay attention	.05	.41	−.03	.51	.13	H/N	.75	.68
44. Easily distracted	−.06	.23	.09	.46	.30	H/N	.64	.59
13. Impulsive	.18	.08	.06	.44	.32	H/N	.73	.76
31. Disrupts group	.25	.05	.01	.44	.34	H/N	.77	.78
56. Ignores directions	.30	.37	−.12	.41	.01	H/N	.74	.71
7. Boisterous	.35	−.13	.11	.37	.31	H/N	.75	.77
9. Talks excessively	.01	−.06	.06	.16	.74	IS	.67	.61
33. Talks to self	.02	.07	.23	.18	.50	IS	.82	.78
43. Does Not communicate ^a	.06	.44	.18	.22	−.48^a	L/SW	.54	.54
22. Repetitive speech	−.02	.01	.32	.13	.46	IS	.82	.86
21. Disturbs others ^a	.33	−.02	.05	.36	.39^a	H/N	.78	.76
EFA subscale correlations								
Irritability	1.0							
Lethargy/social withdrawal	.33	1.0						
Stereotypic behavior	.34	.38	1.0					
Hyperactivity/noncompliance	.50	.18	.31	1.0				
Inappropriate speech	.31	.19	.09	.17	1.0			
Original model CFA subscale correlations (n = 1796 ^b) (calibration below the diagonal and validation above)								
	I	L/SW	S	H/N	IS			
Irritability	1.0	.47	.45	.76	.52			
Lethargy/social withdrawal	.47	1.0	.58	.51	.36			
Stereotypic behavior	.46	.59	1.0	.49	.41			
Hyperactivity/noncompliance	.76	.54	.53	1.0	.54			
Inappropriate speech	.49	.36	.45	.54	1.0			
Original model summed subscale score correlations (n = 1796 ^b)								
Irritability	1.0							
Lethargy/social withdrawal	.45	1.0						
Stereotypic behavior	.43	.54	1.0					
Hyperactivity/noncompliance	.73	.47	.47	1.0				
Inappropriate speech	.48	.31	.37	.49	1.0			

The largest factor loading in the EFA is highlighted in bold. All factor loadings in the CFA are significant at the $p \leq .05$ level
I Irritability, *L/SW* lethargy/social withdrawal, *S* stereotypic behavior, *H/N* hyperactivity/noncompliance, *IS* inappropriate speech

^a Indicates item with highest EFA loading on a new factor

^b After listwise deletion of missing data (maximum sample used for external correlations used in Tables 4 and 5)

Table 3 Model fit comparisons for factor analyses

Model	SB- χ^2	df	SB- χ^2/df	RMSEA [90 % CI]	SMR
Calibration sample (n = 1,130)					
Aman et al. (1985)	14,921.08	1,585	9.4	.086 [.085–.088]	.10
Brown et al. (2002)	26,406.00	1,589	16.6	.12 [.12–.12]	.12
Brinkley et al. (2007) 4-factor	13,985.18	1,589	8.8	.083 [.082–.084]	.10
Brinkley et al. (2007) 5-factor	13,240.26	1,585	8.3	.081 [.079–.082]	.10
Sansone et al. (2012)	12,061.85	1,415	8.5	.082 [.080–.083]	.09
Aman et al. (1985) Younger than age 6 (n = 556)	7,373.79	1,585	–	.081 [.079–.083]	.11
Aman et al. (1985) Age 6 and older (n = 574)	8,238.02	1,585	–	.086 [.084–.087]	.10
Aman et al. (1985) IQ at or below 70 (n = 527)	7,029.04	1,585	–	.081 [.079–.083]	.10
Aman et al. (1985) IQ above 70 (n = 603)	9,711.74	1,585	–	.092 [.091–.094]	.11
Aman et al. (1985) ADOS-CS below 7 (n = 631)	8,789.19	1,585	–	.085 [.083–.087]	.10
Aman et al. (1985) ADOS-CS at or above 7 (n = 499)	7,501.83	1,585	–	.087 [.085–.089]	.11
Validation sample (n = 763)					
Aman et al. (1985)	10,367.33	1,585	6.5	.085 [.084–.087]	.10
1-Factor model	30,081.92	1,595	18.9	.13 [.12–.13]	.14

Speech. The VABS-C associations with Lethargy/Social Withdrawal and with Stereotypic Behavior and the VABS-II Communication domain with Lethargy/Social Withdrawal were moderate in size. The remaining correlations were small.

Autism spectrum disorder severity, as measured by the ADOS, was largely unrelated to the ABC. Some children were administered more than one module, and thus their scores are included separately when reporting within a module. Across modules, the higher CS was used. Increased ASD severity was related to increased Lethargy/Social Withdrawal and increased Stereotypic Behavior. These were also small but statistically-significant effects. Note that the ADOS-CS scores were intended to be comparable across modules (Gotham et al. 2009). However, in this sample, the significant relationship between the CS and the two ABC subscales was primarily driven by children who were administered Module 1. This module is appropriate for children lacking phrase speech. There was no relationship between any ABC subscales and the CS for Modules 2 or 3.

Last, the CBCL was completed for most children in the sample. Not surprisingly, all ABC subscales were significantly related to all CBCL empirically derived subscales, DSM-oriented subscales, and composite scores on both forms. Most of these relationships were small to moderate in size. However, on the younger CBCL form, there was a large relationship between ABC Irritability on the one hand and Emotionally Reactive, Attention Problems, Aggressive Behavior, ODD Problems, Externalizing Problems, and

Total Problems on the other. For ABC Hyperactivity/Noncompliance, there was a large correlation with Attention Problems, ADHD Problems, Externalizing Problems, and Total Problems. ABC Lethargy/Social Withdrawal had a strong relationship with Withdrawn, PDD Problems, and Internalizing Problems.

On the older CBCL form, as with the younger form, there were primarily small-to-moderate correlations. However, ABC Irritability had a large correlation with CBCL Aggressive Behavior, ODD Problems, Conduct Problems, Externalizing Problems, and Total Problems. For ABC Hyperactivity/Noncompliance, there was a large correlation with CBCL Attention Problems, Aggressive Behavior, ADHD Problems, ODD Problems, Conduct Problems, Externalizing Problems, and Total Problems. Lethargy/Social Withdrawal had only one large correlation, with the CBCL Withdrawn/Depressed subscale. No correlations were larger than 0.41 between the Stereotypic Behavior or Inappropriate Speech subscales and any subscale on either CBCL form.

Finally, we present normative data for these children, who were usually rated by their parents, but occasionally by other family caregivers. We used ANOVA to analyze for the effect of sex, age (≤ 6 , 6 to ≤ 12 ; and > 12 years), and IQ (≤ 70 ; > 70), including all 2- and 3-way interactions between demographic variables. Bonferroni-Sidak post hoc tests were used to determine significant age differences. Table 5 presents the normative data by age, IQ, and lumped across age and IQ. No sex-related effects were significant so norms are not presented for boys and girls

Table 4 Bivariate relationships between ABC subscales and demographic/clinical variables

	n	Irritability	Lethargy/social withdrawal	Stereotypic behavior	Hyperactivity	Inappropriate speech
Age (in years)	1,796	-.13**	-.00	-.04	-.16**	.05
Sex	1,796	.03	.04	-.02	-.03	-.01
Dichotomous IQ	1,796	-.03	-.12**	-.19**	-.05	-.09**
SB5 FSIQ	239	.01	-.15	-.26**	-.01	-.13
SB5 NVIQ	239	-.02	-.13	-.28**	-.01	-.13
SB5 VIQ	239	.03	-.15	-.24**	-.01	-.11
Vineland composite	1,669	-.21**	-.33**	-.31**	-.22**	-.05
Vineland communication	1,688	-.18**	-.30**	-.27**	-.21**	.01
Vineland daily living	1,686	-.17**	-.25**	-.28**	-.19**	-.05
Vineland socialization	1,691	-.23**	-.26**	-.28**	-.23**	-.07*
Vineland motor	1,268	-.17**	-.21*	-.25**	-.15**	-.04
ADOS CS ^a	1,796	-.05	.11**	.10**	.02	.04
M1 CS ^a	759	.00	.22**	.16**	.10*	.01
M2 CS ^a	436	-.10	.03	.09	-.03	.08
M3 CS ^a	609	-.06	.03	.04	-.03	.04
CBCL ages 1.5–5 years (<i>t</i> scores)						
Emotionally reactive	954	.57**	.35**	.35**	.43**	.39**
Anxious/depressed	954	.46**	.35**	.27**	.33**	.31**
Somatic complaints	954	.36**	.28**	.26**	.30**	.29**
Withdrawn	954	.26**	.61**	.35**	.34**	.18**
Sleep problems	954	.40**	.31**	.29**	.37**	.28**
Attention problems	954	.51**	.26**	.26**	.63**	.26**
Aggressive behavior	954	.68**	.25**	.25**	.57**	.37**
Affective problems	954	.48**	.45**	.36**	.43**	.30**
Anxiety	954	.43**	.32**	.22**	.33**	.31**
PDD problems	954	.40**	.56**	.40**	.40**	.28**
ADHD problems	954	.41**	.33**	.32**	.61**	.26**
ODD problems	954	.59**	.21**	.19**	.49**	.33**
Internalizing problems	954	.53**	.51**	.40**	.47**	.36**
Externalizing problems	954	.70**	.31**	.30**	.66**	.37**
Total problems	954	.67**	.48**	.41**	.63**	.40**
CBCL ages 6–18 years (<i>t</i> scores)						
Anxious/depressed	801	.31**	.27**	.16**	.22**	.18**
Withdrawn/depressed	801	.29**	.55**	.25**	.21**	.15**
Somatic complaints	801	.23**	.28**	.19**	.17**	.13**
Social problems	801	.33**	.31**	.18**	.29**	.22**
Thought problems	801	.37**	.36**	.37**	.38**	.30**
Attention problems	801	.36**	.20**	.22**	.56**	.35**
Rule-breaking behavior	801	.42**	.18**	.15**	.43**	.20**
Aggressive behavior	801	.65**	.28**	.20**	.56**	.29**
Affective problems	801	.41**	.39**	.24**	.31**	.19**
Anxiety	801	.29**	.31**	.22**	.25**	.22**
Somatic problems	801	.20**	.23**	.18**	.16**	.11*
ADHD problems	801	.33**	.41**	.32**	.51**	.31**
ODD problems	801	.60**	.27**	.16**	.52**	.25**
Conduct problems	801	.52**	.22**	.16**	.51**	.21**
Internalizing problems	801	.36**	.43**	.25**	.28**	.21**

Table 4 continued

	n	Irritability	Lethargy/social withdrawal	Stereotypic behavior	Hyperactivity	Inappropriate speech
Externalizing problems	801	.64**	.28**	.20**	.58**	.30**
Total problems	801	.53**	.42**	.31**	.52**	.33**

Composite = Adaptive Behavior Composite; ADHD = Attention Deficit/Hyperactivity Disorder; ADOS = Autism Diagnostic Observation Schedule; CBCL = Child Behavior Checklist; CS = Calibrated Severity Score; FSIQ = Full Scale IQ; M[#] = ADOS Module number; NVIQ = Nonverbal IQ; ODD = Oppositional Defiant Disorder; PDD = Pervasive Developmental Disorder; VIQ = Verbal IQ

* $p < .01$; ** $p < .001$

^a Spearman correlations were used with the ADOS-CS scores, as they are ordinal rankings

separately. All of the ANOVAs predicting ABC subscales from demographic variables were significant. Nevertheless, there was substantial overlap between the means for each demographic factor, and the effect sizes were very small. For Irritability, there was a main effect for age ($F [2,1,784] = 3.80, p = .022, \omega^2 = .001$). For Lethargy/Social Withdrawal, there was a main effect for age ($F [2,1,784] = 3.70, p = .025, \omega^2 = .001$) and IQ ($F [1,1,784] = 28.10, p < .001, \omega^2 = .007$). There was only a main effect of IQ for Stereotypic Behavior ($F [1,1,784] = 6.20, p = .013, \omega^2 = .001$). For Hyperactivity/Noncompliance, there was a significant main effect of age ($F [2,1,784] = 11.23, p < .001, \omega^2 = .003$). Last, there was a significant interaction between IQ and age for Inappropriate Speech ($F [2,1,784] = 12.90, p < .001, \omega^2 = .005$). Although several post hoc procedures for interactions have been proposed, these are contentious and there is no best way to determine which means are significantly different from each other. Nevertheless, a qualitative interpretation of the interaction shows that whereas scores increased at older age ranges for those with ID, scores decreased at older ages for those with $IQ > 70$.

Discussion

The factor structure of the ABC was robust in this large, heterogeneous sample of children with ASD. The EFA supported a 5-factor solution, which was broadly consistent with the original factor structure (Aman et al. 1985). The original factor structure evidenced marginal fit in CFA analyses, as did other potential structures proposed in the literature. An RMSEA of about .08 is consistent with previous investigations of the ABC (Brinkley et al. 2007; Brown et al. 2002). A lower RMSEA was reported by Sansone et al. (2012), but methodological and sample (ASD vs. Fragile-X) differences prevent direct comparison. Their 6-factor model, when fit in this sample with DWLS, also exhibited marginal fit.

High factor loadings and less-than-optimal model fit may seem like a contradiction, but factor loadings and

model fit are two separate entities. A model could be associated with high factor loadings and still have misfit due to unmodeled covariance. The less-than-optimal fit of any structure for the ABC was unsurprising upon examination of specific items. Several sets of item pairs or triplets evidence a high degree of residual covariance. Allowing correlated residuals or more complex factor structures (e.g., creating bi-factor structures where an orthogonal factor is created for items with local dependence and allowing these items to load on the main factor and the orthogonal one) was considered and could have improved model fit, but this would not be a practical or parsimonious solution. The original 5-factor solution also had appropriate internal consistency for all subscales. As such, fit for the ABC was acceptable from a theoretical and practical perspective.

Nevertheless, several differences have consistently emerged in previous EFAs of the ABC and should be considered. First, various items on the Irritability and Hyperactivity/Noncompliance subscales have periodically moved to the alternative factor, but the nature of the crossover has been idiosyncratic to the study. In the Calibration sample EFA, this occurred for two items (#18 and 24), but there continued to be significant cross-loadings on the Hyperactivity/Noncompliance factor. These differences may be due to sample artifacts and do not suggest a need for scale revisions. Second, a factor for Inappropriate Speech has not always appeared. It appeared in this 5-factor EFA and had strong factor loadings in subsequent CFA analyses. The third difference which has occasionally emerged is the presence of a 3-item SIB factor. It emerged in the 6-factor EFA in this sample. When present, SIB is often highly clinically salient. Because of this clinical significance, some instruments (e.g., the ADOS) examine SIB but do not include it in the diagnostic algorithm (Gotham et al. 2007). Separating the SIB items out to a unique factor, though, is not supported by the 5-factor EFA results and when fit as a 6-factor CFA (available from the corresponding author), does not significantly improve model fit. Last, Sansone et al. (2012) found that the Lethargy

Table 5 ABC subscale means and standard deviations based on age group and IQ (≤ 70 ; > 70)

Age range	Sample Size	Irritability			Lethargy/social withdrawal			Stereotypic behavior			Hyperactivity/noncompliance			Inappropriate speech		
		Full	IQ ≤ 70	IQ > 70	Full	IQ ≤ 70	IQ > 70	Full	IQ ≤ 70	IQ > 70	Full	IQ ≤ 70	IQ > 70	Full	IQ ≤ 70	IQ > 70
≤ 6 years	1,036	558	478		10.2 (8.0) ^{a,b}	11.3 (8.0)	8.9 (7.7)	5.3 (4.9)	6.1 (4.9)	4.4 (4.6)	20.1 (11.7) ^a	20.4 (11.5)	19.6 (11.9)	3.6 (3.2)	3.0 (3.0)	4.3 (3.2)
6–12 years	604	213	391		9.4 (7.1) ^a	10.2 (7.6)	8.9 (6.9)	4.6 (4.2)	5.6 (4.5)	4.1 (4.0)	17.3 (10.9) ^b	17.7 (11.1)	17.1 (10.8)	3.9 (2.9)	4.2 (3.1)	3.7 (2.8)
> 12 years	156	62	94		11.2 (8.4) ^b	10.8 (8.6)	11.4 (8.3)	4.7 (4.5)	6.1 (5.0)	3.8 (3.9)	14.8 (10.0) ^b	15.2 (10.1)	14.5 (10.0)	3.6 (3.0)	4.3 (3.2)	3.1 (2.8)
Full sample	1,796	833	963		10.0 (7.8)	11.0 (8.0) ^f	9.1 (7.5) ^g	5.0 (4.6)	6.0 (4.8) ^h	4.2 (4.3) ⁱ	18.7 (11.4)	19.3 (11.4)	18.1 (11.4)	3.7 (3.1)	3.4 (3.1)	3.9 (3.0)

Significant differences are noted by subscripts. Subscripted letters indicated significant differences by age and are read column-wise. Subscripted symbols indicate significant differences by IQ and are read row-wise within a subscale. Different letters or symbols indicate significant mean differences, whereas equivalent letters or symbols indicate no significant difference

and Social Withdrawal items separated into two factors when ratings from individuals with Fragile-X syndrome were analyzed. This did *not* occur in our EFA results and fitting their model was not clearly superior to the original factor structure. There was no support for splintering Lethargy/Social Withdrawal in this sample, and issues remain as to whether an alternative scoring method should be used when assessing people with Fragile-X or other syndromes (Aman 2012b).

The ABC exhibited appropriate convergent and divergent validity and external correlations with IQ, adaptive behavior, ASD severity and the CBCL. Age appeared to be one of the more interesting correlates. Irritability and Hyperactivity decreased with age. Whereas we investigated age-related changes cross-sectionally, similar results on the ABC have been reported longitudinally with another ASD sample (Anderson et al. 2011). Reductions in ADHD symptoms are also known to occur in typically developing children as they age (Aman and Werry 1984). In this sample, we failed to find a significant relationship between age and Lethargy/Social Withdrawal. In contrast, Anderson et al. (2011) found that Social Withdrawal scores significantly increased during adolescence for young people with ASD diagnoses.

The participants' sex was unrelated to any ABC subscale scores. Among typically-developing children, there is a strong effect for boys to display more physical aggression and meet criteria for ADHD than girls (Card et al. 2008; Ramtekkar et al. 2010). If a similar effect were observed on the ABC, Irritability and Hyperactivity should be elevated for boys, but this was not observed. This is consistent with other research on children with ASD and other developmental disabilities, where sex differences that are expected in typically-developing children for behavior problems and comorbid diagnoses are greatly reduced (e.g. Einfeld et al. 2010; Worley and Matson 2011). Thus, female sex does not seem to be protective among youth with ASD to the extent that it is in neurotypical children.

We also examined the relationship between IQ, adaptive behavior, and ASD severity with the ABC factors. IQ was largely unrelated to all ABC subscales except Stereotypic Behavior, where lower scores (full scale, verbal, and non-verbal IQ) were related to a small increase in Stereotypic Behavior. When IQ scores were dichotomized at 70, we found that those with ID had higher Lethargy/Social Withdrawal and Stereotypic Behavior scores. However, these differences were small. ASD severity on the ADOS-CS was also largely unrelated to ABC subscale scores, showing the independence of constructs. This was somewhat unexpected as the ABC has been moderately correlated with other measures of ASD severity, such as the social domain and total score from the *Autism Diagnostic Interview—Revised* (Lecavalier et al. 2006). There were

very small but significant increases in Lethargy/Social Withdrawal and Stereotypic Behavior with increased ASD severity across modules, and this was largely driven by children administered Module 1 of the ADOS. Within Modules 2 and 3, no relationship emerged between ASD severity and any ABC subscale scores. Individuals administered Module 1 lack phrasal speech. This suggests that they are either younger or minimally verbal (and thus more significantly delayed intellectually).

Adaptive behavior, especially the Communication domain, was more associated than IQ with ABC scores. Lower adaptive behaviors (both the VABS-C and all domain scores) were related to increased Irritability, Hyperactivity, Lethargy/Social Withdrawal, and Stereotypic Behavior. Most of these effects were small, but the VABS-C had a medium association with ABC Stereotypic Behavior, and Communication domain had a medium association with increased Lethargy/Social Withdrawal. This suggests that lower communication abilities may mediate the relationship between ASD severity and Lethargy/Social Withdrawal and that individuals with poor communication skills are more withdrawn on average.

Good convergent validity was demonstrated between numerous ABC and CBCL subscales. For the *younger version* of the CBCL (relevant to ages 1.5–5 years), high correspondence was observed between ABC Irritability and the following CBCL subscales: (a) Emotionally Reactive, $r = .57$; (b) Anxious/Depressed, $r = .46$; (c) Sleep Problems, $r = .40$; (d) Aggressive Behavior, $r = .68$; (e) Affective Problems, $r = .48$; (f) ODD Problems, $r = .59$; (g) Internalizing Problems, $r = .53$; and (h) Externalizing Problems, $r = .70$. All of the associations with CBCL subscales characterizing acting-out problems are easy to reconcile; however, the association between ABC Irritability and CBCL Internalizing is not as intuitively compelling. ABC Lethargy/Social Withdrawal had robust correlations with CBCL Withdrawn ($r = .61$) and with PDD Problems ($r = .56$), suggesting that Lethargy/Social Withdrawal assesses important elements of social impairment as argued by others (see Scahill et al. 2012). ABC Stereotypic Behavior was only associated with CBCL PDD Problems ($r = .40$) and Internalizing Problems ($r = .40$) with any statistical strength. Hyperactivity/Noncompliance was most highly correlated with CBCL Attention Problems ($r = .63$), Aggressive Behavior ($r = .57$), ADHD Problems ($r = .61$), and Externalizing Problems ($r = .66$), which is highly consistent with expectation and literature showing an association between ADHD (e.g., ABC Hyperactivity) and disruptive behavior problems. Inappropriate Speech scores did not correspond with any CBCL subscale scores appreciably, suggesting that none of the latter measure the same construct

(i.e., the CBCL was largely constructed for neurotypical children).

For the school-age version of the CBCL (ages 6–18), the correlations for similar subscale pairs were salient. High correspondence was observed between ABC Irritability and the following CBCL subscales: (a) Aggressive Behavior, $r = .65$; (b) Affective Problems, $r = .41$; (c) ODD Problems, $r = .60$; (d) Conduct Problems, $r = .52$; and (e) Externalizing Problems, $r = .64$. Lethargy/Social Withdrawal was substantially correlated with Social Problems ($r = .31$) and Internalizing Problems ($r = .43$). Hyperactivity/Non-compliance corresponded with the following: (a) Attention Problems ($r = .56$); (b) Aggressive Behavior ($r = .56$); (c) ADHD Problems ($r = .51$); and (d) Externalizing problems ($r = .58$). ABC Stereotypic Behavior and Inappropriate speech were not strongly correlated with any CBCL subscales. Thus, for both the preschool and the school-age versions of the CBCL, there was good convergent and divergent validity, with correspondence between subscales presumed to assess similar constructs.

Finally, we presented normative data for children who are diagnosed as having ASD and who are rated by their parents and other family caregivers. Inspection of age effects did not reveal large differences, but we have endeavored to allow for comparisons with preschool children (≤ 6 years of age), school-age children (6 through 12 years), and adolescents (>12 years of age). We do not recommend using these norms for individuals above 18 years of age. The oldest age group in our sample (12–18 years) contained only 156 subjects, so the norms for this age group may be less reliable than for the other groups.

In conclusion, this study found that the original factor structure of the ABC is robust in children with ASD. We favored parsimony in interpretations, thus accepting a less-than-optimal fit for the original factor structure, over examining complex structures which could model additional residual covariance (e.g., modeling local dependency) but would be difficult to score by hand making them unfriendly for clinical use. The five subscales had acceptable to excellent internal consistency. Other studies have demonstrated that the ABC is sensitive to treatment effects in both medication and psychosocial treatment trials (Aman 2012a; Aman et al. 2004). In this large clinically representative sample of children with ASD, it also evidenced good convergent and divergent validity. Thus, this study supports the ongoing use of the ABC in both research and clinical settings in young people with ASD.

Acknowledgments The authors thank the Autism Treatment Network for access to the Aberrant Behavior Checklist data and related database. We are also grateful to the ATN Scientific Review Committee for comments on an earlier draft of this manuscript.

References

- Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Aman, M.G. (2012a). *Annotated bibliography on the Aberrant Behavior Checklist (ABC)*. (July, 2012, update). Unpublished manuscript, The Nisonger Center UCEDD, Ohio State University, Columbus, Ohio. http://psychmed.osu.edu/media/ABC_Annotated_Bibliography.pdf.
- Aman, M. G. (2012b). Editorial: Aberrant Behavior Checklist: Current identity and future developments. *Clinical and Experimental Pharmacology*, 2, e114. doi:10.4172/2161-1459.1000e114.
- Aman, M. G., McDougle, C. J., Scahill, L., & RUPP Autism Research Group. (2009). Medication and parent training in children with pervasive developmental disorders and serious behavioral problems: Results from a randomized clinical trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 1143–1154.
- Aman, M. G., Novotny, S., Samango-Sprouse, C., Lecavalier, L., Leonard, E., Gadow, K., et al. (2004). Outcome measures for clinical drug trials in autism. *CNS Spectrums*, 9, 36–47.
- Aman, M. G., & Singh, N. N. (1986). *Aberrant behavior checklist manual East Aurora*. New York: Slosson Educational Publications.
- Aman, M. G., Singh, N. N., Stewart, A. W., & Field, C. J. (1985). The Aberrant Behavior Checklist: A behavior rating scale for the assessment of treatment effects. *American Journal of Mental Deficiency*, 89, 485–491.
- Aman, M. G., & Werry, J. S. (1984). The Revised Behavior Problem Checklist in clinic attenders and non-attenders: Age and sex effects. *Journal of Clinical Child Psychology*, 13, 237–242.
- Anderson, D. K., Maye, M. P., & Lord, C. (2011). Changes in maladaptive behaviors from midchildhood to young adulthood in autism spectrum disorder. *American Journal on Intellectual and Developmental Disabilities*, 116, 381–397.
- Brinkley, J., Nations, L., Abramson, R. K., Hall, A., Wright, H. H., Gabriels, R., et al. (2007). Factor analysis of the Aberrant Behavior Checklist in individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 37, 1949–1959.
- Brown, E. C., Aman, M. G., & Havercamp, S. M. (2002). Factor analysis and norms for parent ratings on the Aberrant Behavior Checklist-Community for young people in special education. *Research in Developmental Disabilities*, 23, 45–60.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, 21, 230–258.
- Browne, M. W., Cudeck, R., Tatneni, K., & Mels, G. (2008). CEFA: Comprehensive exploratory factor analysis, Version 3.03 [Computer software and manual]. Retrieved from <http://faculty.psy.ohio-state.edu/browne/>.
- Budimirovic, D. B., & Kaufmann, W. E. (2011). What can we learn about autism from studying Fragile X syndrome? *Developmental Neuroscience*, 33, 379–394.
- Card, N. A., Stucky, B. D., Sawalani, G. M., & Little, T. D. (2008). Direct and indirect aggression during childhood and adolescence: A meta-analytic review of gender differences, intercorrelations, and relations to maladjustment. *Child Development*, 79, 1185–1229.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155–159.
- Einfeld, S. L., Gray, K. M., Ellis, L. A., Taffe, J., Emerson, E., Tonge, B. J., et al. (2010). Intellectual disability modifies gender effects on disruptive behaviors. *Journal of Mental Health Research in Intellectual Disabilities*, 3, 177–189.
- Freund, L. S., & Reiss, A. L. (1991). Rating problem behaviors in outpatients with mental retardation: Use of the Aberrant Behavior Checklist. *Research in Developmental Disabilities*, 12, 435–451.
- Gabriels, R. L., Cuccaro, M. L., Hill, D. E., Ivers, B. J., & Goldson, E. (2005). Repetitive behaviors in autism: Relationships with associated clinical features. *Research in Developmental Disabilities*, 26, 169–181.
- Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39, 693–705.
- Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The Autism Diagnostic Observation Schedule: Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*, 37, 613–627.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- Jöreskog, K., & Sörbom, D. (2007). LISREL 8.8 for Windows [Computer software]. Skokie, IL: Scientific Software International.
- Karabekiroglu, K., & Aman, M. G. (2009). Validity of the Aberrant Behavior Checklist in a clinical sample of toddlers. *Child Psychiatry and Human Development*, 40, 99–110.
- Lecavalier, L., Aman, M. G., Scahill, L., McDougle, C., McCracken, J. T., Vitiello, B., et al. (2006). Validity of the autism diagnostic interview-revised. *American Journal on Mental Retardation*, 111, 199–215.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30, 205–223.
- Marcus, R. N., Owen, R., Kamen, L., Manos, G., McQuade, R. D., Carson, W. H., et al. (2009). A placebo-controlled, fixed-dose study of aripiprazole in children and adolescents with irritability associated with autistic disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 1110–1119.
- Marshburn, E. C., & Aman, M. G. (1992). Factor validity and norms for the Aberrant Behavior Checklist in a community sample of children with mental retardation. *Journal of Autism and Developmental Disorders*, 22, 357–373.
- Norris, M., & Lecavalier, L. (2010). Evaluating the use of exploratory factor analysis in developmental disability psychology research. *Journal of Autism and Developmental Disorders*, 40, 8–20.
- Oliver, C., Arron, K., Sloneem, J., & Hall, S. (2008). Behavioural phenotype of Cornelia de Lange syndrome: Case-control study. *The British Journal of Psychiatry*, 193, 466–470.
- Owen, R., Sikich, L., Marcus, R. N., Corey-Lisle, P., Manos, G., McQuade, R. D., et al. (2009). Aripiprazole in the treatment of irritability in children and adolescents with autistic disorder. *Pediatrics*, 124, 1533–1540.
- Ramtekkar, U. P., Reiersen, A. M., Todorov, A. A., & Todd, R. D. (2010). Sex and age differences in attention-deficit/hyperactivity disorder symptoms and diagnoses: Implications for DSM-V and ICD-11. *Journal of the American Academy of Child and Adolescent Psychiatry*, 49, 217–228.
- Research Units on Pediatric Psychopharmacology (RUPP) Autism Network. (2002). Risperidone in children with autism and serious behavior problems. *New England Journal of Medicine*, 347, 314–321.

- Research Unit on Pediatric Psychopharmacology (RUPP) Autism Network. (2005a). Randomized, controlled, cross-over trial of methylphenidate in pervasive developmental disorders with hyperactivity. *Archives of General Psychiatry*, *62*, 1266–1274.
- Research Unit on Pediatric Psychopharmacology Autism Network. (2005b). Risperidone treatment of autistic disorder: Longer-term benefits and blinded discontinuation after 6 months. *American Journal of Psychiatry*, *162*, 1361–1369.
- Roid, G. (2003). *Stanford-Binet Intelligence Scales* (5th ed.). Itasca, IL: Riverside Publishers.
- Sansone, S. M., Widaman, K. F., Hall, S. S., Reiss, A. L., Lightbody, A., Kaufmann, W. E., et al. (2012). Psychometric study of the Aberrant Behavior Checklist in Fragile X syndrome and implications for targeted treatment. *Journal of Autism and Developmental Disorders*, *42*, 1377–1392.
- Satorra, A., & Bentler, P. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.
- Scahill, L., Hallett, V., Aman, M. G., McDougle, C. J., Arnold, L. E., McCracken, J. T., et al. (2012). Brief report: Social disability in autism spectrum disorder: Results from Research Units on Pediatric Psychopharmacology (RUPP) Autism Network trials. *Journal of Autism and Developmental Disorders*, *43*(3), 739–746.
- Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland Adaptive Behavior Scales* (2nd ed.). Survey Forms Manual. Circle Pines, MN: AGS Publishing.
- Tse, J., Strulovitch, J., Tagalakis, V., Meng, L., & Fombonne, E. (2007). Social skills training for adolescents with Asperger syndrome and high-functioning autism. *Journal of Autism and Developmental Disorders*, *37*, 1960–1968.
- Walley, R. M., & Donaldson, M. D. (2005). An investigation of executive function abilities in adults with Prader-Willi syndrome. *Journal of Intellectual Disability Research*, *49*, 613–625.
- Worley, J. A., & Matson, J. L. (2011). Psychiatric symptoms in children diagnosed with autism spectrum disorder: An examination of gender differences. *Research in Autism Spectrum Disorders*, *5*, 1086–1091.