Published in final edited form as:

Res Dev Disabil. 2013 May ; 34(5): . doi:10.1016/j.ridd.2013.02.024.

Validity and Reliability of the Behavior Problems Inventory, the Aberrant Behavior Checklist, and the Repetitive Behavior Scale-Revised among Infants and Toddlers at Risk for Developmental Disabilities: A Multi-Method Assessment Approach

Johannes Rojahn¹, Stephen R. Schroeder², Liliana Mayo-Ortega^{2,3}, Rosao Oyama-Ganiko³, Judith LeBlanc^{2,3}, Janet Marquis², and Elizabeth Berke¹

¹George Mason University

²University of Kansas

³Centro Ann Sullivan del Peru

Abstract

Individuals with developmental disabilities (DD) are at a heightened risk of developing aberrant behaviors during the course of their lives. Reliable and valid assessment of such behaviors is an important element in empirically verifying successful prevention and intervention. Only a few instruments exist which were developed for behavior problems in children of a young age. The purpose of the current longitudinal study was to examine the performance of three behavior-rating instruments for individuals with developmental disabilities that have been proven useful and psychometrically sound in older populations: the Behavior Problems Inventory (BPI-01), the Aberrant Behavior Checklist (ABC), the Repetitive Behavior Scale – Revised (RBS-R). Data were analyzed for 180 children at risk for a developmental disability in Lima, Peru who were involved in active behavioral treatment for behavior problems. Cronbach's alpha of all three measures showed variable internal consistency across the subscales of the three instruments. BPI-01 test-retest reliability for its three subscales over three times of measurement within approximately a 12 months period showed ICC coefficients ranging between .68 to .77 for frequency ratings and .65 to .80 for severity ratings. Using a multitrait-multimethod matrix approach for the constructs self-injurious behavior (SIB), stereotyped behavior, and aggressive/destructive behavior, we found high levels of convergent and discriminant validity across the three instruments. Sensitivity of the SIB and the Stereotyped Behavior subscales of the BPI-01, the RBS-R, and the ABC showed that while there was some overlap between the three instruments, each scale contributed unique information. This was especially true for stereotyped behavior. In addition, since each one of the three instruments contains subscales that are not represented by the others, it merits to consider using all three scales together.

Keywords

Aberrant Behavior Checklist; Behavior Problems Inventory; Repetitive Behavior Scale – Revised; psychometric properties; multitrait-multimethod matrix; neurodevelopmental disorders; at risk; infants; toddlers; young children

Individuals with developmental disabilities (DD) are at a heightened risk for developing chronic and severe aberrant behaviors during the course of their lives. Reliable and valid assessment of such behaviors is an important element in empirically verifying successful prevention and intervention.

In the past decade, there has been an increased interest in early identification and preventive intervention of behavior problems, such as aggression, self-injurious behavior (SIB), and stereotyped behavior, among infants and toddlers at risk for DD (Schroeder & Courtemanche, 2012). Yet, only a few instruments exist which were specifically developed and validated for behavior problems in children below age three, such as the recently published Baby and Infant Screen for Children with Autism Traits, Part 3 (BISCUIT) (Matson, Wilkins, Sevin, Knight, Bosjoli, & Sharp, 2009). There are, however, several well-validated behavior rating scales originally developed and validated in adult populations with DD which have also been used successfully among younger populations, e.g. the Aberrant Behavior Checklist (Karabokiroglu & Aman, 2009), The Behavior Problems Inventory-01 (Rojahn, Rowe, Sharber, Hastings, Matson, Didden, Kroes, & Dumont, 2012b), and the Repetitive Behavior Scale-Revised (Boyd, McDonough, & Bodfish, 2012).

The current study is a psychometric study involving the latter three behavior-rating scales with a population of infants and toddlers in Peru. In this study we analyze the internal consistency of all three instruments and the test-retest reliability of the BPI-01. Using a multitrait-multimethod approach (Campbell & Fiske, 1959) we also looked at the confirmatory and discriminant validity of the BPI-01 compared with the ABC and the RBS-R. Lastly, we examined the sensitivity and specificity of the BPI-01 compared to the RBS-R and the ABC.

Method

Participants and Recruitment

Participants were recruited in Lima, Peru by the Centro Ann Sullivan del Peru (CASP) via radio, television, and newspaper announcements in Lima and the rest of the country for a family-based early intervention program. Around 1,000 families called CASP in response to advertising. After discussing the project with the CASP Triage and Information Coordinator, 341 families agreed to visit CASP with their child for a screening interview. Children were chosen for clinical trials for their signs of SIB, aggression, and stereotyped behavior, based on parents' answers to the Parental Concerns Questionnaire (PCQ) (for details, see Mayo, Oyama, LeBlanc, Schroeder, Brady, et al., 2012). Participants had to meet several inclusion criteria gleaned from the research literature (Dawson, 1996; Dunlap, Strain, Fox, Carta, Conroy, et al. 2006; Rojahn, Schroeder, & Hoch, 2008) that would classify them as at risk for behavior disorders in developmental or intellectual disability (Mayo, et al., 2012).

Of these 341 families that responded, 262 met the inclusion criteria and 234 were administered the first round of in-depth interdisciplinary assessment measures, involving developmental pediatric exams, with consultations in neurology, genetics, and nutrition as indicated; vision, hearing, and dental screening, cognitive (Bayley, 2006), communication (Wetherby & Prizant, 2002) assessments and autism screening with the Child Autism Rating Scale (Schopler, Reichler, & Renner, 1988). They were then reassessed at six months and at 12 months.

In the end, due to failing to meet inclusion criteria or attrition, complete data on the Behavior Problems Inventory (Rojahn et al., 2001) for all three time points were obtained for 180 participants. This group consisted of 110 boys and 68 girls (2 missing data) with a mean age of 27.4 months ($SD = 10.1$, ranging from 4 to 48).

Measures of Behavior Problems

The key behavioral assessments of current interest included the Aberrant Behavior Checklist (ABC), the Behavior Problems Inventory (BPI-01), and the Repetitive Behavior Scale-Revised (RBS-R).

Aberrant Behavior Checklist (ABC)—The ABC (Aman & Singh, 2006; Aman, Singh, Stewart, & Field, 1985a, 1985b) is a 58 item informant based behavior rating scale that measures the severity of five subscales: Irritability (15 items), Lethargy (16 items), Stereotypy (7 items), Hyperactivity (16 items), and Excessive Speech (4 items). As the subscale labels indicate, this is a broad-band assessment instrument that captures a wide variety of behavior problems and includes generic items of SIB and aggressive behavior. It is one of the most widely used assessment instruments in the IDD literature and has been repeatedly shown to have excellent psychometric properties (e.g., Aman et al. 1985b, 1987; Bihm & Pointdexter, 1991; Paclawski, Bamburgh, Baglio, 1997). The ABC was used as a dependent variable in this study to capture a broad range of behavior problems.

Behavior Problems Inventory (BPI-01)—The BPI-01 (Rojahn et al., 2001; Rojahn et al., 2012a, 2012b) is a 52-item informant based behavior rating scale, which assesses the frequency and severity of problem behaviors in individuals with intellectual disabilities. The BPI-01 was the main dependent measure in the study. The items fall into three subscales: Self-Injurious Behavior (14 items), Stereotyped Behavior (25 items), and Aggressive/Destructive Behavior (11 items). Each item is scored on a three-point scale for severity (1 = slight, 2 = moderate, and 3 = severe) and a five-point scale for frequency (0 = never, 1 = Monthly, 2 = Weekly, 3 = Daily, and 4 = Hourly). Several previous studies were conducted that examined the psychometric properties of the BPI-01. Recently, Rojahn et al. (2012) conducted a comprehensive psychometric analysis on a data set of 1122 cases collated from several different sites all of whom had a BPI-01 total score > 0. Internal consistency (Cronbach's α) of the SIB subscale was .74 and .73 (for the frequency and the severity scales respectively), for the Stereotyped Behavior scale .89 and .90, and for the Aggressive/Destructive Behavior scale .92, and .90. Rojahn et al (2001) reported intra-class correlation (ICC) test-retest reliability for Self-Injurious Behavior of .71, for Stereotyped Behavior of .76, and Aggressive/Destructive Behavior of .64. Evidence for strong convergent and discriminant validity was also found, and the factor structure was endorsed by confirmatory factor analysis (Rojahn et al., 2012b).

Repetitive Behavior Scale – Revised (RBS-R)—The RBS-R (Bodfish, Symons, Parker, & Lewis (2000) is a 43-item informant based behavior rating scale that measures the severity of a variety of behaviors among individuals with DD/ID. It contains six subscales; Stereotyped Behavior (6 items), Self-Injurious Behavior (8 items), Compulsive Behavior (8 items), Ritualistic Behavior (6 items), Sameness Behavior (11 items), and Restricted Behavior (4 items) that are each scored on a four point Likert-type scale (0 = behavior does not occur, 1 = behavior occurs and is a mild problem, 2 = behavior occurs and is a moderate problem, 3 = behavior occurs and is a severe problem). Lam and Aman, (2007) validated a five-factor solution the RBS-R with 320 caregivers with the factors “Ritualistic/Sameness Behavior,” “Stereotypic Behavior,” “Self-injurious Behavior,” “Compulsive Behavior,” and “Restricted Interests.” They reported high measures of internal consistency and interrater reliability for this solution.

The SIB items of the RBS-R and the BPI-01 are contrasted in Table 1. The two instruments have three items that are more or less equivalent (biting, hair scratching, pulling), three RBS-R “hitting” items correspond with two BPI-01 “hitting” items, and three BPI-01 “inserting” items that correspond with one RBS-R “inserting” item. The BPI-01 has seven

SIB items that are not contained in the RBS-R, while the RBS-R has one item (“Skin picking”) that is not featured in the BPI-01. The stereotyped behavior items of the ABC, the RBS-R and the BPI-01 are compared in Table 2.

Procedure

The BPI-01 was administered to all 180 participants at all three time points, while the ABC and the RBS-R were completed for 97 participants at the second time only. The time between measurements was approximately six months. To avoid literacy and comprehension issues, ten well trained interviewers administered all assessment instruments by questioning the parents or caregivers and filled out all of the forms. An effort was made to keep the same interviewers with the same parents. For quality insurance the interviewers were periodically monitored.

The intervention, which is not the focus of the present paper, consisted of six three-hour bi-monthly teaching workshops at CASP and follow-up procedures with bimonthly parent teaching workshops given at CASP as well as monthly telephone follow-up with each family. This treatment lasted throughout the 12-month study period, and is being published elsewhere (Oyama, Mayo, Schroeder, & LeBlanc, 2012).

Results

Descriptive Statistics

Table 3 presents the descriptive statistics of the three behavior assessment instruments. The values for the BPI-01 were comparable to BPI-01 “clinical norms” of the BPI-01 for children from birth to ten years old published by Rojahn et al. (2012a), which were as follows SIB frequency mean = 5.84 ($SD = 5.29$), severity mean = 4.39 ($SD = 4.01$); Stereotyped Behavior frequency mean = 18.39 ($SD = 17.42$), severity mean = 10.5 ($SD = 11.2$); and Aggressive Destructive Behavior frequency mean = 6.81 ($SD = 7.9$), severity mean = 5.38 ($SD = 1.15$). This suggests that the participants’ aberrant behavior was consistent with other groups of peers with DD and severe behavioral concerns.

Internal consistency

A summary of the internal consistency statistics (Cronbach’s α) can be found in Table 4. Internal consistency of the BPI-01 total score across the three times of measurement ranged from .85 to .90 for the total score. For the subscale scores, SIB ranged from .42 to .69, for subscale Stereotyped Behavior, .82 to .86, and for subscale Aggressive/Destructive Behavior, .76 to .81. The ABC total score showed a Cronbach’s α of .96, and the subscales ranged from .81 to .92. The RSB-R total score had an internal consistency coefficient α of .89, while they ranged from .55 to .75 for its subscales. Hair, Anderson, Tatham, and Black (1998) suggest a Cronbach’s α of .70 shows good internal consistency.

Test-retest Reliability

Test-retest reliability for the BPI-01 was estimated in two different ways. First, Spearman’s ρ correlations were computed comparing rounds 1 vs. 2, 1 vs. 3, and 2 vs. 3. All correlations were statistically significant at the .01 level, ranging from .41 to .64 (mean = .53). The mean test-retest coefficient for the BPI-01 subscale SIB was .50 and .46 (for frequency and severity scores respectively), for Stereotyped Behavior it was .58 and .53, and for Aggressive/Destructive Behavior it was .57 and .59. Second, we computed intra-class correlations (ICC) using a two way mixed effects model with three ratings and each participant representing a level of the random person factor (McGraw & Wong, 1996). ICC coefficients ranged from .68 to .80 (see Table 5).

Consistency of BPI-01 Subscale Scores over Time

Three MANOVAs with repeated measures across assessment times 1, 2 and 3 were conducted for each of the three BPI-01 subscale scores with the frequency and the severity scores as the two dependent variables ($n = 180$). For the subscale SIB, the Multivariate Tests found a significant difference between the frequency and the severity scores, Wilks' $\Lambda = .602$, $F(1, 537) = 354.7$, $p = 0.000$. For the subscale Stereotyped Behavior, a significant difference between the frequency and the severity scores was found, Wilks' $\Lambda = .554$, $F(1, 537) = 432.8$, $p = 0.000$. And finally, for the subscale Aggression/Destruction, a significant difference was detected between the frequency and the severity scores (Wilks' $\Lambda = .677$, $F[1, 537] = 256.5$, $p = 0.000$). The interaction effect "factor (frequency vs. severity scores) x time" was not significant in any of the three MANOVAs. The results of the post-hoc analyses comparing the aggregated frequency and severity scores for the three BPI-01 subscale are presented in Table 6. Using Bonferroni adjusted p -values (for nine pair-wise iterations), subscale scores for SIB and Stereotyped Behavior decreased significantly between times 1 and 3 at the .01 and the .05 level respectively. No statistically significant changes were observed with Aggressive/Destructive Behavior subscale (see also Figure 1).

Convergent Validity

Convergent Validity was determined by computing correlations between subscales from the three different instruments that are theoretically analogous. Since none of the subscales of either one of the three instruments had normally distributed data non-parametric Spearman's ρ coefficients were computed (see Table 7). Skewed distributions in behavior problems are common among populations with DD. Numbers in bold boxes represent expected convergent validity for subscales that carry identical labels across instruments; regular-font boxes show expected convergent validity estimates for subscales that measure conceptually related constructs. For instance, self-injurious behavior was assessed by the identical label subscales of the RBS-R and the BPI-01 (bold box in Table 7). However, it was also assessed by a few items of the ABC Irritability subscale (regular-font box). The ABC Irritability subscale was seen as partly analogous with the SIB subscales of the RBS-R, the BPI-01 because three of its 15 items explicitly capture SIB. All of these subscales were significantly correlated with one another at the .01 level (see Table 6). Similarly, the RBS-R Stereotypy subscale, the BPI-01 Stereotyped Behavior subscale, and the ABC Stereotypy subscale were all significantly correlated. Aggressive behavior was measured by the BPI-01 Aggressive/Destructive Behavior subscale and in part by the ABC Irritability subscale which contained three items related aggressive behavior. They were all significantly correlated at the .01 level.

Discriminant Validity

Discriminant validity is reflected by weak or non-significant correlations between the scores of theoretically independent subscales or by negative correlations between subscales measuring incompatible constructs. In Table 7 underlined numbers point to relationship between unrelated constructs. Identifying evidence for discriminant validity is more difficult than finding evidence for confirmatory validity because various constructs of aberrant behavior and psychopathology are not entirely independent. Therefore, it was difficult to find subscales that measured theoretically independent or incompatible constructs among the three rating instrument that assess aberrant behavior in individuals with ID. We decided that the ABC subscale Excessive Speech should be independent of the BPI-01 subscales SIB and Aggressive/Destructive Behavior and the RBS-R subscale SIB. Indeed, we found that these subscales were not significantly correlated (see Table 4). We found similar evidence for discriminant validity between the ABC subscales Lethargy the BPI-01 subscales SIB and

Aggressive/Destructive Behavior. The ABC subscales Lethargy and the RBS-R subscales was significant, but low ($\rho = .24$).

However, several of the subscales that measured *prima facie* unrelated constructs were significantly correlated, which is not consistent with the notion of discriminant validity. As mentioned before, this may be accounted for by the fact that many of the constructs underlying those subscales within instruments are non-orthogonal.

Sensitivity and Specificity

Sensitivity and specificity was computed for the BPI-01 vis-à-vis the RBS-R, and the ABC. Only those subscales were chosen that had the same label. This means the SIB and the Stereotype Behavior subscales of the BPI-01, the RBS-R, and the ABC. Table 8 shows the sensitivity and specificity estimates of the BPI-01 subscales SIB and Stereotyped Behavior Compared to the Analogous Subscales of the RBS-R and the ABC. The BPI-01 was treated as “test” variable and the ABC and the RBS-R were treated as the “condition” variable. Table 8 indicates that the BPI-01 sensitivity scores were very high (ranging between .92 and .98) but that the specificity was low (ranging from .26 to .29). This means that the BPI-01 identified most cases that were also identified by the ABC and the RBS-R (specificity), but that the BPI-01 identified more individuals as having SIB and stereotyped behaviors that were missed by the ABC and the RSB-R (specificity).

Figure 2 presents the scatter plots of the subscale raw scores of the BPI-01, the ABC and the RSB-R. Linear regression estimates ranged from .17 to .29. Panel 1 contrasts the SIB subscales of the BPI-01 and the RBS-R. As shown in Table 8, there were many cases ($n = 28$) that received positive frequency scores on the BPI-01 SIB subscale (up to a score of 16) that received zero ratings on the equivalent subscale of the RBS-R. A closer look at the types of SIB topographies that were missed during the RBS-R interviews but that were identified by the BPI-01 shows that there were 16 cases with teeth grinding (frequency ratings ranging from 1 to 3 and severity ratings from 1 to 2), 14 cases of pica (frequency 1 to 4, severity 1 to 3), six cases of inserting fingers into body openings (frequency 1 to 3, severity 1 to 3), five cases of head hitting (frequency 2 to 3, severity 1 to 2), four cases of vomiting (frequency 2 to 3, severity 1 to 3), three cases of hair pulling (frequency 1 to 2, severity 1), three cases of extreme drinking (frequency 1 to 4, severity 1 to 3), two cases of biting (frequency 2, severity 1), two cases of pinching (frequency 1 to 2, severity 1), two cases of nail pulling (frequency 3 to 4, severity 2 to 3), one case with inserting objects (frequency 3, severity 1), and one case with air swallowing (frequency 3, severity 1). Conversely, there was only one case that received a score >0 on the RBS-R and a zero on the BPI-01 (self hitting and inserting fingers).

Panel 2 and 3 compare the number of cases the three instruments identified with regard to stereotyped behavior. While there were a good number of the same cases identified, the instruments also identified cases that were not identified by the other instruments.

Discussion

The BPI-01 descriptive statistics were comparable to BPI-01 “clinical norms” of the instrument for children from birth to ten years old published by Rojahn et al. (2012a). This result suggests that the occurrence of aberrant behaviors in this sample was consistent with other peer groups with DD in the U.S. and Europe, who are at risk for severe behavioral concerns.

As far as this sample of participants is concerned, internal consistency was excellent for all ABC subscales according to the interpretive guidelines proposed by Sparrow and Cicchetti

(1981). The RBS-R and the BPI-01 had lower internal validity scores that ranged between fair and excellent. The lower scores on the RBS-R are in part due to the small number of items per subscale, which inherently lowers Cronbach's alpha. Somewhat surprising was the steady decline of the BPI-01 SIB subscale reliability across the three assessment times. BPI-01 test-retest reliability was found to be decent, which was expected, considering the long time intervals and the putative influence of treatment.

The concurrent validity analysis shows that the three instruments measure a number of similar and dissimilar constructs. Primary convergent validity was examined by comparing subscales across instruments that carried the same label, namely SIB (on the BPI-01 and RBS-R) and Stereotypies (on the ABC, BPI-01, and RBS-R). Despite the different item content as shown in Tables 1 and 2, these subscales were significantly correlated across the instruments. Evidence for discriminant validity is harder to come by than for confirmatory validity because many aberrant behavior constructs are known not to be entirely independent from one another.

In summary, each one of the three instruments showed very decent psychometric properties to the extent that they were examined in this study, and there is some merit in using all three scales together, as their subscales differ and may add a more complete picture of other aberrant, but perhaps less severe, behavior problems that often accompany SIB, aggression, and stereotyped behavior, and may be precursors to them sometimes. Another impact of this study is that these three measures are now further validated and have shown versatility in assessment in infant populations at risk for developmental disabilities. Future projects should be done in order to test the validity of these measures in other age populations such as toddlers or children.

Acknowledgments

This study is part of an ongoing international collaborative research project between the University of Kansas, Centro Ann Sullivan del Peru, and George Mason University which was funded by the Fogarty International Research grant No. HD 060500.

The authors wish to gratefully acknowledge the pro bono volunteer work of many professionals in Peru.

References

- Aman, MG.; Singh, NN. Aberrant Behavior Checklist: Manual. East Aurora, NY: Slosson Educational Publications; 1986.
- Aman MG, Singh NN, Stewart AW, Field CJ. The Aberrant behavior Checklist: A behavior rating scale for the assessment of treatment effects. *American Journal on Mental Deficiency*. 1985a; 89:495–491.
- Aman MG, Singh NN, Stewart AW, Field CJ. Psychometric characteristics of the Aberrant Behavior Checklist. *American Journal on Mental Deficiency*. 1985b; 89:492–502.
- Bayley, N. Manual for the Bayley Scales of Infant Development. 3. Psychological Corporation; San Antonio, TX, USA: 2006.
- Bihm E, Pointdexter AR. Cross-validation of the factor structure of the Aberrant Behavior Checklist. *American Journal on Mental Retardation*. 1991; 96:209–11. [PubMed: 1930950]
- Bodfish JW, Symons FS, Parker DE, Lewis MH. Varieties of repetitive behavior in autism: Comparisons to mental retardation. *Journal of Autism and Developmental Disorders*. 2000; 30:237–243. [PubMed: 11055459]
- Boyd BA, McDonough SG, Bodfish JW. Evidence-based behavioral interventions for repetitive behaviors in autism. *Journal of Autism and Developmental Disorders*. 2012; 42:1236–1248. [PubMed: 21584849]

- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*. 1959; 56(2):81–105.10.1037/h0046016 [PubMed: 13634291]
- Cicchetti DV, Sparrow S. Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Retardation*. 1981; 86:127–137.
- Clinical Calculator. n.d. Retrieved from <http://vassarstatseveritynet/clin1.htm>
- Dawson G. Brief report: Neuropsychology of autism: A report on the state of the science. *Journal of Autism and Developmental Disorders*. 1996; 26:179–184. [PubMed: 8744481]
- Dunlap G, Strain PS, Fox L, Carta JJ, Conroy M, et al. Prevention and intervention with young children's challenging behavior: Perspectives regarding current knowledge. *Behavior Disorders*. 2006; 32:29–45.
- Hair, JF, Jr; Anderson, RE.; Tatham, RL.; Black, WC. *Multivariate data analysis*. 5. Prentice-Hall; Upper Saddle River, Nj: 1998.
- Lam KSL, Aman MG. The Repetitive Behavior Scale-Revised: Independent validation in individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2007; 37(5): 855–866.10.1007/s10803-006-0213-z [PubMed: 17048092]
- Karabokiroglu K, Aman MG. Validity of the ABC in a sample of toddlers. *Child Psychiatry and Human Development*. 2009; 40:99–110. [PubMed: 18600444]
- Matson JL, Wilkins J, Sharp B, Knight C, Sevin JA, Bosjoli JA. Sensitivity and specificity of the Baby and Infant Screen for Children with aUtism Traits (BISCUIT). Validity and cutoff scores for autism and PDD-NOS in toddlers. *Research in Autism Spectrum Disorders*. 2009; 3:924–930.
- Mayo L, Oyama R, leBlanc J, Schroeder SR, Brady N, Butler M, et al. Mass screening for severe behavior problems among infants and toddlers in Peru. *Journal of Mental Health Research in Intellectual Disabilities*. 2012; 5:246–259. [PubMed: 22962573]
- McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*. 1996; 1:30–46. (Correction, 1, 390).
- Oyama R, Mayo L, Schroeder SR, LeBlanc JM. Early distance intervention and follow-up for families of infants and toddlers at risk for developmental disabilities and severe behavior problems in Peru. 2012 Submitted for review.
- Paclawski TR, Matson JL, Bamburg JW, Baglio CS. A comparison of the Diagnostic Assessment for the Severely Handicapped-II (DASH-II) and the Aberrant Behavior Checklist (ABC). *Research in Developmental Disabilities*. 1997; 18:289–98. [PubMed: 9216028]
- Rojahn, J.; Schroeder, SR.; Hoch, TA. *Self-injurious behavior in intellectual disabilities*. New York: Elsevier; 2008.
- Rojahn J, Matson JL, Lott D, Esbensen AJ, Smalls Y. The *Behavior Problems Inventory: An Instrument for the assessment of self-injury, stereotyped behavior and aggression/destruction in individuals with developmental disabilities*. *Journal of Autism and Developmental Disorders*. 2001; 31:577–588. [PubMed: 11814269]
- Rojahn J, Rowe EW, Sharber AC, Hastings RP, Matson JL, Didden R, Kroes DBH, Dumont ELM. The Behavior Problems Inventory-Short Form (BPI-S) for Individuals with Intellectual Disabilities I: Development and Provisional Clinical Reference Data. *Journal of Intellectual Disability Research*. 2012b; 56:527–545. [PubMed: 22151184]
- Rojahn J, Rowe EW, Sharber AC, Hastings RP, Matson JL, Didden R, Kroes DBH, Dumont ELM. The Behavior Problems Inventory-Short Form (BPI-S) for Individuals with Intellectual Disabilities II: Reliability and Validity. *Journal of Intellectual Disability Research*. 2012a; 56:546–565. [PubMed: 22221562]
- Schopler, E.; Reichler, RJ.; Renner, BR. *The Child Autism Rating Scale*. Western Psychological Services Corporation; Los Angeles, CA: 1988.
- Wetherby, AM.; Prizant, BM. *Communication and Symbolic Behavior Scales Developmental Profile*. Baltimore, MD: Paul Brookes Publishing Co., Inc; 2002.

Figure 1.
Treatment sensitivity of the BPI-01 across three assessment times

Figure 2.

Scatter plots with linear regression lines between corresponding SIB and stereotypic behavior BPI-01 (frequency ratings), ABC, and RBS-R subscales (as observed at the 2nd point of measurement). Note: Dots in the scatter plots can represent one or more cases

Table 1

SIB Items (abbreviated)

RBS-R	BPI-01
Bites self	Self-biting
Hits with body part Hits against surface or object Hits with object	Hitting head with hand or body part Hitting body (not head)
Rubs or Scratches	Self-scratching
Inserts finger or object Inserting fingers/objects	Inserting objects Inserting fingers Inserting objects
Pulls hair or skin	Hair pulling
	Self-pinching Pica Pulling nails Air swallowing Extreme drinking Teeth grinding Vomiting and rumination
Skin picking	

Table 2

Stereotyped Behavior Items (abbreviated)

RBS-R	BPI-01	ABC
Body rocking	Rocking back and forth Repetitive body movements	Body movements Rocks body back and forth
Rolls, nods, turns head	Rolling head	Moves or rolls head
Locomotion (turns, jumps)	Bursts of running Pacing Bouncing around Spinning own body Whirling, turning around	
Sensory (gazes, sniffs, ...)	Sniffing objects Gazing at hands or objects Sniffing own body	
Hand/finger	Complex hand/finger movements Sustained finger movements Repetitive hand movements Clapping hands	
Object use (spins, twirls ...)	Twirling things Waving or shaking arms Spinning objects Manipulating objects repeatedly	
	Waving hands	Waves/shakes extremities
	Yelling/screaming Rubbing self Maintaining bizarre postures Grimacing	
		Stereotyped behavior Odd, bizarre behavior Repetitive movements

Table 3

Descriptive Statistics for the BPI-01, the ABC, and the RBS-R

	Time 1			Time 2			Time 3		
	mean	SD	n	mean	SD	n	mean	SD	n
BPI-01									
Self-Injurious Behavior (freq.)	6.8	6.1	180	5.8	4.9	180	5.1	4.1	180
Self-Injurious Behavior (sev.)	4.7	4.5	180	3.8	3.5	180	3.2	2.7	180
Stereotyped Behavior (freq.)	13.7	13.1	180	12.5	11.5	180	10.8	11.8	180
Stereotyped Behavior (sev.)	9.4	10.2	180	7.3	7.6	180	6.1	6.8	180
Agg/Dest (freq.)	7.7	7.4	180	7.2	6.6	180	7.0	7.1	180
Agg/Dest (sev.)	6.1	6.5	180	5.5	5.6	180	5.2	5.4	180
ABC									
Irritability				12.6	9.4	97			
Lethargy				8.4	8.7	97			
Stereotypy				3.6	4.6	97			
Hyperactivity				18.1	10.9	97			
Excessive Speech				1.9	2.8	97			
RBS-R									
Stereotypy				2.8	3.1	97			
Self-Injurious Behavior				2.0	2.5	97			
RBS				3.5	4.0	97			
Rituals				3.3	3.5	97			
RBS				6.0	5.0	97			
Restricted				2.0	2.1	97			

Table 4

Internal consistency(Cronbach's α)

Instruments and subscales	Items	Time 1	Time 2	Time 3
BPI-01				
Self-Injurious Behavior	14 freq.	.69	.55	.42
	sev.	.69	.58	.45
Stereotyped Behavior	24 freq.	.86	.82	.86
	sev.	.88	.84	.86
Aggressive/Destructive	11 freq.	.81	.76	.80
	sev.	.84	.80	.80
Total score	49 freq.	.90	.85	.86
	sev.	.92	.88	.87
ABC				
Irritability	15		.90	
Lethargy	16		.91	
Stereotypy	7		.87	
Hyperactivity	16		.92	
Excessive Speech	4		.81	
Total score	58		.96	
RBS-R				
Stereotyped Behavior	6		.75	
Self-Injurious Behavior	8		.55	
Compulsive Behavior	8		.72	
Ritualistic Behavior	6		.68	
Sameness Behavior	11		.74	
Restricted Behavior	4		.30	
Total Score	43		.89	

Note: Clinical significance of reliability levels: < .40 = poor agreement; .40 – .59 = fair; .60 – .74 = good; .75 – 1.00 = excellent (Cicchetti and Sparrow, 1981).

Table 5

BPI-01 Test-Retest Reliability (ICC Correlation Coefficients) Across Three Measurements

Subscales	Ratings	
	frequency	severity
Self-Injurious Behavior	.68	.65
Stereotyped Behavior	.70	.70
Aggressive/Destructive Behavior	.77	.80

Table 6

Post-hoc Multiple Comparisons (LSD) for the MANOVAs for the BPI-01 Subscales

	Time		Mean Difference	<i>p</i>
SIB	1	2	0.95	0.03549
	1	3	1.54	0.00063
	2	3	0.60	n.s.
Stereotypy	1	2	1.63	n.s.
	1	3	3.07	0.00389
	2	3	1.44	n.s.
Agg/Dest	1	2	0.51	n.s.
	1	3	0.76	n.s.
	2	3	0.25	n.s.

Bonferroni adjusted *p* values for nine pairwise comparisons: *p* = .05 is .0056; *p* = .01 is .0011

Table 7

Multitrait-Multimethod Matrix for the Convergent and Discriminant Validity: Spearman ρ Correlations Across Subscales at Time 2

		BPI-01 (sev.)			ABC					RBS-S					
		SIB	Stereotypies	Agg/Dest	Irritability	Lethargy	Stereotypy	Hyperactivity	Exc. Speech	Stereotypy	SIB	Compulsions	Rituals	Sameness	Restricted
BPI-01 (freq.)	SIB	.90**	.43**	.49**	.50**	.20*	.26**	.24**	.17*	.19*	.54**	.07	.22*	.12	.11
	Stereotypy	.46**	.92**	.29**	.39**	.38**	.50**	.33**	.27**	.41**	.36**	.26**	.09	.26**	.34**
	Agg/Dest	.53**	.32**	.94**	.51**	.07	-.16	.39**	.08	-.01	.09	.27**	.15	.21*	.16
ABC	Irritability	.49**	.36**	.52**		.67**	.39**	.77**	.47**	.27**	.38**	.36**	.33**	.35**	.31**
	Lethargy	.15	.32**	.10			.51**	.71**	.61**	.36**	.24**	.24**	.16	.25**	.23*
	Stereotypy	.17	.35**	-.12				.33**	.34**	.55**	.31**	.20*	.17*	.25**	.27**
	Hyperactivity	.22*	.27**	.39**					.52**	.21*	.07	.32**	.17	.25**	.20*
	Exc. Speech	.16	.26**	.14						.23*	.09	.27**	.03	.19*	.11
RBS-R	Stereotypy	.12	.36**	.02						.42**	.36**	.38**	.52**	.51**	
	SIB	.49**	.27**	.13							.17*	.27**	.22*	.16	
	Compulsions	.11	.28**	.28**								.52**	.64**	.44**	
	Rituals	.15	.08	.16									.67**	.53**	
	Sameness	.18*	.29**	.25**										.58**	

* $p < .05$;

** $p < .01$

Note: Numbers in bold boxes represent expected convergent validity for subscales with identical labels; regular boxes show expected convergent validity for subscales that measure related constructs; underlined number signal expected discriminant validity.

Table 8

Sensitivity and Specificity Estimates of the BPI-01 SIB and Stereotyped Behavior Subscale Scores (Treated as Test Variable) Compared to the Analogous Subscales of the RBS-R and the ABC (Treated as Condition Variable)

Subscale	Test	Condition		95% CI	
SIB	BPI-01	Absent ²	Present ¹	Totals	
		RBS-R			
		28	58	86	
		Sensitivity =.98		.90	1.00
		10	1	11	
Stereotypy	BPI-01	Specificity =.26		.14	.43
		38	59	97	
		RBS-R			
		19	65	84	
		Sensitivity =.92		.82	.97
Stereotypy	BPI-01	7	6	13	
		Specificity =.27		.12	.48
		26	71	97	
		ABC			
		25	59	84	
Stereotypy	BPI-01	Sensitivity =.95		.86	.99
		10	3	13	
		Specificity =.29		.15	.47
		35	62	97	

¹ score of 1 or higher;
² score of 0; CI = confidence interval