

The behavior flexibility rating scale-revised (BFRS-R): Factor analysis, internal consistency, inter-rater and intra-rater reliability, and convergent validity

Nienke Peters-Scheffer^a, Robert Didden^{a,*}, Vanessa A. Green^b,
Jeff Sigafos^b, Hubert Korzilius^a, Keenan Pituch^c,
Mark F. O'Reilly^c, Giulio Lancioni^d

^a *Department of Special Education, Radboud University Nijmegen, P.O. Box 9104,
6500 HE Nijmegen, The Netherlands*

^b *School of Education, University of Tasmania, Australia*

^c *Department of Special Education, University of Texas at Austin, USA*

^d *Department of Psychology, University of Bari, Italy*

Received 20 July 2007; accepted 30 July 2007

Abstract

We examined the psychometric properties of the behavior flexibility rating scale-revised (BFRS-R), a new scale intended for assessing behavioral flexibility in individuals with developmental disabilities. Seventy-six direct care staff members and 56 parents completed the BFRS-R for 70 children with developmental disabilities. Factor analysis revealed three factors (i.e., Flexibility towards objects, Flexibility towards the environment, and Flexibility towards persons) and results of several analyses indicated an excellent internal consistency and good intra-rater and inter-rater reliability of the total scale. These data suggest that the BFRS-R may provide a reliable rating of behavioral flexibility when used by direct-care staff and parents of children with developmental disabilities.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Behavior flexibility rating scale-revised (BFRS-R); Children with developmental disabilities; Factor analysis; Internal consistency; Inter-rater reliability; Intra-rater reliability; Convergent validity

* Corresponding author.

E-mail address: R.Didden@pwo.ru.nl (R. Didden).

Children with autism and related developmental disabilities are often said to insist on sameness and resist change (Rutter, 2005). The insistence on sameness and resistance to change may be viewed as a lack of behavioral flexibility (Wahlberg & Jordan, 2001). Thus many such children appear to have substantial difficulty in being flexible in their behavior and appear unable to cope with unpredictable changes in the environment (e.g., Lewis & Bodfish, 1998). Recent studies that have focused explicitly on the assessment of behavioral flexibility are lacking, as are contemporary instruments for measuring behavioral flexibility in specific situations in individuals with developmental disabilities.

Recently, Green et al. (2006) and Pituch et al. (2007) developed and evaluated the discriminant validity and factor structure of the behavior flexibility rating scale (BFRS). Based on the results of these two initial studies, a revised version of the scale (i.e., behavior flexibility rating scale-revised [BFRS-R]) was developed, which includes some wording and item changes (Green et al., 2007). Given these revisions, it would seem timely and important to assess the reliability and validity of the BFRS-R.

The BFRS and BFRS-R were developed as research tools for exploring issues related to behavioral flexibility and as clinical tools for identifying specific situations in which children with autism and related developmental disabilities insist on sameness or resist change. Systematic identification of these situations may be helpful in enabling clinicians to target interventions to scenarios that are problematic for the child.

In an initial study, Green et al. (2006) showed that the BFRS has good discriminant validity. In this study, 726 parents used the BFRS to assess severity of behavioral inflexibility in children with autism, Asperger syndrome, and Down syndrome. Results showed that individuals with Asperger syndrome showed more problems in relation to the insistence on sameness as measured by the BFRS, followed by individuals with autism, and Down syndrome, respectively. A subsequent factor analysis (Pituch et al., 2007) revealed that the BFRS appears to comprise two main factors: (a) Interruption/Disruption, and (b) Position/Location, with the remaining items comprising a third potential factor.

The above studies suggest the instrument has potential for research and possibly also for clinical assessment. While these data are promising, further research is required to evaluate various psychometric properties of this instrument for assessing problems in behavior flexibility. The present study, therefore, was designed to examine the internal consistency, intra- and inter-rater agreement and convergent validity of the revised version of the BFRS. Data of this type may assist in evaluating the clinical utility of this measure.

1. Method

1.1. Participants and respondents

Respondents were 76 direct care staff members working in the Dutch day care centers where the children received early intervention services. At the time of the study, primary ($n = 43$) and secondary ($n = 33$) respondents had cared for participants for a mean of 2.2 years (range: 0.6–11 years) and 2.3 years (range: 0.7–11 years), respectively. Also parents (of whom 7 were fathers) of 56 children participated in this study. The scale was completed for 70 children (of whom 41 were boys) with developmental disabilities who attended two Dutch day care centers for children with disabilities. Their mean age was 6.7 years (range: 2–17 years). In addition to severe to mild intellectual disability, the children had a range of other diagnoses, including: (a) autism/PDD ($n = 26$), (b) Angelman syndrome

($n = 4$), (c) ADHD ($n = 1$), (d) Down syndrome ($n = 3$), (e) physical impairment ($n = 11$), and (e) other ($n = 6$).

1.2. Instruments

The instruments included the BFRS-R and the Sameness Questionnaire which served for the assessment of problems in behavioral flexibility. The BFRS-R is a revised version of the BFRS (Green et al., 2007). Compared to the original BFRS, the revised version includes several wording changes on individual items, the removal of one item and the inclusion of two new items. The BFRS-R is a 16-item rating scale for measuring behavioral flexibility in children with autism and related developmental disabilities in specific situations. (A copy of the BFRS-R is included in the [Appendix A](#).) Items in the BFRS-R refer to a number of specific and unexpected events and changed routines that may prove problematic for the individual. The severity of each potentially problematic situation is rated on a three-point Likert scale, ranging from '0' ('Not a problem at all') to '2' ('The situation causes severe problems'). The scale is completed by proxies who have known the individual for at least 6 months.

The Sameness Questionnaire is a 28-item questionnaire for assessing a child's resistance to change (Prior & MacMillan, 1973). Examples of items are: 'Does he become very upset if interrupted in what he is doing?', and 'Does he object to visiting new places?' Each item is rated on a 3-point Likert scale, ranging from '0' ('Behavior is not shown by child') to '2' ('Behavior is present to a considerable degree'). In a sample of 32 children (who were between 3 and 11 years of age) with various types of developmental disabilities Prior and MacMillan (1973) found that children with autism had a significantly higher mean flexibility score than children without autism indicating that these children had more problems with behavioral flexibility than those without autism.

1.3. Procedure

The BFRS-R was provided to the respondents with a letter explaining the aim of the study. Each respondent independently completed a Dutch-language translation of the BFRS-R without the involvement of an interviewer. However, they were informed that they could contact the first author in case they had any questions on how to fill in the scale. Three weeks after the completion of the first assessment, each respondent was asked to complete a second BFRS-R as well as the Sameness Questionnaire.

2. Results

2.1. Exploratory factor analysis

An exploratory factor analysis to identify possible subscales was conducted using all 132 BFRS-Rs completed by parents and staff members at the first assessment. A principle axing analysis revealed three factors, which were labelled as (a) Flexibility towards objects, (b) Flexibility towards the environment, and (c) Flexibility towards persons. Due to the correlation between factors one and three (i.e., $r = .59$), and between factors two and three (i.e., $r = .46$), direct oblimin rotation was used. In the pattern matrix factor loadings needed to be .30 or higher,

Table 1
Factor loadings for each item

Item	Factor		
	1	2	3
1 A commonly used object is misplaced and cannot be found	.50		
2 A planned event is delayed or cancelled with little warning because of unforeseen circumstances	.67		
3 The person is required to move from their current location and go to another location	.38		
4 An object in the environment has been moved or repositioned from its usual location or position		.79	
5 The person wants something what is not available	.82		
6 An object or some materials that the person was using breaks or malfunctions	.65		
7 A usual routine is altered or changed, for example the parent takes a new route home from school	.63		
8 An unexpected interaction occurs with another person, for example a stranger tries to talk to the person			.67
9 The person becomes momentarily separated from his/her family or group			.57
10 Materials break, causing a premature end to an activity	.72		
11 Another person is doing something annoying, for example making noise		.54	
12 Objects or materials are not returned to their proper place at the end of an activity		.45	
13 A new object, item, or person has been added to the environment.		.42	
14 An activity is interrupted before the person was able to finish the task	.83		
15 A new activity is introduced into the person's routine		.60	
16 Another person tries to use favorite possessions of the person	.68		

and an item was retained on the factor on which it loaded most strongly. The percent of variance explained by the factors was 37.11, 3.16, and 7.54 for the factors 1, 2 and 3, respectively. Eigen values were 6.42 for factor 1, 1.04 for factor 2 and 1.76 for factor 3. Factor loadings are described in Table 1.

2.2. Internal consistency

Cronbach's α coefficient was calculated for internal consistency of the total scale using the data from all the respondents of the first assessment ($\alpha = .90$). Alpha values for two of the subscales (i.e., Factor 1 [Items 1–3, 5–7, 10, 14 and 16] and Factor 2 [Items 4, 11–13, and 15]) were .90 and .73, respectively. The alpha value of the third [Factor 3] subscale (Items 8 and 9) was .58. The alpha value for the total scale was excellent, suggesting that the total scale was homogeneous in content.

Mean item-total correlation for the total scale was .55 (S.D. = 0.13) with a range from .28 to .74. Mean item-total correlations for the subscales (1) Flexibility towards objects, (2) Flexibility towards the environment, and (3) Flexibility towards persons were .66 (S.D. = 0.07; range: .50–.75), .50 (S.D. = 0.04; range: .37–.61), and .41 (S.D. = 0.00; range: .41–.41), respectively. Intraclass correlation coefficients (calculated on the total sum of scores) for each item ranged from .34 to .79 ($M = .55$, S.D. = .11) and were .87, .65, and .54, for the three subscales, respectively (all correlations were statistically significant with $p < .01$). Intraclass correlation was considered excellent for 1 item, good for 3 items, fair for 11 items and poor for 1 item.

2.3. Intra-rater reliability

Intra-rater reliability was calculated for the BFRS-R, the three factors and the items for all respondents as well as for both parents and staff members. For each rater, percentage of exact and adjacent agreement between the first and second assessment was calculated (see Table 2). Mean percentage of exact agreement for the total scale was 74.64 (S.D. = 4.82; range: 68.6–88.6%) and mean percentage of adjacent agreement was 99.55 (S.D. = 0.7; range: 98.6–100%). For the staff members, mean percentage of exact agreement was 75.3 (S.D. = 6.9, range: 64.4–91.1%) and mean percentage of adjacent agreement was 100. Mean exact agreement and adjacent agreement for parents were 73.5 (S.D. = 8.3; range: 60–84%) and 99 (S.D. = 1.8; range: 96–100%), respectively.

Mean Cohen's kappa coefficient for exact agreement was .49 (S.D. = .10; range: .34–.80) for all respondents, .51 (S.D. = .14; range: .28–.84) for staff members, and .46 (S.D. = .18; range: .14–.72) for parents.

For each item, Spearman rank correlation coefficients were calculated (see Table 2). Mean Spearman correlation was .56 (S.D. = 0.12; range .34–.84) for all respondents, .58 (S.D. = 0.14; range .33–.91) for staff members and .51 (S.D. = 0.15; range .18–.77) for parents. Based on the criteria established by Cicchetti (1994), intra-rater agreement for the total scale was fair and ranged from poor to good for the items.

Total severity score was calculated as the sum of the severity ratings for the 16 items for each assessment. Spearman correlation for total severity score was .81 ($p < .01$) for all respondents, .82 ($p < .01$) for staff members and .87 ($p < .01$) for parents. Severity score for each factor was also calculated. Spearman correlations ($p < .01$) for Factors 1, 2 and 3 were .84, .67, and .55 for all respondents, .82, .74 and .49 for staff members and .92, .54, .67 for parents, respectively.

Table 2
Mean intra-rater reliability and internal consistency for staff members and parents

Item	Total				Staff members				Parents			
	EA	AA	Kappa	Spearman Rank	EA	AA	Kappa	Spearman rank	EA	AA	Kappa	Spearman rank
1	75.7	100	.53	.62	71.1	100	.47	.61	84.0	100	.64	.65
2	88.6	98.6	.80	.84	91.1	100	.84	.91	84.0	96.0	.72	.72
3	75.7	100	.53	.57	77.8	100	.55	.55	72.0	100	.59	.50
4	75.7	98.6	.50	.50	75.6	100	.50	.54	76.0	96.0	.50	.43
5	74.3	98.6	.52	.59	71.1	100	.47	.64	80.0	96.0	.59	.49
6	78.6	100	.56	.67	75.6	100	.51	.62	84.0	100	.67	.76
7	75.7	100	.53	.61	80.0	100	.61	.64	68.0	100	.41	.59
8	68.6	100	.40	.50	64.4	100	.32	.40	76.0	100	.55	.70
9	71.4	98.6	.47	.51	68.9	100	.39	.47	76.0	96.0	.57	.53
10	71.4	98.6	.45	.51	77.8	100	.59	.58	60.0	100	.18	.38
11	75.7	100	.56	.70	84.4	100	.70	.82	60.0	100	.32	.54
12	77.1	100	.48	.48	80.0	100	.54	.55	72.0	100	.38	.38
13	71.4	100	.34	.34	75.6	100	.49	.50	64.0	100	.14	.18
14	70.0	100	.41	.53	73.3	100	.49	.59	64.0	100	.23	.41
15	68.6	100	.38	.41	64.4	100	.28	.33	76.0	100	.52	.52
16	75.7	100	.43	.54	73.3	100	.44	.58	80.0	100	.36	.46

NB. EA, exact agreement; AA, adjacent agreement.

Table 3
Inter-rater reliability and intraclass correlation for staff members and parents

Item	Staff members					Staff member–parent				
	EA	AA	Kappa	Spearman rank	ICC	EA	AA	Kappa	Spearman rank	ICC
1	66.7	97.0	.41	.42	.43	68.2	95.5	.41	.50	.43
2	72.7	100	.51	.69	.63	63.6	100.0	.30	.39	.40
3	63.6	100	.26	.26	.27	77.3	95.5	.58	.53	.40
4	72.7	100	.42	.49	.55	68.2	95.5	.39	.36	.23
5	57.6	100	.22	.41	.44	63.6	100	.29	.48	.44
6	51.5	100	.07	.17	.21	59.7	100	.27	.48	.45
7	72.7	100	.48	.51	.57	72.7	100	.44	.45	.46
8	60.6	100	.25	.29	.30	59.1	100	.23	.35	.32
9	36.4	100	.24	.16	.14	59.1	100	.22	.30	.32
10	72.7	100	.52	.62	.63	68.2	100	.37	.49	.43
11	54.6	100	.12	.04	.06	59.1	95.5	.20	.21	.20
12	66.7	100	.05	.05	.04	72.7	95.5	.27	.27	.15
13	54.6	100	.06	.06	.08	59.1	100	.18	.20	.16
14	72.7	100	.49	.55	.58	77.3	95.5	.54	.44	.39
15	63.6	100	.29	.35	.36	50.0	95.5	.04	.04	.05
16	60.6	100	.24	.41	.43	77.3	100	.43	.56	.54

NB. EA, exact agreement; AA, adjacent agreement; ICC, intraclass correlation coefficient.

2.4. Inter-rater reliability

On each item, percentages of both exact and adjacent agreement were calculated between pairs of staff members and between pairs of a staff member and a parent (see Table 3). Between staff members, mean percentage of exact agreement was 62.5 (S.D. = 10.07; range: 36.4–72.7%) and mean percentage of adjacent agreement was 99.8 (S.D. = 0.76; range: 97–100%). Mean Cohen's kappa coefficient for exact agreement was .29 (S.D. = .17; range: .05–.52). For the staff member and the parent, mean percentage of exact agreement was 65.9 (S.D. = 8.13; range: 50–77.3%), and mean percentage of adjacent agreement was 98 (S.D. = 2.33; range: 95.5–100%). Mean Cohen's kappa coefficient for exact agreement was .32 (S.D. = .14; range: .04–.58).

Then, Spearman rank order correlation coefficients were calculated for each item (see Table 3). Mean Spearman correlation was .34 (S.D. = 0.20; range .04–.69) and .38 (S.D. = 0.14; range .04–.56) between staff members and between a staff member and a parent, respectively.

Finally, inter-rater agreement of item severity score was calculated using the intraclass correlation coefficient. Mean intraclass correlation was .35 (S.D. = 0.21; range: .04–.63) for staff members and .33 (S.D. = .14; range: .05–.54) for parents. For staff members, intraclass correlations were statistically significant at $p < .05$, except for Items 3, 6, 9, and 11–13. For parents, all items were significant at $p < .05$, except for Item 4, 8, 9, 11–13 and 15.

The intraclass correlation for total severity score between pairs of staff members was .58 ($p < .01$) and Spearman correlation for total severity score was .58 ($p < .01$). For the three factors, correlations were .67 ($p < .01$), .45 ($p < .01$), and .13 (ns.), respectively. Spearman correlations were .69 ($p < .01$), .49 ($p < .01$), and .10 (ns.), respectively. Between pairs of a staff member and a parent, intraclass correlation for total severity score was .55 ($p < .01$); Spearman correlation for total severity score was .64 ($p < .01$). Intraclass correlations for the three factors

were .62 ($p < .05$), .28 (ns.), and .45 ($p < .01$), respectively, with Spearman correlations of .45 ($p < .05$), .78 ($p < .01$), and .32 (ns.).

2.5. Convergent validity

The convergent validity was calculated by correlating total scores on the BFRS-R and the Sameness Questionnaire (SQ). A statistically significant (Pearson) correlation was obtained between the total scores of the BFRS-R and the SQ, $r = .51, p < .01$. Also, a significant correlation was found between the SQ-score and Flexibility towards objects (Factor 1), $r = .37, p < .05$, and between the SC-score and Flexibility towards the environment (Factor 2), $r = .56, p < .01$.

3. Discussion

The present study was the first to investigate internal consistency, intra-rater and inter-rater agreement, and convergent validity of a new scale (i.e., the BFRS-R) for the assessment of behavioral flexibility in children with various types of developmental disabilities. The total scale has an excellent internal consistency and consistency for individual items ranged from poor to excellent. Factor analysis revealed three factors (i.e., Flexibility towards objects, Flexibility towards the environment, and Flexibility towards persons), with internal consistency ranging from modest to excellent. Inter-rater and intra-rater reliability were assessed using exact and adjacent agreement. Adjacent agreement was excellent, whereas exact agreement was modest to good. Inter-rater reliability was good for the total scale. Finally, convergent validity was assessed using the Sameness Questionnaire of [Prior and MacMillan \(1973\)](#). Significant moderate correlations were obtained between the total Sameness Questionnaire and the total BFRS-R and between the SQ and Factors 1 and 2 of the BFRS-R.

The present data suggest that this scale has good potential for assessing situations related to behavioral flexibility in children with developmental disabilities. In clinical practice, the BFRS-R might be used to identify the type of situations and extent to which individuals show a resistance to change or an insistence on sameness. The BFRS-R is a reliable tool that may be helpful in enabling clinicians to target interventions for scenarios that are problematic for the child. The BFRS-R can be completed independently by the parent or staff member and is easier to administer than more comprehensive yet elaborate tools measuring related aspects of behavioral flexibility (such as for example, the Autism Diagnostic Interview-Revised; [Lord, Rutter, & Le Couteur, 1994](#)).

The results should be interpreted with caution given the fact that the ratings were provided by parents and direct care staff working in two Dutch day care centers for children with developmental disabilities. It is not clear if the BFRS-R would produce similarly reliable ratings when used by teachers or clinicians in other settings and with other samples of children with developmental disabilities. Clearly, further research is necessary to extend these preliminary findings.

Acknowledgements

We thank parents and direct care staff affiliated with the day care centers for their participation. This study was supported by a grant from Stichting De Driestroom (Elst, Netherlands) and completed while the second author was a visiting Endeavour Research Fellow at the University of Tasmania.

Appendix A. Behavior flexibility rating scale-revised

Demographic information

1. Name/gender of person: _____ Male or Female
2. This person is _____ years old
3. The person's main diagnosis is: (Check the one item that best describes the person's diagnosis)
 Autism Asperger syndrome Down syndrome Other _____ (Please specify)
4. The person's level of speech is: (Check the one item that best describes the person's level of speech development)
 Does not speak Speaks only a few single words/sentences Mainly nonfunctional speech/echolalia Fluent speech
5. The person's level of Intellectual Disability is: (Check the one item that best describes the person's level of Intellectual Disability)
 Superior (IQ above 115) Average (IQ 85-115) Below average (IQ 75-84) Mild (IQ 55-74) Moderate (IQ 35-54) Severe (IQ less than 34)
6. I am the person's: Parent Teacher Other _____ (Please specify) and have known this person for _____ years and _____ months

A.1. Background

The behavior flexibility rating scale-Revised (BFRS-R) seeks to assess to what extent various SITUATIONS AND CHANGES cause problems for the person. The BFRS-R can be completed by individuals who know the individual well enough to provide accurate information about the person's ability to be flexible and cope with change. Generally, anyone who has known and cared for the individual for at least 6 months could complete the BFRS-R.

A.2. Directions

Please rate the extent to which each of the following 16 situations is a problem for the individual by marking the best answer.

0—means that the situation is *not at all a problem* for the person. He or she copes easily with the situation.

1—means that the situation causes only *mild problems* and that these are only short-lived. The person might complain or fuss a little bit and for a short period of time (for 1–2 min). He/she might even tantrum mildly (e.g., stomp their feet, cry), but eventually the person accepts the situation and calms down and copes with the situation.

2—means that the situation causes *severe* problems. The situation may lead to a major tantrum. The tantrum might include aggression, screaming, and/or self-injury. The person never accepts the situation and things have to be returned to how they were before or the person has to be removed from the situation to calm down

To what extent is each of the following situations a problem for the person?		Severity of the problem		
		No	Mild	Severe
1	A commonly used object is misplaced and cannot be found	0	1	2
2	A planned event is delayed or cancelled with little warning because of unforeseen circumstances	0	1	2
3	The person is required to move from their current location and go to another location	0	1	2
4	An object in the environment has been moved or repositioned from its usual location or position	0	1	2
5	The person wants something what is not available	0	1	2
6	An object or some materials that the person was using breaks or malfunctions	0	1	2
7	A usual routine is altered or changed, for example the parent takes a new route home from school	0	1	2
8	An unexpected interaction occurs with another person, for example a stranger tries to talk to the person	0	1	2
9	The person becomes momentarily separated from his/her family or group	0	1	2
10	Materials run out, causing a premature end to an activity	0	1	2
11	Another person is doing something annoying, for example making noise	0	1	2
12	Objects or materials are not returned to their proper place at the end of an activity	0	1	2
13	A new object, item, or person has been added to the environment	0	1	2
14	An activity is interrupted before the person was able to finish the task	0	1	2
15	A new activity is introduced into the person's routine	0	1	2
16	Another person tries to use favorite possessions of the person	0	1	2

References

- Cicchetti, D. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6, 284–290.
- Green, V., Sigafoos, J., O'Reilly, M., Pituch, K., Didden, R., Lancioni, G., & Singh, N. (2007). Behavior flexibility in individuals with autism: Theory, assessment, and intervention. In L. Zhao (Ed.), *Autism research advances* (pp. 63–77). New York: Nova Science Publishers.
- Green, V., Sigafoos, J., Pituch, K., Itchon, J., O'Reilly, M., & Lancioni, G. (2006). Assessing behavioral flexibility in individuals with developmental disabilities. *Focus on Autism and Other Developmental Disabilities*, 21, 230–236.
- Lewis, M., & Bodfish, J. (1998). Repetitive behavior disorders in autism. *Mental Retardation and Developmental Disabilities Research Reviews*, 4, 80–89.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24, 659–685.
- Pituch, K. A., Green, V. A., Sigafoos, J., Itchon, J., O'Reilly, M., Lancioni, G., & Didden, R. (2007). Factor structure of the behavior flexibility rating scale. *Research in Autism Spectrum Disorders*, 1, 55–66.
- Prior, M., & MacMillan, (1973). Maintenance of sameness in children with Kanner's syndrome. *Journal of Autism and Childhood Schizophrenia*, 3, 154–167.
- Rutter, M. (2005). Aetiology of autism: Findings and questions. *Journal of Intellectual Disability Research*, 49, 231–238.
- Wahlberg, T., & Jordan, S. (2001). A case study in the dynamics of autism. In T. Wahlberg, F. Obiakor, S. Burkhardt, & A. Rotatori (Eds.), *Autism spectrum disorders: Educational and clinical interventions* (pp. 53–65). Oxford, UK: Elsevier Science.