Brief Report

Standardization of the Behavioral and Emotional Rating Scale: Factor Structure, Reliability, and Criterion Validity

Michael H. Epstein, EdD
Gail Ryser, PhD
Nils Pearson, PhD

Abstract

The present study reports on the standardization of the Behavioral and Emotional Rating Scale and examines its factor structure, reliability, and criterion validity. Data on a national sample of children without disabilities (n = 2,176) and children with emotional and behavioral disorders (n = 861) were collected. Analysis of the data from the first sample identified five factors: interpersonal strengths, family involvement, intrapersonal strength, school functioning, and affective development. The factors appeared to be highly stable and reliable (.79 to .99). No statistically significant age or gender differences were noted, although females were rated higher on each factor and the overall score. The second sample was rated significantly lower than the first across the factors and total score. The article discusses future research issues and practical implications.

Introduction

In the social sciences, a primary goal of researchers, policy makers, and direct service providers is the development of assessment instruments that afford reliable and valid information on human behavior. *Assessment* is a process of gathering information about an individual or group in order to make important decisions.[1] These decisions may include identifying children, placing children into specialized education and treatment programs, and evaluating the outcomes of such specialized efforts. Most typically, assessment instruments identify deficits or problems in an individual's functioning. Several assessment instruments are available to gather information for decision-making purposes. While these instruments differ with respect to assumptions, items, and data collection procedures, they focus on and are oriented toward identifying an individual's deficits. In the area of children's emotional and behavioral disorders (EBD), there are several instruments (eg, Child Behavior Checklist,[2] Child and Adolescent Functional Assessment Scale,[3] Revised Behavior Problem Checklist[4]) that possess adequate psychometric properties and are helpful in understanding the functioning of children.

Address correspondence to Michael H. Epstein, EdD, William Barkley Professor, Center for At-Risk Children's Services, 202 Barkley Center, University of Nebraska, Lincoln, NE 68582-0732; e-mail: mepstein1@unl.edu.
Gail Ryser, PhD, is Assistant Director of Research, PRO-ED.
Nils Pearson, PhD, is Director of Research, PRO-ED.

Journal of Behavioral Health Services & Research, 2002, 29(2), 208–216.

Overall, these instruments document the problems, pathologies, and deficits of the individuals being assessed. However, the field of assessment does not need to be restricted to the identification and measurement of deficits, problems, and pathologies. From a holistic perspective, information about an individual's strengths, assets, or competencies is equally as important as information about limitations. To this end, several test developers have begun to include items that assess the strengths of children in addition to their deficits. The Behavior Assessment System for Children (BASC),[5] for example, is a comprehensive approach to assessing behavior that includes forms for teachers, parents, and youths. The parent and teacher versions include both adaptive and problem behaviors. Also, the Strengths and Difficulties Questionnaire (SDQ),[6] a 25-item scale, is a brief screening instrument that includes positive and negative items that are divided into five domains of five items each. Other instruments have been developed to measure a specific aspect of child competence such as social skills. For example, the Social Skills Rating System (SSRS)[7] includes separate rating forms for teachers, parents, and children. Each instrument includes items that assess social skills. While the BASC, SDQ, and SSRS include a focus on adaptive behaviors or social skills, meet acceptable levels of test psychometrics, and are used in research and applied settings to assess child behavior or mental health status, they are not solely strength-based instruments nor were they developed specifically to assess strengths and competencies.

Within the past decade, several mental health, social service, and educational initiatives have advocated for alternatives to deficit-focused assessments. Specifically, the Child and Adolescent Service System Program (CASSP),[8] the *National Agenda for Achieving Better Results for Children and Youth with Serious Emotional Disturbance*,[9] the wraparound approach,[10] and multisystemic therapy[11] have noted the need for and importance of documenting the strengths and resources of children and families in treatment planning and outcome monitoring. In addition, professionals in mental health,[12] education,[13] child welfare,[14] and family services,[15] and the families of children who receive these services, have articulated the need to include strengths in the assessment of children and families.

As practiced by mental health workers, school psychologists, social workers, and other direct service workers, strength-based assessment has been conducted in an informal manner.[10] Indeed, these assessments have been referred to as "strength chats," which involve an informal interview between a professional and the child's parent or caregiver. The informal approach has been helpful in furthering the concept of strength-based assessment and gathering useful data for planning purposes. Nonetheless, the informality has raised critical questions about the fidelity of the data collection process, the reliability and validity of the data, and the relevance of the data as a clinical service planning or outcome monitoring measure. Without sound psychometric characteristics, it is unclear whether the informal strength-based assessments are measuring the variable(s) of import in a consistent manner. Many strength-based measures also have lacked normative data that allow clinicians and researchers to compare an individual's performance with that of others from a specific population, thus restricting use for screening, identification, and evaluation purposes.

If strength-based assessment is to serve as an alternative, or at least as a supplement, to deficit-oriented assessment, formal assessment instruments that are normed-referenced and possess adequate technical adequacy (ie, reliability and validity) must be developed. Within this framework, the procedures and results of a national study to determine the factor structure of an instrument that purports to assess the emotional and behavioral strengths of children 5 to 18 years of age are presented. Specifically, this study seeks to determine the factor structure of the Behavioral and Emotional Rating Scale (BERS).[16] Notably, the content validity of the BERS has already been determined.[17] The content validation process included several item development, identification, and discrimination studies that resulted in a 68-item pilot version.

The goal of the present study was to continue to develop and refine the BERS. This was achieved by obtaining a nationally representative sample of school-age children without EBD (NEBD) and a national sample of students with EBD and (1) validating the factor structure of the BERS, (2) deriving

age and gender norms, (3) assessing whether any age and gender differences exist, (4) determining the internal consistency of the BERS factors and the interrelationship of its factors, and (5) evaluating the criterion validity of the BERS.

Methods

Participants

The participants included 2,176 children who ranged in age from 5 years 0 months through 18 years 11 months. The sample selection procedures resulted in a normative sample that was representative of children nationwide. The characteristics of the sample with regard to geographic area, gender, race, residence, ethnicity, family income, educational attainment of parents, and disabling condition are reported as percentages in Table 1. The percentages of these characteristics were compared with those reported in the *Statistical Abstract of the United States*[18] for the school-age population. No statistically significant differences were reported between the groups on any of the demographic variables. Thus, the sample ($n = 2,176$) appeared to be representative of school-age children nationwide. A second national sample of 861 children with EBD was obtained to determine the criterion validity of the scale. All of the children with EBD were school-system-identified as having EBD, had an active individualized education plan (IEP), and were receiving special education services. The demographic features of this sample are presented in Table 1.

Procedures

Data on these children were gathered from adults (ie, teachers, parents, counselors) from 32 states in 1996. Parents and professionals who provided the ratings were approached by the author either by telephone or mail and asked to participate in the norming process. Teachers who agreed to participate were asked to randomly select up to 10 students from their class rosters. Parents who agreed to participate were asked to rate all of their school-age children. Counselors were asked to rate a representative number of children from their caseload.

The adults were given the same instructions. Specifically, they were asked to read each statement on the scale and provide the rating that best described the child over the past 3 months. The ratings employed a 4-point Likert scale (0, not at all like the child; 1, not like the child; 2, like the child; 3, very much like the child). The 68 items were presented in random order.

Results

Factor structure

The correlation matrix of the 68-item pool was subjected to a principal-axis factoring and scree test and eigenvalue criteria of greater than 1.0. These criteria indicated the relative suitability of five factors. These five factors were rotated to a Promax solution. Items were eliminated if they failed to load above .40 on any factor, were redundant to an item with a higher loading, or failed to contribute to the understanding of the factor. The remaining items were factor-analyzed, and five factors were rotated to a Promax solution. This procedure continued until 52 items were retained. The factor loadings and eigenvalues for each factor for this analysis are presented in Table 2.

Factor 1, interpersonal strengths (14 items; eg, considers consequences of own behavior), measures a child's ability to control his or her emotions or behaviors in social situations. Factor 2, family involvement (10 items; eg, maintains positive family relationships), measures a child's participation and relationship with his or her family. Factor 3, intrapersonal strengths (11 items; eg, demonstrates a sense of humor), assesses a child's outlook on his or her competence or accomplishments. Factor 4,

Table 1

Demographic characteristics of the NEBD sample

Characteristics	NEBD percentage of sample ($n = 2,176$)	EBD percentage of sample ($n = 861$)	Percentage of school-age population[a]
Geographic area			
Northeast	19	21	19
Midwest	24	24	24
South	35	38	36
West	22	17	21
Gender			
Male	54	74	51
Female	46	26	49
Race			
White	80	72	80
Black	12	23	15
Other	8	5	5
Residence			
Urban	74	71	78
Rural	26	29	22
Ethnicity			
Native American	1	1	1
Hispanic	8	5	12
Asian	5	1	3
African American	12	23	15
Caucasian	74	70	69
Family income			
Under $15,000	19	17	16
$15,000–$24,999	21	17	16
$25,000–$34,999	19	17	15
$35,000–$49,999	17	22	19
$50,000–$74,999	15	19	20
$75,000 and over	9	8	14
Educational attainment of parents			
Less than bachelor's degree	75	75	76
Bachelor's degree	19	19	15
Master's, professional, and/or doctoral degree	6	6	9
Disability status			
No disability	89		89
Learning disability	5	2	5
Speech-language disorder	2	1	3
Mental retardation	2	1	2
Other handicap	2	4	1
Emotionally disturbed[b]		62	
Behaviorally disordered[b]		45	

[a]*Source*: Data on school-age population derived from US Bureau of the Census, *Statistical Abstract of the United States*, Washington, DC, 1990.

NEBD, no emotional or behavioral disorders; EBD, emotional and behavioral disorders

[b]In some cases students were labeled as both ED and BD.

Table 2

Factors and loadings of the items of the BERS by subscale

Interpersonal strength		Family involvement		Intrapersonal strength		School functioning		Affective strength	
Item	**Loading**	**Item**	**Loading**	**Item**	**Loading**	**Item**	**Loading**	**Item**	**Loading**
10. Uses anger management skills	.83	1. Demonstrates a sense of belonging to family	.79	5. Is self-confident	.78	14. Completes a task on first request	.81	3. Accepts a hug	.74
12. Expresses remorse for behavior that hurts or upsets other	.79	2. Trusts a significant person with his or her life	.71	8. Demonstrates a sense of humor	.58	24. Completes school tasks on time	.86	6. Acknowledges painful feelings	.64
16. Reacts to disappointment in a calm manner	.78	4. Participates in community activities	.66	20. Demonstrates age-appropriate hygiene skills	.53	31. Completes homework regularly	.82	9. Asks for help	.57
17. Considers consequences of own behavior	.84	7. Maintains positive family relationships	.86	21. Requests support from peers and friends	.62	39. Pays attention in class	.76	13. Shows concern for the feelings of others	.66
18. Accepts criticism	.76	11. Communicates with parents about behavior at home	.69	22. Enjoys a hobby	.58	40. Computes math problems at or above grade level	.62	23. Discusses problems with others	.60
28. Accepts responsibility for own actions	.83	15. Interacts positively with parents	.86	26. Identifies own feelings	.70	41. Reads at or above grade level	.57	25. Accepts the closeness and intimacy of others	.78
30. Loses a game gracefully	.75	19. Participates in church activities	.58	27. Identifies personal strengths	.80	47. Studies for tests	.79	34. Expresses affection for others	.79
33. Listens to others	.80	29. Interacts positively with siblings	.69	32. Is popular with peers	.68	51. Attends school regularly	.50		
35. Admits mistakes	.80	36. Participates in family activities	.82	38. Smiles often	.57	52. Uses note-taking and listening skills in school	.76		
37. Accepts "no" for an answer	.78	45. Complies with rules at home	.70	42. Is enthusiastic about life	.74				
43. Respects the rights of others	.86			48. Talks about the positive aspects of life	.72				
44. Shares with others	.74								
46. Apologizes to others when wrong	.78								
49. Is kind toward others	.81								
50. Uses appropriate language	.71								
Eigenvalues	27.01		2.43		2.15		1.67		1.29

Number appearing with each item corresponds to the item's location on the scale.

school functioning (9 items; eg, completes school tasks on time), focuses on a child's competence on school and classroom tasks. Factor 5, affective strengths (7 items; eg, acknowledges painful feelings), measures a child's ability to give and receive affection from others.

Age and sex differences

Standard scores were used for further analyses. For each of the subscales, the raw scores for the entire normative sample were converted to standard scores with a mean of 10 and a standard deviation (SD) of 3. Then the sum of the subscale standard scores was converted into an overall strength quotient with a mean of 100 and an SD of 15.

Standard scores were used to determine any age and gender differences. With respect to age, no differences were noted between the age intervals of 5 years 0 months and 16 years 11 months. With respect to gender, no statistically significant differences were reported for any of the subscale scores or the overall score. Overall, females demonstrated a somewhat higher standard score on each of the scores.

Internal consistency and intercorrelations

To determine the homogeneity of the BERS with the normative sample, internal consistency reliabilities of the subscales and the overall score were calculated. Cronbach[19] coefficient alphas were computed for the entire normative sample at each age interval between 5 years 0 months and 16 years 11 months. The alphas for the entire sample were determined by averaging the alphas across age levels using the z-transformation procedure. The average alphas for the subscales and strength quotient were highly acceptable. Coefficient alphas for the five subscales and overall quotient ranged from .79 to .99.

The standard scores for the subscales and the overall strength quotient were intercorrelated for the normative group. The effects of age were controlled by the "partialing" procedure. The coefficients that depicted relationships among the subscales ranged from .62 to .87; the median was .79. The coefficients that depicted relationships among the five subscales with the overall strength quotient ranged from .75 to .95. All of the relationships were significant at the $p = .01$ level.

Criterion validity

One way of establishing an instrument's criterion validity is to assess the performance of different groups of individuals. Each group's results should make sense, given what is known about the instrument's content to the group. In the case of the BERS, which assesses emotional and behavioral strengths, one would expect that children with EBD would be rated lower by their teachers, counselors, and parents than were children without EBD. Data from the two national samples were used to assess criterion validity. To test for group differences, six t tests were conducted. The groups were significantly different on each of the subscales and overall score ($p = .0001$).

Discussion

The construction of the BERS was performed in a systematic and logical manner, in accordance with the psychometric standards established by the American Psychological Association.[20] The collection of data from a representative sample of children nationwide resulted in the identification of a factor structure of emotional and behavioral strengths and representative national norms. The five factors—interpersonal strengths, family involvement, intrapersonal strengths, school functioning, and affective strengths—appear to be related to important areas of adjustment and functioning for children and adolescents.

Gender differences on the BERS indicated higher, although not significant, ratings for females than males. The standard scores for each of the subscales and the overall strength score showed that adults rated females as possessing slightly more strengths than males. Given that previous research on rating scales has demonstrated that significant differences do exist between males and females, the present results were surprising. Specifically, in large normative studies of rating scales that assess childhood pathology males are generally rated higher than females.[21] The lack of a statistically significant gender effect suggests that strengths or assets of children may develop in a similar manner in males and females. Also, the failure to find age differences was unexpected. This may have been related to the manner in which items were selected. The item selection and scale construction process resulted in the removal of items that would be more age specific (eg, vocational strengths). Nonetheless, this finding is contrary to other rating scale studies that indicate a developmental trend at least with respect to problem behavior. The lack of statistically significant differences suggests that the development of emotional and behavioral strengths may be different than the development of emotional and behavioral problems. Clearly, further research is needed on age and gender issues.

Previous research on the reliability[22] and validity[23] of the BERS indicated that the instrument has sound psychometric properties. Specifically, the research demonstrated that the BERS met acceptable levels of test-retest and inter-rater reliability and convergent validity. The data reported herein are congruent with those findings. Specifically, the reported Cronbach coefficient alphas were well within the acceptable range and indicate that the BERS and its subscales are highly stable and reliable. Also, the strong correlation coefficients among the subscales are indications that these subscales all measure some aspect of emotional and behavioral strength. The correlation coefficients of the five subscales with the overall score suggest that all five subscales together form a behavioral and emotional strength composite score. However, the strong subscale correlations also may indicate some overlap and redundancy across subscale and suggest that some items could be deleted.

The criterion validity of the BERS was established by the scale's ability to discriminate between groups of children. The scores made by children with EBD were as one would expect. Overall, this group scored about 1 SD lower than children without EBD. The results of the *t* tests indicated that across the five subscale scores and total score the two groups were rated as significantly different.

Several limitations should be noted concerning the development of the BERS. First, the samples used in the content validation studies were restricted to professionals and parents who volunteered to participate in the process. Thus, the participant selection process raises questions about external validity. Also, in the pilot studies to determine the items, the data included ratings on secondary-age students, which may have oriented the test more toward adolescents. Second, the children with EBD who were rated were, for the most part, school-system-identified as having EBD. Given that the definition of EBD varies from state to state, this group of children may not have been a homogeneous sample. Nonetheless, the characteristics (see Table 1) of these children appear to be consistent with national reports of children with EBD.[24] Third, the national sample of children without EBD was not randomly selected. The sample comprised adults who volunteered to complete scales on a group of children. This convenience sample may have led to some rater bias and does not inform us about children not rated. Finally, the parent and professional ratings of children were combined in the present study even though previous researchers[25] reported significant cross-informant differences. However, other researchers have reported that the BERS ratings of parents and teachers of children with EBD do not differ.[26]

The strong points of the present research on the BERS are the care taken in selecting the items, the size and representativeness of the national sample, and the size of the EBD sample. The identification of five conceptually sound factors is encouraging and is congruent with research conducted in the areas of resiliency and protective factors.[27,28] Moreover, the reliability and validity of the BERS and its subscales are as high as generally reported in large-scale rating scale research. As with any new instrument, further research needs to be conducted on its psychometric characteristics. First, additional criterion validity research needs to be done with other instruments of child and family

functioning. For example, the BERS and its subscales could be correlated to measures of self-esteem, school attendance and performance, family functioning, and other areas of childhood adjustment. Second, the BERS should be used in longitudinal research to determine whether it is an appropriate instrument for measuring status over time. Finally, future research should investigate the reliability of ratings across informant groups (eg, parent and teacher) to determine whether adults with different roles and experiences judge the strengths of children in a different manner.

Implications for Behavioral Health Services

Despite the need for further research, the BERS appears to possess a logical factor structure, adequate criterion validity and reliability, and nationally representative norms. Based on these strong psychometric characteristics the BERS is recommended for use to practitioners and researchers. The BERS appears to have several uses. First, it can be used as part of a comprehensive assessment to identify children for mental health or special education services for children with EBD. Specifically, the data in the present study indicated that children with EBD scored significantly lower—almost 1 SD—than students without disabilities. Second, the BERS can be used for planning treatment, particularly in determining the goals and objectives for a child's individual treatment plan. Data from the BERS can be used to identify and quantify an emotional or behavioral strength to be achieved as a result of services or to identify a strength a child already possesses that can be used to enhance or reinforce other, less well-developed skills. Third, BERS scores can be used to identify for the child, parents, other family members, and professionals what is going well in the life of a child. The focus on strengths may eventually lead to a more positive parent-professional relationship. Finally, BERS data can be used to assess the outcomes of a specialized treatment program of an individual child or group of children. For example, children who enter a mental health or special education program can be rated at intake and at regular intervals (eg, 6 months) to document change in emotional and behavioral strengths.

A comprehensive mental health assessment should provide information about a child's overall emotional and behavioral functioning. As such, evaluating children's emotional and behavioral strengths should be an important part of a comprehensive evaluation. However, current mental health and special educational practices focus almost exclusively on the assessment of deficits. The inclusion of strength-based information has the potential of changing, or at least adding to, the assessment and treatment planning processes in very fundamental ways. Specifically, strength-based assessments would provide a more balanced view of the child's functioning, treatment goals would include efforts to build on resources and competencies, and communication between parents and professionals would very likely be enhanced.

References

1. Salvia J, Ysseldyke J. *Assessment.* 7th ed. Boston: Houghton Mifflin; 1997.
2. Achenbach TM. *Manual for the Child Behavior Checklist/4–18 and 1991 Profile.* Burlington, VT: University of Vermont, Department of Psychiatry; 1991.
3. Hodges K. *Child and Adolescent Functional Assessment Scale.* Ypsilanti, MI: Eastern Michigan University; 1989. Unpublished manuscript.
4. Quay HC, Peterson D. *Revised Behavior Problem Checklist.* Coral Gables, FL: University of Miami, Department of Psychology; 1987.
5. Reynolds C, Kamphaus R. *Behavior Assessment System for Children.* Circle Pines, MN: American Guidance; 1992.
6. Goodman R. The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry.* 1997;38:581–586.
7. Gresham FM, Elliott SN. *The Social Skills Rating System.* Circle Pines, MN: American Guidance; 1990.
8. Stroul BA, Friedman RM. *A System of Care for Children and Youth with Severe Emotional Disturbances.* Washington, DC: Georgetown University, CASSP Technical Assistance Center; 1994.
9. US Dept of Education. *National Agenda for Achieving Better Results for Children and Youth with Serious Emotional Disturbance.* Washington, DC: Author; 1994.
10. VanDenBerg JE, Grealish EM. Individualized services and supports through the wraparound process: philosophy and procedures. *Journal of Child and Family Studies.* 1996;5:7–22.
11. Henggler SW, Schoenwald SK, Bourdin CM, et al. *Multisystemic Treatment of Antisocial Behavior in Children and Adolescents.* New York: Guilford Press; 1998.
12. Lourie IS, Katz-Leavy J, Stroul BA. Individualized services in a system of care. In: Stroul BA, ed. *Children's Mental Health: Creating Systems of Care in a Changing Society.* Baltimore, MD: Paul H. Brookes; 1996.
13. Nelson CM, Pearson CA. *Integrating Services for Children and Youth with Behavioral Disorders.* Reston, VA: Council for Exceptional Children; 1991.
14. Saleeby D, ed. *The Strengths Perspective in Social Work Practice.* New York: Longman; 1992.
15. Dunst CJ, Trivette CM, Deal AG, eds. *Supporting and Strengthening Families: Methods, Strategies, and Practice.* Cambridge, MA: Brookline Books; 1994;1.
16. Epstein MH, Sharma J. *Behavioral and Emotional Rating Scale: A Strength-Based Approach to Assessment.* Austin, TX: PRO-ED; 1998.
17. Epstein MH. The development and validation of a scale to assess the emotional and behavioral strengths of children and adolescents. *Remedial and Special Education.* 1999;20:258–262.
18. US Bureau of the Census. *Statistical Abstract of the United States.* Washington, DC: Author; 1990.
19. Cronbach L. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16:297–334.
20. American Psychological Association. *Standards for Educational and Psychological Testing.* Washington, DC: Author; 1985.
21. Achenbach TM, Howell CT, Quay HC, et al. National survey of problems and competencies among four- to sixteen-year olds: parents' reports for normative and clinical samples. *Monographs of the Society for Research in Child Development.* 1991;56(3):1–131. Serial no. 37-976X.
22. Epstein MH, Harniss MK, Pearson N, et al. The Behavioral and Emotional Rating Scale: test-retest and inter-rater reliability. *Journal of Child and Family Studies.* 1999;8:319–327.
23. Harniss MK, Epstein MH, Ryser G, et al. The Behavioral and Emotional Rating Scale: convergent validity. *Journal of Psychoeducational Assessment.* 1999;17:4–14.
24. Kauffman JM. *Characteristics of Emotional and Behavioral Disorders of Children and Youth.* 6th ed. Upper Saddle River, NJ: Prentice Hall; 1997.
25. Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: implications of cross-informant correlations for situational specificity. *Psychological Bulletins.* 1987;101:213–232.
26. Friedman KA, Leone PE, Friedman P. Strengths-based assessment of children with SED: consistency of reporting by teachers and parents. *Journal of Child and Family Studies.* 1999;8:169–180.
27. Rutter M. Psychosocial resilience and protective mechanisms. *American Journal of Orthopsychiatry.* 1987;57:316–331.
28. Rutter M. Pathways from childhood to adult life. *Journal of Childhood Psychology and Psychiatry.* 1989;30:23–51.