Evaluation and Critique of the *Behavioral and Emotional Rating Scale-2*

Name Name

Chapman University

Evaluation and Critique of the *Behavioral and Emotional Rating Scale-2*

Description

The *Behavioral and Emotional Rating Scale-2* (BERS-2) is the assessment that will be discussed and critiqued in this paper. Michael Epstein, Ed.D. developed the BERS-2, and PRO-ED Inc. published it in 2004. The BERS-2 is used to identify children's strengths as well as determine if there are absences in any personal strength areas. The BERS-2 collects data on a child's Interpersonal Strength, Family Involvement, Intrapersonal Strength, School Functioning, Affective Strength, and Career Strength. Data are collected using rating scales from three different sources: the child, parent, and teacher.

Content and Use

The BERS-2 includes a total of 52 scaled items in each of the rating scales: Youth Rating Scale (YRS), Parent Rating Scale (PRS), and Teacher Rating Scale (TRS). The YRS and PRS also include five additional, scaled items related to Career Strength. All rating scales also incorporate eight open-ended statements to gather supplemental information about the child. Each rating scale can be administered in approximately 10 minutes. The BERS-2 also includes a Summary Form to be completed by the examiner that can provide a profile of the standard scores.

It is recommended that all three rating scales be used for evaluation if possible; however, some circumstances can prevent the use of one or two of the scales. The YRS is suitable to be completed by children between the ages of 11 and 18. Children younger than 11 or who cannot read at a fifth-grade level should not use the rating scale. The PRS is designed to be completed by a child's primary caretaker, whether this is a parent, another family member, or a foster parent. The manual deems that it is essential for the person filling out the PRS to know the child

for the last 30 days, however, the instructions on the protocol ask the caretaker to think about the child over the last three months. The TRS can be administered to one or more individuals who teach the child in the classroom. However, the TRS can also be completed by other school personnel who have regular contact with the student, such as a school counselor or school administrator.

The protocols all have a high degree of readability, and each protocol is a different color in order to differentiate the YRS, PRS, TRS, and Summary Form. Section One of each protocol asks for identifying information of the student, which can be filled out by the rater. Sections Two, Three, and Four of the protocols are sections that the examiner, not the rater, will complete. Having these sections on the front page may confuse a student, parent, or teacher. Instructions should be clear to all raters that the rater should not complete sections Two, Three, and Four; however, there is no instruction of that on the first page. Section Two includes the scoring section, which the examiner will use to score the ratings once all the statements have been rated. Section Three asks for interpretation and recommendations, and Section Four asks for other pertinent information (e.g., "Who referred the student?" and "What was the reason for the referral?"). The order of the sections may be improved by switching Sections Three and Four since interpretations and recommendations should be the final step after collecting all of the data.

The directions on the second page of each protocol are understandable, and anchors for each number on the four-point Likert are operationally defined (e.g., "3 = If the statement is very much like your child" and "0 = If the statement is not at all like your child"). The anchors for each number are repeated on top of the third page, where the second half of the statements are listed. This is helpful to the rater so they do not have to refer to the previous page. Next to each statement, the numbers are listed in descending order (3, 2, 1, 0), and the rater must select one of

the numbers as a rating for each statement. Some raters may expect a rating scale to list numbers in ascending order instead, so this may cause confusion or errors among some raters. Also, all of the statements are written positively, so no reverse-scoring needs to be conducted. This can pose a challenge to examiners if raters select the same number for all of the items because they are not reading the statements thoroughly.

The backside of each protocol includes eight open-ended, fill-in-the-blank statements to gather supplemental information about the student. Examples of the open-ended statements from the YRS include "My favorite hobbies or activities are," and "The most important people in my life are," and equivalent statements are used on the PRS and TRS (e.g., "The student's favorite hobbies or activities are"). The placement of this section is not ideal because the student, parent, or teacher may forget to complete this section if the examiner is not present because this section is on the backside of the protocol. Because three out of the four sections on the front page of the protocol are only to be used by the examiner, it may be ideal to place the fill-in-the-blank statements about the student on the front page and move the other sections to be used by the examiner on the backside.

The majority of the statements used in the rating scale are relevant to obtaining information about the student's Interpersonal Strength, Family Involvement, Intrapersonal Strength, School Functioning, and Affective Strength. Some of the statements, on the other hand, are not as relevant for determining the student's strength. For example, a low rating of the statement, "Participates in religious activities" may lower a student's Family Involvement score if the student and their family are not religious. Statements such as "Participates in religious activities" and others in the Family Involvement section (e.g., "Participates in community

activities" and "Interacts positively with siblings") may not be applicable for teachers to answer if they do not know much about the student's personal life.

Some of the fill-in-the-blank statements in the final section of the protocol are not relevant to all students either. The statement "This student's favorite sports are," may not be suitable to all students if sports are not their hobby of choice. Another statement, "In the community, this student has worked or volunteered at," may not apply to younger students if they have not had the opportunity to work or volunteer anywhere. The fill-in-the-blank statements should be used as a guide to what questions to ask the student, parent, or teacher in an interview; however, raters should be required to fill in all of the statements if they are not relevant.

The test manual for the examiner provides relevant information about possible uses for the BERS-2. The manual explains that the BERS-2 can be used to identify emotional or behavioral disorders in students. However, it can also be used to target specific goals and document progress. The manual is easy to use, and the table of contents clearly indicates where to find information about interpreting results, normative samples, test reliability, and validity of the test results. The section regarding administration, however, does not clearly state whether the examiner must be present when the student, parent, or teacher completes the rating scale. There are situations, however, when it is instructed that the examiner must read the statements to the rater (e.g., if the rater cannot read the statements). The examiner's manual also provides helpful figures when demonstrating a completed rating scale, and the graphs throughout the manual for the normative sample, reliability studies, and validity studies are clear and easy to read. Epstein (2004) also explains in the manual that the examiner should familiarize themselves with the BERS-2, its theoretical background, and have solid knowledge on how to interpret results.

Standardization Sample and Norms

The BERS-2 contains four sets of normative data, including two sets for the TRS, one set for the PRS, and one set for the YRS. The original BERS rating scale was only to be used by teachers, and the BERS-2 now includes rating scales for the parent and student. Because the BERS-2 TRS is derived from the original BERS, Epstein (2004) states that the normative data gathered for the TRS was collected in 1996 for the original BERS. Therefore, no new normative data has been collected for the TRS. Epstein (2004) indicates toward the beginning of the manual that minimal changes were made from the original BERS to the BERS-2 TRS, and the changes mostly include the additional open-ended statements about the student. The items in the BERS-2 TRS have not changed from the original BERS; therefore, Epstein (2004) focused on gathering normative data for the PRS and YRS for the BERS-2.

The normative data samples for the PRS sample and both sets of the TRS samples include children between the ages of 5 and 18 years old, and YRS includes children between 11 and 18 years old. Other demographic data were collected for the samples, including geographic area, gender, race, Hispanic ethnicity, family income, educational attainment of parents, and disability status. All of the demographic data were compared to the national percentages from *The Statistical Abstract of the United States*, and the majority of the sample percentages in all normative data samples are representative based on the national percentage. Race norms in all of the samples are extremely limited because the only options are White, Black, and Other. Participants were asked in a separate question if they identified as Hispanic. Although the percentage of participants who identified as Other were low (5% to 13%), other race options should have been included in order to gain a better perspective of participants. Broader race options are also necessary because the BERS-2 is meant to be conducted on any student, meaning that students of all races should be represented in the normative sample.

The TRS includes two sets of normative data, and they are based on two different samples of children: children not identified with emotional or behavioral disorders (NEBD) and children diagnosed with emotional or behavioral disorders (EBD). The normative sample for the TRS NEBD contained 2,176 total participants and an adequate number of participants in almost every age group. Only participants in the 16-year-old and 18-year-old groups had less than 100 participants, which included 98 and 82 participants, respectively. Other age groups ranged from having 108 to 297 participants. Data would have been stronger for this normative sample if there were at least 100 participants in each age group, and data would have been superior if there were at least 200 participants in each age group. Each age group then stratified the sample by other characteristics such as geographic region, gender, race, and Hispanic ethnicity.

The TRS EBD normative sample included 861 participants, which is less than half the number of participants in the TRS NEBD sample. Because this sample has a more specific population being diagnosed with an emotional or behavioral disorder, it is expected for this sample to be smaller. The male to female ratio, however, is skewed because 74% of the students were male. No comparison data are given to determine if gender percentages of participants are representative of the national population with an emotional or behavioral disorder. The number of students in each age category ranged from 13 to 125. Only ages 15 through 17 included over 100 participants.

The PRS and YRS included 927 and 1,301 participants, respectively. Neither of these normative samples provides information as to how many students in the sample have been diagnosed with an emotional or behavioral disorder. Unlike the TRS, the PRS and YRS already provide a more limited normative sample because there is a lack of information on emotional or behavioral disorders. Therefore, it may be difficult to compare students who have EBD in the

schools to these normative samples. The PRS includes over 100 participants for the 11 and

12-year-old age groups, and it includes less than 100 children for all other ages. The YRS

includes over 100 children for all ages except for 18-year-olds ($N = 64$), and ages 12 through 14

include over 200 children.

Overall, Epstein (2004) made significant efforts in order to gather normative data for the

BERS-2; however, there are several limitations to the normative samples that may cause

restrictions.

Scores and Interpretation

A detailed description of how to score the BERS-2 on each of the rating scales is clearly

provided in the examiner's manual. Examples of completed rating scales are provided in the

manual to show examiners exactly how to score. Epstein (2004) emphasizes that raw scores must

be accurate in order for them to be appropriately converted into standard scores. The scoring

procedure for the BERS-2 is straightforward overall. The examiner must add all of the raw

scores for each item in a subscale for all of the subscales on the rating scale. The raw scores for

each subscale are then converted into percentile ranks and scaled scores, and a sum of all of the

scaled scores is calculated to then determine the BERS-2 Strength Index for the student. This

procedure is completed for each of the rating scales. The Summary Form includes a graph on the

backside that an examiner can use to plot the scaled scores for each subscale in each of the rating

scales. This visual can be extremely helpful for the examiner when interpreting the results and

conveying the results to teachers and parents.

The standard scores for each of the subscales are called scaled scores for the BERS-2. For

the subscales and scaled score distribution, the mean is set at 10, and the standard deviation is set

at 3. Epstein (2004) says that the scaled scores for each of the subscales are more valuable

compared to the percentile ranks because scaled scores can be used for comparison to the normative sample as well as to the same student's other scores.

Epstein (2004) provides information on percentile ranks for the BERS-2 and stresses that percentile ranks should be used with caution due to their inherent advantages and disadvantages. One of the disadvantages mentioned by Epstein (2004) is that percentile ranks are not interval data; therefore, they do not have equal differences between them. Percentile ranks are provided in many educational and psychological tests in order to highlight significant strengths or areas for improvement. Percentile ranks are more appropriate for behavioral measures, such as the BERS-2, than other measures.

The BERS-2 includes a Strength Index that is a converted, composite score of all of the subscales. Epstein (2004) claims that the Strength Index is the most reliable of all the scores generated from the BERS-2, which is explained further in the reliability section of the manual. The Strength Index is also placed on a normal distribution with a mean of 100 and a standard deviation of 15. A normal distribution, however, is not the most appropriate distribution for the BERS-2 because it measures behavior with non-equal interval data. A skewed distribution using the median as the central tendency would make the Strength Index stronger.

Information about omitted and multiply-marked items in the rating scale is also provided for the examiner. Two items are allowed to be omitted or multiply-marked for each subscale (except the Career Strength subscale) in order for the subscale composite to still be computed and usable. When an item is omitted or multiply-marked, the examiner must calculate the average raw score for the other items in the subscale and use the calculated average for the item that has a rating error. Any subscale that has more than two omitted or multiply-marked items is

considered unscorable. Overall, a total of ten unscorable items are allowed for the entirety of one BERS-2 rating scale as long as there are no more than two unscorable items in each subscale.

Since the TRS includes two norms, the EBD and NEBD norms, the examiner must select the appropriate norms to use when scoring the TRS. If examiners are using the BERS-2 to determine if the student has an emotional or behavioral disorder, then the NEBD norms should be used. In this case, if the student has lower scores than the NEBD sample, this may indicate that the student has several areas for improvement in the strength categories and may have EBD. Epstein (2004) says that the EBD norms are intended to be used when evaluating change or progress of a student already identified with EBD as a result of class placement changes or interventions implemented with the student. EBD norms can also be used to determine strengths and possible goals for a student with EBD. These norms are helpful for examiners to use in order to interpret the data accurately; however, the two separate norms can only be used for the TRS. The PRS and YRS do not have specific normative samples. For interpretation of the BERS-2 PRS and YRS to be significantly more accurate and helpful for students, the PRS and YRS should have NEBD and EBD norms as well.

Epstein (2004) provides a thorough section in the manual regarding interpreting results from the BERS-2. There is a chart that shows the ranges of scores for each subscale and Strength Index that indicates the possibility of the student having EBD. Epstein (2004) also emphasizes that the scores from the BERS-2 alone should not be used to diagnose or classify a child as having EBD. As with any assessment, one particular assessment alone should not label a student. Also, Epstein (2004) provides a section in the manual with direction on with whom results are allowed to be shared and keeping the scores of the rating scales as confidential as possible.

Psychometric Properties

Reliability

The BERS-2 uses a Cronbach's coefficient alpha in order to calculate internal consistency for each subscale and composite score of each rating scale. The Cronbach's coefficient alpha was the ideal selection to calculate internal consistency because the rating scale has a multiple response Likert format. The internal consistency for the TRS is calculated separately using NEBD and EBD norms. The internal consistency for the TRS EBD and TRS NEBD used 11 and 12 age categories, respectively. Ages 16 to 18 are combined for each, and ages 5 to 6 are combined for the TRS EBD group because of smaller sample sizes in those age groups. The Cronbach's alpha for the TRS subscales in both the EBD and NEBD norms ranged from 0.84 and 0.96. The Cronbach's alpha for the Strength Index of the TRS NEBD was 0.98, and the alpha for the Strength Index of the TRS EBD was 0.97. All of these internal consistencies meet general guidelines for ideal internal consistency. The internal consistency for both Strength Indexes is particularly high, accounting for more than 97% of method error.

The internal consistency for the subscales and Strength Index of the PRS was particularly high, as well. The Cronbach's alpha for the subscales on the PRS ranged from 0.80 to 0.93, and the alpha for the Strength Index was 0.97. The internal consistency for the subscales and Strength Index on the YRS was slightly lower than the TRS and PRS; however, the internal consistency was still sufficient. The Cronbach's alpha for the subscales on the YRS ranged from 0.79 to 0.88, and the alpha for the Strength Index was 0.95.

Duppong Hurley, Lambert, Epstein, and Stevens (2015) conducted a study using the BERS-2 with children in the residential care group population. This was the first study using the BERS-2 in this particular setting, and they used the TRS. When evaluating internal consistency, Duppong Hurley et al. (2015) found that internal consistency measured with Cronbach's alpha

was between 0.76 and 0.94 in the subscales, and the alpha coefficients for the Strength Index among gender and race categories were all above 0.90. This study confirms that the BERS-2 maintains high internal consistency across settings (Duppong Hurley et al., 2015).

Epstein (2004) also provides a detailed explanation of the Standard Error of Measurement (SEM) and how it is used to calculate confidence intervals. Charts are provided to display the SEM for each subscale on the rating scales by age category. The TRS is again separated by EBD and NEBD samples. The SEM for all of the subscales on the PRS, YRS, and TRS (EBD and NEBD) was 1. The average Strength Index SEM for the PRS, YRS, and TRS (EBD) was 3, and the average SEM for the TRS (NEBD) was 2. Epstein (2004) says in the manual that a small SEM can be interpreted as having higher confidence in the assessment results. Since the SEM tells the examiner the average amount of error around the score, it can be interpreted that the subscales and Strength Index all generally have small SEM, and Epstein (2004) suggests that there is high confidence surrounding the subscale and Strength Index scores.

In order to measure consistency in test performance over time, reliability was calculated using the test-retest method. Six studies were conducted in order to measure reliability over short and long periods of time. Study 1 was conducted to measure the reliability of the TRS using a teenage EBD sample in DeKalb, Illinois over a two-week interval. The sample size was small, which is typical for test-retest studies, consisting of 59 students between the ages of 14 and 18. The correlation between the first and second testing periods for the subscales of the TRS ranged from 0.85 to 0.99, and the correlation for the Strength Index was 0.99, indicating a high correlation rate.

Studies 2 and 3 measured Pearson's correlation using the TRS for students between ages 5 and 9 in Lincoln, Nebraska. Study 2 measured general education students, while Study 3

measured at-risk of EBD students. Race and ethnicity for the participants were not gathered for Study 2, and Strength Index correlations were calculated for neither Study 2 nor Study 3. Both of these studies were conducted over a long-term period of six months. The Pearson's correlation for all of the subscales in Study 2 and 3 ranged from 0.53 to 0.68, with only the correlation for the Intrapersonal Strength subscore falling below 0.60. For a long-term study, correlations above 0.60 indicate strong correlations.

Studies 4 and 5 measured correlation for the PRS, and Study 6 measured correlation for the YRS. All of the studies were conducted on general education students and small sample sizes for a two-week interval in Lincoln and Beatrice, Nebraska. All of the subscale and Strength Indexes for all three studies are shown to be highly correlated with all of the subscore and Strength Index correlations being above 0.80.

The majority of the correlations determined to measure test-retest reliability were strong; however, there were some limitations in these studies. The main limitation for the test-retest studies is the location in which the test-retest studies were completed. Five out of the six studies were completed in two cities in Nebraska, and one study was completed in DeKalb, Illinois. These limited locations indicate that test-retest results are not representative of the students in the nation. Another limitation was race, similar to the normative samples. Race is only indicated with White, Black, and Other options, which is also not representative of the students in the country with whom the BERS-2 will be used. To strengthen the data even further, a dependent *t*-test could have been used in order to determine if there were statistically significant differences among the means.

Interrater reliability was used to calculate the variability in ratings among the various raters. Three separate studies were conducted to assess interrater reliability. Study 1 involved the

TRS among 96 teenage males who were identified with having serious EBD in DeKalb, Illinois, but no race or ethnicity data were provided. Two separate educators completed the TRS for each student, and the correlation between raters for each subscale ranged from 0.83 to 0.96, and the Strength Index correlation was 0.98. Study 2 compared ratings from parent-to-teacher in Washington, D.C. among 20 students with severe EBD. No gender, race, or ethnicity information was provided for this study. The correlation between raters for each subscale ranged from 0.20 to 0.67. The Intrapersonal Strength subscore correlation was an outlier at 0.20, and Epstein (2004) contributes this to parents seeing more of their child's intrapersonal strengths than their teachers. All of the other correlations for the subscores were above 0.50, which is sufficient for teacher-to-parent correlations. No data are provided for the Strength Index correlation. Lastly, Study 3 used the normative sample of 296 general education students between the ages of 10 and 18 to compare ratings between the PRS and YRS. All data regarding gender, race, and ethnicity are provided. The correlations among the subscales ranged from 0.50 to 0.63, and the Strength Index correlation is 0.54. All of the correlations in Study 3 are at or above 0.50, which is considered ideal for child-to-adult ratings.

Epstein (2004) provides a summary of all of the reliability results at the end of the chapter. These graphs are helpful in comparing internal consistency, test-retest reliability, and interrater reliability for the TRS, PRS, and YRS. It should be emphasized, however, that the numbers provided in these graphs for each of the rating scales are from different samples. Therefore, the numbers may be slightly different if they were all from the sample. In addition, the locations from which the samples came are limited as well, with only the third study in interrater reliability to use the normative sample.

Validity

Epstein (2004) went to great lengths in order to measure content-description validity for the BERS-2. Other assessments such as the *Child Behavior Checklist, Behavior Rating Profile-Second Edition,* and *Revised Behavior Problem Checklist* were analyzed for their item content, and over 200 professionals in the education, mental health, and child welfare professionals were sent a questionnaire asking about possible items to measure emotions and behavior in children.

After researching possible items and ending up with over 1,200 possible behaviors, item discrimination was used in order to narrow down the possible item choices. Epstein (2004) explains that present-absent dichotomy was used instead of the traditional right-wrong dichotomy used in other assessment measures. A study for the original BERS was used to determine if the BERS had the ability to discriminate between children with serious emotional disturbance (SED) and those without (NSED). ANOVAs and chi-square analyses were used, and they identified items on which the SED children and NSED children did not differ. These items were removed and further narrowed the number of items. Factor analysis was also used to narrow down items even further, and the analysis helped identify the five core significant strength areas now used in the BERS-2.

Cross-validation was measured by exploratory factor analysis by Furlong, Sharkey, Boman, and Caldwell (2007). Exploratory factor analysis investigates how items are fit into particular clusters or subscales. Furlong et al. (2007) conducted the BERS-2 YRS on a general education population of students at a comprehensive high school and a juvenile probation department sample. The results determined four-main factors in which the items of the BERS-2 YRS should be grouped. All of the School Functioning items grouped together because they were significantly unique from the rest.

On the other hand, Furlong et al. (2007) found that items in the Intrapersonal Strength category were spread out among factors, including Interpersonal Strength and Affective Strength. Therefore, the exploratory factor analysis conducted by Furlong et al. (2007) differed from the five-factor model developed in the BERS-2 in factor analysis by Epstein (2004). This difference may be because Furlong et al. (2007) focused on the YRS only, and students may view aspects such as Intrapersonal Strength differently than teachers or parents.

Criterion-prediction validity procedures indicate how an assessment predicts an individual's performance on particular measures. Concurrent validation is measured in different studies evaluating the TRS, PRS, and YRS. Study 1 correlated the scores from the BERS-2 TRS to the *Walker-McConnell Scale of Social Competence and School Adjustment-Adolescent Version.* This study was conducted on severely emotionally disturbed teenagers in DeKalb, Illinois. All scores between the subscales on both assessments were statistically significant and positively correlated. Other studies were conducted in order to establish criterion-prediction validity on general education students using the PRS and YRS. Epstein (2004) provides tables displaying the correlations between each subscale of each assessment used in the studies, and almost all of them indicate appropriate direction between them (e.g., negative direction between the maladaptive behavior subscale of the *Systematic Screening for Behavior Disorders* and the Interpersonal Strength subscale of the BERS-2).

Construct-identification validity was assessed for the BERS-2 in order to determine if the assessment was measuring what it was intended to measure. One way of determining construct-identification validity is group differentiation. Epstein (2004) explains that one must compare students with and without EBD, and it would be expected that students with EBD receive lower scores on each of the subscales when using the TRS, PRS, and YRS compared to

those without EBD (NEBD). When the study was conducted to establish this construct validity,

students with EBD scored lower on every subscale on the TRS, PRS, and YRS than their NEBD

counterparts and below the mean of the normative sample. Their average scores for each of the

subscales ranged between 6 and 9, and the averages for the normative sample were set at 10.

Students with EBD received the highest scores on the YRS compared to the PRS and TRS.

A second study under group differentiation was conducted to determine if there was a

differentiation on the TRS among students with no disability, a learning disability, and a

behavioral disorder. As expected, students with no disability were rated the highest among

teachers in all subscale categories compared to students with a learning disability and behavioral

disorder with a median score of 11. Students with a behavioral disorder were rated lower on the

five subscales than the students with a learning disability as well on a statistically significant

level. Students with a learning disability scored a median of 9 among the subscales, and students

with a behavioral disorder scored a median of 7 among the subscales.

Subscale interrelationships are also measured in order to establish validity. Epstein (2004)

goes on to explain that subscales should be correlated because they measure strengths; however,

they should not be too correlated because they are intended to measure unique strengths. The

table listing the correlations indicates that correlations range between 0.37 and 0.87 when

comparing averages of all the subscales on each of the rating scales. The highest correlations are

between Interpersonal Strength and Intrapersonal Strength ($r = 0.87$) as well as Interpersonal

Strength and School Functioning ($r = 0.87$). The least correlated subscales included School

Functioning and Career Strength ($r = 0.37$) and Affective Strength and Career Strength ($r = 0.37$).

Convergent validity was not discussed in the BERS-2 manual, but Duppong Hurley et al. (2015) measured convergent validity among a group of students in a residential care setting. Convergent validity is determined by comparing an assessment to related constructs that are expected to be correlated. To determine convergent validity, the BERS-2 subscales were compared to subscales and total scores for the *Symptoms and Functioning Severity Scale* (SFSS) and the *Child Behavior Checklist* (CBCL). The Strength Index of the BERS-2 was positively correlated with the total score on both the SFSS ($r = 0.60$) and the CBCL ($r = 0.57$). Both of these correlations indicate adequate convergent validity, and all subscores except those comparing Intrapersonal Strength on the BERS-2 also had statistically significant correlations (Duppong Hurley et al., 2015).

Conclusions

The BERS-2 is an easy-to-use assessment in order to gather information about a student's behavioral and emotional strengths and areas for improvement. Overall, all of the BERS-2 rating scales can be administered and scored easily; however, there are some areas for improvement in the layout of the BERS-2. The assessment also underwent norming procedures for all three rating scales. The TRS could use updated norming samples because it was normed for the original BERS, and the PRS and YRS should be normed with EBD and NEBD samples like the TRS for consistency and accuracy. Norming data would also be stronger with more specific race and ethnicity information. Epstein (2004) went through many studies to develop reliability and validity for the BERS-2. Although overall reliability and validity for the BERS-2 were shown to be strong, the studies that they were based on were extremely limited in terms of geographic region because only three states were used, and only one study used the normative sample.

The BERS-2 is a reasonable assessment to gain information about a student's strengths and areas for improvement; however, it should not be the defining factor in diagnosing or classifying a student with an emotional or behavioral disorder. Practitioners should use the BERS-2 with caution based on the limitations stated. The TRS is sufficient to be used in measuring the strengths of students already identified with EBD using the EBD norms, but there is no normative data provided for EBD students in PRS and YRS normative samples. Therefore, the PRS and YRS should only be used to identify strengths and areas for improvement in students who do not have EBD or have yet to be classified as such. Other forms of assessment should be used in conjunction with the BERS-2 when classifying a student with EBD and developing interventions.

References

Duppong Hurley, K., Lambert, M., Epstein, M., & Stevens, A. (2015). Convergent validity of the

    strength-based behavioral and emotional rating scale with youth in a residential setting.

    Journal of Behavioral Health Services & Research, *42*(3), 346–354.

    https://doi.org/10.1007/s11414-013-9389-0

Epstein, M. H. (2004). Behavioral and Emotional Rating Scale-2: A strength-based approach to

    assessment. Austin, TX: PRO-ED.

Furlong, Sharkey, Boman & Caldwell (2007). Cross-validation of the Behavioral and Emotional

    Rating Scale-2 youth version: An exploration of strength-based latent traits. *Journal of*

    Child & Family Studies, *16*(5), 696–711. https://doi.org/10.1007/s10826-006-9117-y