Behavior Evaluation Scale–2, Home Version (BES-2HV)

The Behavior Evaluation Scale–2, Home Version (J. McCarney, 1994) is a norm-referenced rating scale intended to provide a measurement of behavior in the home environment; to identify the characteristics of emotionally disturbed (or behaviorally disordered) students, as defined in federal legislation; to measure changes in behavior over time, and to ascertain areas needing behavioral intervention. The BES-2HV consists of 73 items arranged into five subscales:

1. Learning/Self-Control (8 items)
2. Interpersonal/Social (14 items)
3. Inappropriate Behavior Under Normal Circumstances (35 items)
4. Unhappiness/Depression (7 items)
5. Physical Symptoms/Fears (9 items)

Each item is rated on a 7-point scale:

1. Not in the rater's presence
2. One time
3. Several times
4. More than one time a month, up to one time a week
5. More than one time a week, up to once a day
6. More than once a day, up to once an hour
7. Once an hour (or more)

Scores

Raw scores on each subscale are converted to standard scores with a mean of 10 and a standard deviation of 3. Subscale standard scores are summed to convert to quotients (presumably a standard score with a mean of 100 and a standard deviation of 15) and percentiles.

Norms

The BES-2HV was standardized on 1,769 individuals ranging in age from 4.5 years to 21 years. However, separate norms are provided for males and females in four age ranges (5 through 7 years, 8 through 10 years, 11 through 14 years, and 15 through 18 years); no norms are provided for individuals younger than 5 or older than 18. An unreported number of these students had been identified as behaviorally disordered. The students resided in 18 states. The author compares the total standardization sample to the U.S. population (presumably the 1990 census). However, the information of interest (that is, the representativeness of each age–sex normative sample) is missing. As a whole, the normative sample greatly overrepresents European Americans, blue-collar workers, individuals from rural areas, and individuals from the North Central region of the United States.
**Reliability**

Stability was estimated using 201 individuals from the normative sample who were rerated within 30 days of their initial rating. Correlations, across all ages, between the first and second ratings ranged from .88 to .93 for the five subscales and was .90 for the total score.1

Interrater agreement for the total score was estimated using 147 pairs of parents and 200 students. Correlations between raters were calculated at 13 age groups (including individuals who were older or younger than those used for the normative conversion tables).2 Interrater agreement ranged from .83 to .91; in 3 of the 13 age groups, the agreement equaled or exceeded .90.

Using the total normative sample, internal consistency was estimated by KR-20 for each subscale.3 Estimates ranged from .74 to .93; only one subscale had an estimated reliability greater than .90.

**Validity**

---

1 The author does not report whether the correlations are based on raw or standard scores. Correlations based on raw scores would be inflated to the extent that raw scores and age are correlated.

2 The number of students in each age group is not reported. However, if the students were distributed equally among age groups, there would be about 15 students in each age group.

3 As you will recall, KR-20 (a special case of coefficient alpha) is used with dichotomously scored items. The items on this scale are scored on a 7-point scale. Thus, how KR-20 could be used is unclear. Moreover, because internal consistency was estimated across ages, these reliabilities would be inflated to the extent that raw scores and age are correlated.
The technical manual is poorly organized, and validity information is incompletely reported. Some evidence for content validity comes from the way in which the scale was developed. Items for the BES-2HV were derived from the literature and from input from education professionals and parents and guardians. In addition, item–total correlations and the distributions of responses to each item were examined. Moderate correlations between the BES-2HV and the BRP for 49 children provide some evidence for criterion-related validity.

Construct validity of the BES-2HV was addressed in several ways. First, a series of factor analyses suggests that the scale is factorially complex. However, these analyses do not suggest that the subscales of the BES-2HV correspond to the instrument's factor structure. Second, the ratings for a group of 104 students previously identified as having behavioral disorders were compared to those for a randomly selected subsample of 104 individuals from the normative sample. The students with behavioral disorders received ratings that were about 1 to 1.3 standard deviations below the mean, while the randomly selected students had ratings at the mean. However, variances are not reported, and so it is impossible to ascertain the degree of overlap between the groups.

Summary

---

4 The author reports 17 factors with Eigenvalues greater than 1 but fails to report the percentage of variance accounted for by each factor. A skree plot of factors suggests, to us, a single-factor solution. When individual subscales are factor-analyzed, some have several factors.
The BES-2HV is a behavior rating scale for use with individuals from 5 through 18 years of age. The 73 items are clustered into five subscales that parallel the criteria for identification of students who are seriously emotionally disturbed as provided in federal regulations. Although reliability estimates are usually sufficient for screening purposes, the BES-2HV's norms do not appear representative of the U.S. population. Information about the scale's validity is presented incompletely and is therefore limited.