Access Mental Measurements Yearbook:

- Social Work library guide - [http://libguides.usc.edu/socialwork/measurements](http://libguides.usc.edu/socialwork/measurements)
- Home page of library - [https://libraries.usc.edu/](https://libraries.usc.edu/)
  - Select Databases
  - On next page, type the name of the database
After typing the name of the database, select Go.
Basic search in Mental Measurements Yearbook (MMY)

Start with a broad search. For example: the variable you are trying to measure and the population.
Not a lot of search results?

Try synonyms or related terms like: aggression, aggressive behavior, bullying.
Review of the Test

Review of the Clinical Assessment of Behavior by MIKE BONNER, Assistant Professor, Department of Psychology, University of Nebraska at Omaha, and METTA K. VOLKER-FRY, Graduate Student, School Psychology Program, University of Nebraska at Omaha, Omaha, NE.

DESCRIPTION. The Clinical Assessment of Behavior (CAB) is a broad-band behavior inventory-report rating scale for use in clinical, educational, and research settings for children ages 2 through 18 years. There are three rating forms: the Teacher Rating Form (CAB-T; 70 questions), Parent Rating Form (CAB-P; 70 questions), and Parent Extended Rating Form (CAB-PX; 170 questions). Completing the rating scales is appropriate for any adult with "a minimal literacy level" (professional manual, p. 9). All three scales require the respondent to rate the frequency of behaviors displayed "daily" using a 5-point Likert scale. Scoring must be done using the CAB Scoring Program (CAB-SP), which generates scores for all three of the CAB Rating Forms, which includes raw scores, T-scores, percentile rank, confidence intervals, qualitative classifications, and a graphical profile display. The manual provides raw-score, T-score, percentile rank, and confidence interval data, although the CAB forms are not designed to be hand-scored. The CAB-PX will prorate raw scores when at least 80% of the items in a particular scale have appropriate completion.

Interpretation is organized using Clinical scales, Adaptive scales, and a CAB Behavioral Index (CBI), which combines both scales. The Clinical scales are composed of indices for Internalizing Behaviors, Externalizing Behaviors, and for the CAB-PX only, Critical Behaviors (CBR). The Internatizing Scale includes measures of Anxiety and Depression (called "clusters"). The Externalizing Scale clusters include Anger, Aggression, Bullying, and Conduct Problems. The Adaptive scale consists of measures for Social Skills, Competence, and for the CAB-PX, Adaptive Behaviors.

DEVELOPMENT. Items were developed from a review of content and psychometric properties of existing behavior rating scales and, in a subsequent analysis, four preschool behavior rating scales, scales for items that had high agreement in four different biological and/or environmental studies (Kazdin et al., 1989; Walker, 1989). Their goal was to develop items with high inter-rater agreement and/or high internal consistency as well as items that had relevance for DSM-V diagnosis. A process called "logical reanalysis" for best-fit items was used to develop the CAB-SP scales. From this, the abbreviated parent and parent scales were developed so they would have matching procedures.

The CAB was normed and standardized over a 2-year period using two sampling procedures. The normative group is composed of 2,111 parents of children aged 2 to 18 and 1,689 teachers of 5- to 18-year-old students. The first method was a traditional direct solicitation for participation in standardization from schools, youth organizations, and mental health professionals. Data from the remaining data were generated through an Internet survey research company. Parents (of children aged 2 to 18) and teachers (of children aged 5 to 18) who were members of the company's online survey panel were selected based on a stratified sampling plan designed to match the U.S. Bureau of the Census (2001) in terms of: age, gender, race/ethnicity, parental education, and geographic region. This method yielded respondents from 49 states.

Univariate analyses of variance were conducted controlling for the child's age, gender, grade, race/ethnicity, and urban/suburban/rural residency with results demonstrating that the two sampling methods were equivalent for mean CAB scale scores in the parent sample (accounting for up to 2.2% of the variance). The teacher sample demonstrated more variance due to the normative sampling method (between 9.1% and 12.2% of variance). The manual states this difference is likely the result of a systematic orientating procedure for paper-pencil standardization (they were told to base ratings on a randomly selected student) and the Internet standardization (who could have selected the target student to rate based on unknown criteria).

The normative sample was matched to 2001 U.S. Census data for only age, gender and race/ethnicity, geographic region, and parent and teacher education level. There is a fair amount of discrepancy for some of the variables. For example, the CAB-T norms overestimate males. There are small to moderate differences across race/ethnicity and age as well.

TECHNICAL.

Reliability. Internal consistency coefficients for the Clinical, Adaptive, and Behavioral Indices across the three forms are between .89 and .96. For the subscales (clusters), coefficients range from .88 to .97. There are some patterns of strength and weakness for internal consistency across forms and subscales for age and gender that need to be considered by potential users. Test-retest reliability (time intervals varied across forms but were on average about 18 days) was reported for a small (N = 122) sample with correlations between .75 and .92 for indices and subscales. Inter-rater reliability was evaluated for 288 children and adolescents. The Parent Forms were administered to 31 pairs of biological parents yielding a correlation of .62 on the CBI and a range of .70 to .90 across the more specific subscales. Teacher and parent inter-rater reliability correlations across the scales and subscales ranged from .40 to .66, typical for cross-context comparisons. The Standard Error of Measurement (SEM) is between 1.4 and 0.45 T-score points. Low SEVs, such as these, are indicators that the test has good evidence of reliability.

Validity. Evidence for content validity comes from the empirical approach used to derive items in the scale development. Criterion-related validity evidence is offered through comparisons with the Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1999) and the Devereux Scales of Mental Disorders (DSMD; Naglieri, LeBuffe, & Pfeiffer, 1994). The CAB and BASC correlations between theoretically equivalent scales and clusters ranged from .57 to .77. There are a few components of the CABs that do not correlate as well as would be expected with the BASC. In comparisons with the DSMD, the coefficients between similar scales and clusters ranged from .57 to .79. These findings give compelling evidence for criterion-referenced validity. Construct validity evidence was offered with correlations between specific clusters and the corresponding index, with a range between .71 to .95. Item-level factor analysis was conducted using the normative sample and provided support for four factors within the CAB-T and CAB-P and six factors within the
Example of Citation from MMY

Last name, first initial of review author. (Year of MMY publication). Test review of the name of test. In name(s) of MMY editors, *The edition mental measurements yearbook* [electronic version]. Retrieved from [URL].

Example: