Systematic Screening for Behavior Disorders (SSBD)

The Systematic Screening for Behavior Disorders (H. M. Walker & Severson, 1992) is a series of three interrelated measures of behavior for children in grades 1 to 6. This three-stage sequential screening system incorporates a procedure known as gating. Gating refers to the progressively more detailed and precise assessment procedures that identify students who may be at risk, in this case for behavioral problems. Assessment in Stages 1 and 2 of the SSBD relies solely on teacher judgments of child behavior across externalizing and internalizing dimensions. The final assessment step, Stage 3, consists of repeated observations conducted by a trained observer (someone other than the teacher) to validate teacher concerns about students who have passed through the previous two gating procedures.

The authors state that "the SSBD provides for mass screening of all students enrolled in a regular classroom. The system also gives each child an equal chance to be screened and identified for either externalizing or internalizing behavior disorders and problems" (p. 3). Referral for further assessment is contingent on a child's passing through the entire multiple-gating process. A description of the three-stage process follows.

Stage 1: Rank-Ordering of All Students Enrolled in General Classrooms

Teachers rate (for a minimum of 30 days) all students enrolled in their classroom on both externalizing and internalizing dimensions of problem behavior. Clear operational
definitions are provided to facilitate teacher ratings of children's behavior. Initially, teachers nominate ten children according to the externalizing and internalizing descriptions provided. Next, teachers rank-order the ten students from one to ten for each dimension. The three top-ranked students who best match the behavioral profile for either the internalizing or the externalizing dimension may be passed through to Stage 2.

*Stage 2: Teacher Rating Scales*

The teacher rating scales include a 33-item critical events index (CEI) to rate low-frequency, high-intensity events. For example, items include "Steals," "Sets fires," or "Damages property." Two open items are provided on the CEI scale so that teachers can provide information on critical events that may not appear on the prepared list. Although these teacher-provided critical events are included in the total score, no criteria or guidelines on acceptable or unacceptable items are provided to teachers. This may invite an unnecessary degree of subjectivity that could skew results or decisions.

The second part of the Stage 2 procedure produces a combined frequency index (CFI), which is calculated by summing scores on Adaptive Behavior and Maladaptive Behavior scales. The Adaptive Behavior section is a 12-item scale with specific behaviors that are rated by a teacher to describe a child's current functioning. Items include statements such as "Follows established classroom rules" and "Initiates positive social interactions with peers." The Maladaptive Behavior scale is composed of 11 items. Examples from this scale include statements such as "Refuses to participate in games and
activities with other children at recess" and "Creates a disturbance during class activities."

Stage 3: Observations of Academic and Social Behavior

All students who pass through Stages 1 and 2 are further observed using two different observational measures. The purposes of these observations are to (a) verify teachers' rankings of student behavior, (b) provide a direct measure of children's required behavioral adjustments to teachers and peers, and (c) evaluate students' normative levels of performance regarding their adjustment to teachers and peers.

The first measure is a record of a child's Academic Engaged Time (AET). An observer using a stopwatch records the total amount of time a child is engaged on an instructional task. The second measure is a rating of Peer Social Behavior (PSB) during playground interaction, using a ten-second-interval recording system. Coders are asked to record children's social behavior according to five categories: social engagement, participation, parallel play, alone, and no code. Each observation is scheduled for a total of 15 minutes and the recording of behavior on the two measures occurs across two days. A significant amount of training is required for observers of Stage 3 measures. Quizzes and videotaped practice observations are used to train observers to reach what is called an "expert" criterion.

Scale Development
The development of the SSBD and the collection of data to support its technical adequacy involved five years of research. Instrument development and preliminary testing were conducted during an initial phase. Extensive research on the reliability and validity of the measures was done in a secondary phase. Normative studies were conducted in the latter phase. A technical development section of more than 50 pages is included in the manual. Interested readers should consult this source for the complete report on all research and development activities.

Scores Produced

Stage 1 produces no scores, as teachers rank-order children on internalizing and externalizing dimensions of problem behavior. Stage 2 comprises two indexes of behavior, the CEI and the CFI. The CEI is scored by adding the number of items endorsed. The range of scores on the CEI is from 0 to 35. The CFI comprises two scales: Adaptive Behavior and Maladaptive Behavior. These two scales use a 1-to-5 continuous rating scale (1, never; 3, sometimes; 5, frequently). Although these anchor points are not defined further, teachers are encouraged to choose any number between 1 and 5 that best describes the frequency of a child's behavior during the preceding six months. In addition, teachers are instructed to endorse items that they may not have witnessed but that can be confirmed by a reliable source. Total subscale scores are computed by summing the points assigned to all items within the subscale. The ranges of scores are from 12 to 60 for the Adaptive scale and from 11 to 55 for the Maladaptive scale.
In Stage 3, the score for the AET measure is calculated by recording the total seconds a child is engaged in instructional tasks. An average of the two observations is computed and compared, to separate norm scores by group (normal, externalizers, internalizers, and combined) and by grade (1–3, 4–6).

Scores for the five social behavior categories on the PSB measure are transferred to an observation summary sheet. Summary scores are reported as a percentage of intervals observed. Both individual code categories and summary scores (composed of code-category combinations) are reported on this summary form.

Decision Rules

Decision rules for determining whether a child continues to progress through Stages 1, 2, and 3 are described thoroughly in the administration manual for each level. Straightforward criteria are presented separately for children rated on the externalizing and internalizing dimensions of problem behavior. Flowcharts are presented in the manual, with separate cutoff scores for all scales (by behavioral dimension). Stage 3 decision rules for the PSB scale are provided by grade level. The authors suggest that an additional level of evaluation can be conducted at Stage 3 by collecting comparable observation data on a same-gender, nonreferred peer. In this manner, peer-referenced norms can be created for specific classrooms.

Additional Score Comparisons
Means, standard deviations, and standard errors for Stage 2 and Stage 3 are presented in the administration manual. Additional conversions of raw scores to \( T \)-scores and percentile ranks are also available for these measures. All norm tables are provided separately for children initially ranked by teachers on externalizing and internalizing dimensions. Norms are presented separately for males and females for AET and PSB measures. Finally, norm tables are provided for nonranked students from the standardization sample for Stage 3 measures.

**Norms**

The SSBD was normed on a national sample of children enrolled in general education classrooms from 18 school districts across eight states. Stage 2 included 4,463 students, with 72 percent of the sample coming from three states (Kentucky, Oregon, and Utah). Stage 3 observations were conducted on 1,275 students, with 70 percent of the sample drawn from the same three states. No additional information is provided on the gender, age, grade distribution, or educational classifications of subjects participating in the national sample. Demographic data are included for only 66 percent of the standardization sample. In addition, demographic data are reported only by the total enrollment of participating school districts, not by the actual students included in the standardization sample.

**Reliability**
Test–retest, internal-consistency, and interrater reliability data were collected during both phases of development of the SSBD. Data on these reliabilities are summarized here.

**Test–Retest Reliability**  For Stage 1 measures (teacher rankings), stability of behavioral classifications (internalizing, externalizing) was calculated during early tests of the SSBD. In a study involving 168 students, 78 percent of the students were classified in the same category on two occasions.

For Stage 2 measures (ratings on the CEI and CFI), short-term stability coefficients are provided. Reported correlations are in the .80s; however, the authors warn about possible inflation of these values because categories of students were collapsed for data analysis. No test–retest data for the Stage 3 (observational) measures are provided in the technical manual.

**Internal-Consistency Reliability**  Data related to the internal consistency of Stage 2 measures are reported. Reliability coefficients for the CFI Adaptive Behavior scale range from .85 to .94 and for the Maladaptive scale, from .82 to .92. Item–total correlations are provided for the scales of the CFI. Coefficients obtained in early testing of the measures were lower than those obtained during the standardization/validation studies.

**Interrater Agreement**  The authors made extensive use of interrater agreement data in the development of the Stage 1 measures. Rankings of students by two teachers, or by a teacher and a teacher's aide, on early versions of the SSBD yielded coefficients ranging from .60 to .94 for externalizing students and from .35 to .72 for internalizing students. For the final version (after group membership definitions were clarified), coefficients obtained by eight pairs of raters were .89 to .94 for externalizers and .82 to .90 for internalizers.
Stage 3 measures by trained observers have repeatedly resulted in high interrater agreement. The authors state that "interrater agreement levels have not been as yet established for Stage 2 measures" (H. M. Walker & Severson, 1992, p. 34).

Overall, the reliabilities of the SSBD measures are adequately supported by appropriate data from multiple-trial testing and validation studies. Some gaps in evidence are noticeable, and the potential user of the SSBD is not provided with comprehensive reliability data for all measures and for the full range of ages for which the SSBD is intended (grades 1–6). Because some portions of the device are age-sensitive (such as Stage 3), reliability of these measures for these ages would be better substantiated if supporting data were made available.

Validity

Content Validity  Publication of the SSBD was preceded by extensive development and research procedures. The selected behavioral items and the internalizing and externalizing dimensions have undergone a thorough testing, refinement, and validation process that has included both general education and special education teachers. For the Stage 3 observation codes, a previous observation code developed by H. M. Walker, Hops, and Greenwood (1984) served as a model. The authors provide adequate data to support their claim that these observation codes have been used and validated across numerous school settings.

Criterion-Related Validity  Two types of criterion-related validity (concurrent and predictive) are referred to in the technical manual. Support for concurrent validity is
provided by a series of correlational studies conducted during the instrument-
development phase and during the research phase of the development process. In one
such study, correlation coefficients between the SSBD–CFI scales and the CBCL were
calculated. On two occasions, correlations between the CBCL externalizing and the
SSBD Adaptive Behavior scale were -.63 and -.68, and for the Maladaptive Behavior
scale, .81 and .77 (compared at two different times). All correlations were significant at
the .001 confidence level.

In addition, Stage 2 measures were correlated with the Walker–McConnell Scale
of Social Competence and School Adjustment and the Classroom Adjustment Code.
Total score correlations were computed between the W-M and the CEI, the Adaptive
Behavior and the Maladaptive Behavior scales. All three comparisons were found to be
statistically significant at the .001 confidence level (correlations were -.57, .79, and -.44,
respectively). Correlations obtained between the CAC scales of on-task and unacceptable
behavior with the CEI were -.45 and .15, respectively. The Adaptive Behavior and
Maladaptive Behavior scales correlated in the low to moderate range in expected
directions with the CAC measures.

To assess the predictive validity of the SSBD, a total of 155 students (grades 1–5)
were assessed on all measures and reassessed one year later by different raters and
observers. Students initially rated as externalizers were found to be rated in the top three
externalizers one year later, 69 percent of the time. Internalizers were rated among the top
three at a lower level one year later, 52 percent of the time. Correlations between first-
year and second-year scores on Stage 2 measures for combined internalizing and
externalizing groups ranged from .32 (CEI) to .70 (Maladaptive Behavior scale). Overall,
correlations between Time 1 and Time 2 were in the low to moderate range. In addition, the criterion of teacher ratings at different times, which was used to suggest predictive validity, seems more appropriate for evidence of long-term stability. It might have served consumers better if an independent standard of comparison had been included.

**Construct Validity** To establish construct validity, the authors present data explaining factor-analytic and discriminant-function analyses conducted on the SSBD. Evidence of construct validity in the form of convergent and divergent correlational studies is also presented. Correlations between the CBCL internalizing and externalizing dimensions and Stage 3 (AET and PSB scales) were calculated. Stage 3 AET correlated -.42 with the CBCL at a .01 confidence level, and the PSB produced two significant correlations with the CBCL: negative social interaction and positive social interaction, with correlations of .29 and -.35, respectively. No SSBD observation measures were found to be significantly correlated with the CBCL internalizing scales.

In another study, a total of 66 students from Washington enrolled in grades 1, 3, and 5 were assessed on the SSBD Stage 2 measures. Additional instruments were used to collect information regarding the children's status on sociometrics, direct observation, school records, and social-skills ratings. The authors created a "deviance index" (p. 59) by combining selected variables from within the SSBD and from other measures judged to be best-evidence variables for each group. Then, Stage 1 rankings of the students were compared with the deviance index. Correlations between Stage 1 rankings and these multiple instruments for internalizers and externalizers were .71 and .76 (p < .001); these students were correctly identified most of the time (82 percent and 73 percent, respectively).
The authors also describe a study in which a total of 40 general education teachers, who had been assigned 54 children identified previously as having severe emotional disorders (45 externalizers and 9 internalizers), completed the SSBD. The teachers, having no prior knowledge of the students' educational history, were able to identify all 54 students by using the SSBD gating procedures. Overall, these correlations and classifications appear to be in the moderate to high range and to provide support for the SSBD as a measure of school-adjustment problems.

**Discriminant Validity** Discriminant validity of the SSBD system and its component measures is very important to the authors, given that one of the main constructs of the SSBD is the notion of differentiating children on a bipolar continuum of externalizing and internalizing problem behaviors. In prevalidation studies of the SSBD, almost 90 percent of students were correctly classified as internalizers or externalizers using Stage 2 and Stage 3 measures.

In one validation study, 170 teachers (for grades 1 to 5) completed Stage 1 and Stage 2 measures for students enrolled in classrooms in Oregon. Analysis of variance (ANOVA) on group differences for Stage 2 measures were highly significant. Post hoc T-tests indicated that all possible combinations of comparisons among three groups (internalizers, externalizers, and controls) were significant at the .01 confidence level.

In a second study conducted to replicate the first, a total of 40 general education teachers from the state of Washington completed Stage 1 and Stage 2 measures for 270 students. All post hoc T-tests identified mean differences among all comparisons of these groups, which were significant at the .05 level of confidence. These two studies suggest
that the SSBD has sufficient discriminant power to differentiate among these three
groups.

An additional study examined a more clinical sample of 106 students who were
assigned to a residential facility serving severely disturbed or abused children in
kindergarten through grade 12. Students were enrolled in four different programs that
included two residential and two day-treatment programs: secure unit \( (N = 17) \),
residential \( (N = 52) \), day treatment \( (N = 20) \), and community based \( (N = 17) \). Separate
one-way ANOVAs were conducted for the CEI, the Maladaptive Behavior scale, and the
Adaptive Behavior scale. Results indicated significant differences among the four
categories of students on the CEI. No statistical differences were observed on the
Maladaptive Behavior or Adaptive Behavior rating scale. Post hoc \( T \)-tests identified
differences between the secure unit and the residential students. No other mean
differences were found to be significant. This may suggest that, although the SSBD has
sufficient discriminant power to differentiate classification between clinical and general
population samples, further specificity within clinical populations is not as well
established.

Taken together, discriminant-validity studies demonstrate that the SSBD is
effective at distinguishing internalizing and externalizing students from the larger
population of students without behavioral and adjustment problems. Students classified
as externalizers on the SSBD were found to engage in less adaptive behavior and more
maladaptive behavior than controls and internalizers. In addition, this group spent less
time academically engaged and more time in negative interactions than did internalizers
and control children. Children rated as internalizers on the SSBD were found to engage in
less adaptive behavior and more maladaptive behavior and to spend more time alone than control students.

Summary

The SSBD is a well-conceived and well-researched instrument for screening and identifying children in need of further assessment for behavioral disorders. In fact, the SSBD has been nominated as an example of an effective instrument by the Program Effectiveness Panel of the U.S. Department of Education. However, to date, the SSBD has been used largely in a research capacity that has consistently produced discrete evidence of reliability and validity. Normative and demographic data and defining characteristics of the students who participated in the standardization sample are difficult to discern. The authors provide substantial evidence of the SSBD scales' ability to differentiate between students exhibiting internalizing or externalizing problem behavior and well-adjusted students. Although reported reliability and validity studies appear adequate, there are gaps in some areas (such as interrater reliability and grade-specific reliabilities). These gaps may ultimately restrict the types of interpretations that can be made from the results of the SSBD instrument. The essential strength of the SSBD is the conceptual framework of multiple-gating procedures, which serves to organize and standardize what teachers and practitioners have been doing informally for years.