Investigation of a Multiple-Gated Screening System for Preschool Behavior Problems

Edward G. Feil and Wesley C. Becker

ABSTRACT

The problem of behavioral disorders in preschool-age children is ever increasing. With the exponential rise in the utilization of child care compounded by growing social problems such as poverty and child abuse, methods for the early identification and remediation of behavioral disorders in preschool children are needed. In the May 1993 issue of Behavioral Disorders, Sinclair, Del'Homme, and Gonzalez reported a pilot study using the Walker/Severson Systematic Screening for Behavior Disorders (SSBD) with preschool children. While their results were encouraging, they found that changes were needed to make the SSBD more appropriate for the preschool population. The present research extensively revised the SSBD for preschool children. The revision consists of three hierarchical stages of increasingly time-consuming methodologies: (a) teacher rankings, (b) teacher ratings, and (c) direct behavioral observations. Subjects for this study were 121 children, aged 3 to 6 years old, enrolled in several typical and specialized preschools (e.g., programs for children with behavior problems). The Behar Preschool Behavior Questionnaire and the Conners Teacher Rating Scale were included to examine concurrent validity. The results show significant reliability and validity coefficients. The screening procedures select those children with emotional/behavioral problems accurately.

This research evaluates a screening system for emotional/behavioral problems among preschool children using an adaptation of the Walker/Severson Systematic Screening for Behavior Disorders (SSBD) procedures (Walker & Severson, 1990). The (SSBD) is a multiple-gated screening instrument for use with elementary schoolage children. This adaptation is based on empirical findings from past studies, expert judgment, and reliability testing. Since a closely related study was published in the May 1993 issue of *Behavioral Disorders* — Systematic Screening for Preschool Behavioral Disorders (Sinclair, Del'Homme, & Gonzalez, 1993) — the authors will attempt to reduce overlap by focusing on their procedures and outcomes and how they differ from the Sinclair et al. report.

As Sinclair et al. (1993) note, the prevalence of behavioral disorders in young children is estimated to be between 2 and 10%, but less than 1% are typically identified for special services. Walker and Severson's (1990) SSBD was designed to help identify children with behavior problems described as externalizing (aggressive, hyperactive, noncompliant, antisocial, etc.) and internalizing (shy, timid, depressed, isolated from peers, etc.). Identification is achieved through a three-staged procedure. Stage I uses teacher nominations of children with externalizing and internalizing behavior; Stage II uses teacher ratings of the children most troubled; and Stage III uses direct observations of academic and social behavior. This multiple-gated procedure was devised to be cost effective by reducing the number of children on whom ratings and observations are made.

In revising the schoolaged-based SSBD, it was necessary to consider the appropriateness of the various procedures used in the SSBD for younger children. Sinclair et al. (1993) used the SSBD intact except that the direct observations of academic engaged time were eliminated and the direct observations of peer social behavior during free play in the classroom and on the playground were doubled to four 10-minute sessions. In addition, in Stage I the teacher were asked to nominate and rank only 7 externalizers and 7 internalizers (out of classes of 15) rather than 10 of each. The three top ranked externalizers and internalizers were followed up with Stage II rating scales and Stage III observations. Also, the cut-off cri-

This research was supported by the U.S. Department of Education, Research in Education of the Handicapped Program: Student-Initiated Projects (84.023B) Competition, Office of Special Education and Rehabilitative Services, Grant No. H023B10004. The views expressed are not necessarily those of the USDOE. Requests for reprints and/or a technical report should be sent to the first author.

teria for defining problem children were adjusted to take into account the behavior of younger and older children (e.g., younger children show more parallel play). The scores of the 36 selected children were then compared with the norm data for school-level externalizers and internalizers reported by Walker and Severson (1990).

The present adaptation of the SSBD had the preschool teachers nominate and rank only 5 externalizers and 5 internalizers (per classroom) during Stage I. All of the children in the preschools were followed up through Stages II and III so that appropriate item analyses, reliabilities, and construct and criterion validities could be examined. The Stage II behavior checklist measures were substantially modified to make them appropriate for ratings of preschool level children. For example, approximately half of the occurrence/nonoccurrence items were changed to a five-point scale (to get a better report on frequency and/or intensity of problems), items regarding academics were omitted due to their inapplicability to preschool activities, and wording was changed to make the items more easily answered. Stage III observations used a slightly modified peer social behavior (PSB) code. Observations of academic engaged time were replaced with structured-activity engaged time (SAET). This code assessed the duration of engaged time during such structured activities such as storytime and arts 'n' crafts.

METHOD

Subjects

Subjects were 121 children between 3 and 6 years of age, enrolled in five preschools. The preschools included programs for (a) typical children, (b) children identified with serious emotional/behavioral disorders, (c) children with developmental and language delays, and (d) children of families with low incomes (i.e., Head Start children). The study included 17 teachers and assistant teachers of the children. In the second data collection period (March - May, 1992)16 subjects withdrew from their respective preschools, so the final number fell to 105 preschoolers. Table 1 summarizes the descriptive statistics for the sample.

TABLE 1

Number of Subjects by Age Showing Gender and Eligibility Status

	Age at Beginning of Study (Years)				Total
	3	4	5	6	No. (%)
Gender					
Female	4	31	14	1	50 (41)
Male	6	34	23	8	71 (59)
Special Education Eligibility Status					
SED/BD	0	3	4	1	8 (7)
Developmental delay	1	0	1	0	2 (2)
Language delay	2	8	5	0	15 (12)
At-risk (Head Start)	0	32	11	0	43 (43)
Other (e.g., ESL)	1	2	0	0	3 (2)
None	6	20	16	8	50 (41)
Total	10 (8)	65 (54)	37 (31)	9 (7)	121 (100)

Measures and Procedures

The Preschool Screening for Behavior Problems. The PSBP is a three-stage, multiple-gating procedure to screen for behavioral disorders among preschool children 3 to 5 years old. Stage I is based on teachers rankings of their students on externalizing and internalizing behavior dimensions. Teachers were asked to list the 5 children who best exemplified externalizing characteristics and the 5 children who best exemplified internalizing characteristics. The two lists were mutually exclusive; therefore, a child could only be put on one list. Then the teacher ranked each list from most characteristic to least characteristic of the externaliz-

ing or internalizing dimension. As noted above, this procedure was modified from the SSBD procedure because of smaller class sizes.

Stage II is a behavior checklist consisting of four measures: Critical Events Index Parts A and B (CEI-PART A and CEI-PART B, respectively) and Adaptive and Maladaptive Behavior Indexes. Stage II of the PSBP differs substantially from the Walker/Severson SSBD. The SSBD Critical Events Index consists of 33 occurrence/nonoccurrence items. Therefore, a teacher will check an item if a child has exhibited the behavior. Since all preschool children will exhibit problem behaviors at one time or another (Campbell, 1990; Paget, 1990), the frequency and intensity of the behaviors were most likely the important discriminative features. Consequently, 17 occurrence/nonoccurrence items of the SSBD were converted to frequency ratings. The CEI-PART A contains 13 occurrence/nonoccurrence items and the CEI-PART B consists of 17 five-point Likert response scales for sensitivity to frequency and intensity. Items converted to a five-point frequency ratings were: steals, sets fires, vomits after eating, tantrums, physically assaults an adult, physically aggressive with other children, damages property, has nightmares, exhibits inappropriate sexual behaviors, suddenly cries, physical complaints, ignores teacher warnings, makes lewd gestures, swears, is teased/neglected by peers, enuretic, and encopretic.

The Adaptive Behavior Scale contains 8 items representing overall prosocial behavior (e.g., cooperation and positive social interactions). Four items were omitted from the SSBD Adaptive Behavior Scale due to their developmental inappropriateness to preschool children regarding cognitive skills or academic work: (a) considerate of others' feelings, (b) produces acceptable work, (c) is socially perceptive, and (d) does seat-work assignments. The Maladaptive Behavior Scale consists of 10 items representing overall anti- or nonsocial behavior (e.g., defies teacher requests and creates disturbance). Two items were combined from the SSBD Maladaptive Behavior Scale — "Uses coercive tactics to force the submission of peers; manipulates, threatens, etc." and "Manipulates other children and/or situation to get his/her own way" were combined as "Manipulates or threatens other children to follow him/her" which is a better representation of preschool behavior.

Since preschool teachers frequently do not have a bachelors degree, the wording was changed to ease responding. Further, items regarding academics were omitted because of their inapplicability to most preschool curricula.

Stage III measures consist of direct observations of a child's structured-activity engaged time (SAET) in the classroom and peer social behavior (PSB) in the classroom and on the playground. SAET was a duration recording of the child's attending to a teacher-led, structured-group activity such as storytime or gymnastics. SAET was defined by the observation codes for (a) attending to the teacher, (b) making appropriate motor responses (e.g., following directions), and (c) asking for assistance in an appropriate manner. PSB was a record of the quality, level, and distribution of a child's social behavior during free play settings. The PSB code categories included social engagement (SE), participation (P), parallel play (PLP), solitary play (SP), alone (A), adult (Adlt), and no codeable response (N). This adaptation added the solitary play code due to the more frequent occurrence of solitary play with preschoolers. The first two categories, social engagement and participation, were coded as a qualitative judgment of either positive (+) or negative (-). Parallel play, solitary play, alone, adult, and no codeable response were simply recorded as occurrence/nonoccurrence. The observation intervals were 10 seconds in duration. The rater recorded the type of behavior hierarchically. So, if two behaviors occurred during the same interval, the behavior higher in the hierarchy was recorded. For example, if positive social engagement (SE+) and solitary play (SP) occurred in the same interval, *only* social engagement (SE+) was recorded. The children were each observed for 40 minutes — 10 minutes for structured-activity engaged time and peer social behavior, respectively, on two separate occasions. If the total time per observation code was under 20 minutes, a third observation was conducted to bring the time up to 20 minutes.

Behar Preschool Behavior Questionnaire. The Behar Questionnaire was developed as a screening instrument for identifying preschoolers showing potential behavioral disorders (Behar & Stringfield, 1974). Behar and Stringfield found that the scale significantly differentiated the groups labeled *normal* ($N = 496$) and *disturbed* ($N = 102$). In the present study, the

teachers and assistant teachers completed the Behar Preschool Behavior Questionnaire at the same time they completed the Stage II questionnaire.

Conners Teacher Rating Scale. The Conners Questionnaire is widely used to assess children ages 3 to 17 for behavioral problems, especially for attention deficit hyperactivity disorder (ADHD). The Conners has seven subscales: hyperactivity, conduct problems, emotional indulgence, anxious-passive, asocial, daydream-inattention, and hyperactivity index. Test-retest reliabilities of the Conners' questionnaire range from .70 to .90 from a sample of 578 children ages 3 to 17 years of age (Goyette, Conners, & Ulrich, 1978). Over a year, stability was reduced to a range of .33 to .55 (Conners, 1989). The concurrent validity of the Conners was tested with a comparison to parent ratings. Correlation coefficients ranged from .33 for the conduct problem factor to .45 for the inattentive-passive factor (Goyette et al., 1978). Conners (1989) states that "generalized hyperkinesis at 7 years of age was highly predictive of hyperactivity at 10 years" although no specific data are cited. The teachers and assistant teachers in this study completed the Conners Teacher Rating Scales at the same time they completed the Stage II questionnaire.

Data Collection Schedule

Subjects were administered the PSBP in the Fall (November-December) and Spring (March-May) of 1991-1992. No gating criteria (i.e., exclusion of normal subjects) were used in the administration of the Stage III observational data since data on all children were needed for reliability and validity statistical procedures such as interrater reliability and discriminant function analysis. Stage III observations were completed in the remaining month(s) by a group of 17 trained personnel. Training sessions were conducted until each observer could code videotaped scenes at a 90% accuracy (average training time was 6 hours).

Data Analysis

Item aggregation into scales. The first task in data analysis was to aggregate the individual Stage II rating items and Stage III observational categories into functional scales. The Stage II items were correlated with total scores. Items with low correlations with scale scores were eliminated and adjusted total scores were used in subsequent analyses. The teacher ratings (as opposed to the assistant teacher ratings) were used for all comparisons with the exception of interrater agreement (see Reliability section). Two items from CEI-PART A (i.e., Significant weight change and Loss of interest) and two items from CEI-PART B (i.e., Vomiting and Nightmares) were eliminated due to their low correlation with their respective total scale score. The remaining item-total correlations for all Stage II measures ranged from .23 to .89 with a median of .67.

The Stage III observational data means were compared by session in order to decide whether to aggregate the sessions into totals per code category. There were few session three observations because subjects usually met the 20-minute requirement by the end of session two. Since the percentages of the coded behaviors did not change appreciably over sessions, aggregated totals were used for further analyses.

Once sessions were combined, peer social behavior code categories were correlated with Stage II ratings, Behar, and the Conners to help decide on aggregation into PSB-Pos and PSB-Neg categories (see Table 2). The highest correlating positive and negative code categories, respectively, were combined to an overall positive (PSB-Pos) and negative (PSB-Neg) aggregated score. The social engagement-positive, participation-positive, and parallel play categories were the highest positive code categories, and social engagement-negative, solitary play, and adult categories were the highest correlating negative code categories. The adult category (i.e., the amount of time the child interacts with an adult) was significantly correlated with all measures of deviance.

TABLE 2

Peer Social Behavior Code Categories Pearson Correlation to Stage II, Behar, and Conners scales

PSB Code Categories	CEI-Part A	CEI-Part B	Adaptive	Mal-adaptive	Behar	Conners-Hyper	Conners-Inattention
Social engagement-positive[1]	-.13	-.12	.24*	-.09	-.12	-.14	-.25*
Social engagement-negative[2]	.12	.29**	-.21	.18	.23*	.18	.03
Participation-positive[1]	-.06	-.06	.12	-.11	-.07	-.12	-.06
Participation-negative	-.05	-.03	-.00	-.07	.01	-.00	.01
Parallel play[1]	-.06	-.10	-.14	-.02	-.07	0.01	-.04
Solitary play[2]	.08	.09	-.08	.09	.14	.13	.31**
Alone	.17	.06	.09	-.06	-.01	-.03	.07
Adult	.45***	.51***	-.25**	.27**	.32***	.25*	.32***

*$p < .05$
**$p < .01$
***$p < .001$
[1]Included in the global PSB positive scale (PSB-Pos)
[2]Included in the global PSB negative scale (PSB-Neg)

RESULTS

Reliability Findings

Interrater Reliability of Teacher-Assistant Teacher Ratings

Externalizers. A correlation of teacher and assistant teacher ranks on the externalizing list (coded 10 for highest ranked, 6 for lowest ranked, and 5 for not ranked) resulted in a highly significant coefficient of .71 (see Table 3). A correlation of this magnitude is equal to the correlations obtained between Stage II measures (i.e., Adaptive, Maladaptive, Behar, and Conners).

TABLE 3

PSBP Stage I, PSBP Stage II, Behar, and Conners Scales' Interrater Reliability Coefficients

Scale	*N*	Teacher A Mean	Teacher A SD	Teacher B Mean	Teacher B SD	*r*
Externalizer ranking	99	6.04	1.71	5.97	1.64	.71*
Internalizer ranking	99	3.84	1.74	3.91	1.73	.55*
Critical events-A	98	.71	1.24	.73	1.21	.79*
Critical events-B	98	20.93	7.62	19.71	6.72	.83*
Adaptive	98	32.03	6.80	31.40	6.74	.60*
Maladaptive	98	19.85	8.19	19.85	8.55	.59*
Behar	98	12.37	9.17	11.31	9.84	.61*
Conners hyperactive	98	11.42	10.43	10.86	10.71	.59*
Conners inattention	98	1.84	1.93	1.77	1.99	.56*

*$p < .001$

Internalizers. The interrater reliability correlation for the internalizing list (coded 0 for highest ranked, 4 for lowest ranked, and 5 for not ranked) resulted in a lower, yet highly significant, coefficient of .55. This correlation confirms the frequent findings that it is more difficult to get rater agreement on internalizing than on externalizing behavior (Achenbach, 1982).

CEI PARTS A and B. Comparing the teachers and assistant teachers' scale scores resulted in highly significant correlations of .79 and .83 (see Table 3). Interrater reliability was probably higher for these scales than the more global rating scales because they deal with more specific behaviors.

Adaptive and Maladaptive. Correlations of teachers' and assistant teachers' scale scores resulted in highly significant correlations of .60 and .59 (see Table 3). While these were not as great as in the Critical Events Indexes, these correlations were of sufficient magnitude to be useful. These coefficients were equal to the Behar and Conners, two published measures used for the identification of preschool behavior problems.

Behar and Conners. Comparisons of rater scores gave highly significant correlations between .56 and .61 (see Table 3). These correlations were within expectations from past studies (Goyette et al., 1978).

Stage III observations — structured-activity engaged time and peer social behavior. The interrater reliabilities of the observational measures were calculated from a random sample of 20% of the observations. For both measures, extensive training was conducted in advance. The structured-activity engaged time (SAET) interrater reliability was based on the sum of the smaller scores across children divided by the larger scores (Hartman, 1977; Wolery, Bailey, & Sugai, 1988). This provides a percentage indicator of rater differences *weighted for length of observation* and resulted in a coefficient of .97 which indicates that on average the raters only differed by 3% in recording SAET.

The peer social behavior (PSB) interrater reliability was analyzed by interval agreement. For an agreement to occur, both raters would have to mark the same category and same valance if appropriate (i.e., positive or negative) in the identical time interval (Hartman, 1977; Kratochwill & Wetzel, 1977; Wolery et al., 1988). This procedure resulted in an average percentage agreement of .87 which is within acceptable limits for a screening device of this type (Salvia & Ysseldyke, 1988).

Test-Retest Reliability

Considering the externalizers and internalizers lists, correlations were computed between the teacher scores in the fall and spring data collection periods. The highly significant correlations were .64 and .49 for externalizers and internalizers, respectively (see Table 4). The teacher rankings were more stable for externalizers than internalizers, yet both coefficients were adequate for the initial stage of this screening system. Teacher fall and spring scores on the CEI-PARTS A and B, the Adaptive. and the Maladaptive scales were compared and resulted in highly significant correlations between .77 and .91 (see Table 4). The correlations were comparable to the Behar and Conners scales for test-retest correlations. These coefficients were quite acceptable for a screening instrument. Stage III measures were significant but of lower magnitude than other measures (see Table 4). A low correlation was expected from observational data since there is situation specific variability (Skiba, 1989). Correlations of fall and spring teacher ratings on the Behar and Conners scales were significant with coefficients between .61 and .82 (see Table 4).

Validity Findings

Content Validity

Content validity is the degree to which a measure tests the domain to be studied, which in this study is externalizing and internalizing behaviors. Content validity was deduced from three data sources: (a) empirical findings from past studies, (b) the judgments of a panel of experts, and (c) preschool teacher feedback. The formulation phase of this research occurred from October 1990 to June 1991. A draft of the PSBP was presented to a panel of experts and a few changes were implemented before any data were collected. An initial pilot study was completed in the Spring of 1991 to assess teacher perceptions of the study, resulting in a positive review by the two participating teachers.

TABLE 4

Test-Retest Reliability Correlations – Fall and Spring Data

	Fall (Nov. - Dec. 1991)			Spring (Mar. - May 1992)			
	N	Mean	SD	N	Mean	SD	*r*
Stage I							
Externalizer	121	6.04	1.70	105	6.04	1.71	.64**
Internalizer	121	3.97	1.67	105	3.84	1.74	.49**
Stage II							
CEI-PART A	121	.65	1.04	105	.71	1.21	.77**
CEI-PART B	121	19.43	7.49	105	20.77	7.53	.91**
Adaptive	121	31.91	7.18	105	31.69	6.84	.75**
Maladaptive	121	18.85	7.73	105	19.78	8.08	.79**
Stage III							
SAET	98	.90	.11	98	.86	.10	.20*
PSB-Positive	100	.77	.18	100	.77	.15	.22*
PSB-Negative	100	.17	.15	100	.17	.13	.22*
Rating Scales							
Behar	121	10.92	8.74	104	12.17	9.06	.79**
Conners-Hyper	120	10.95	10.53	105	11.29	10.22	.82**
Conners-Inattention	120	2.24	2.12	105	1.87	1.91	.61**

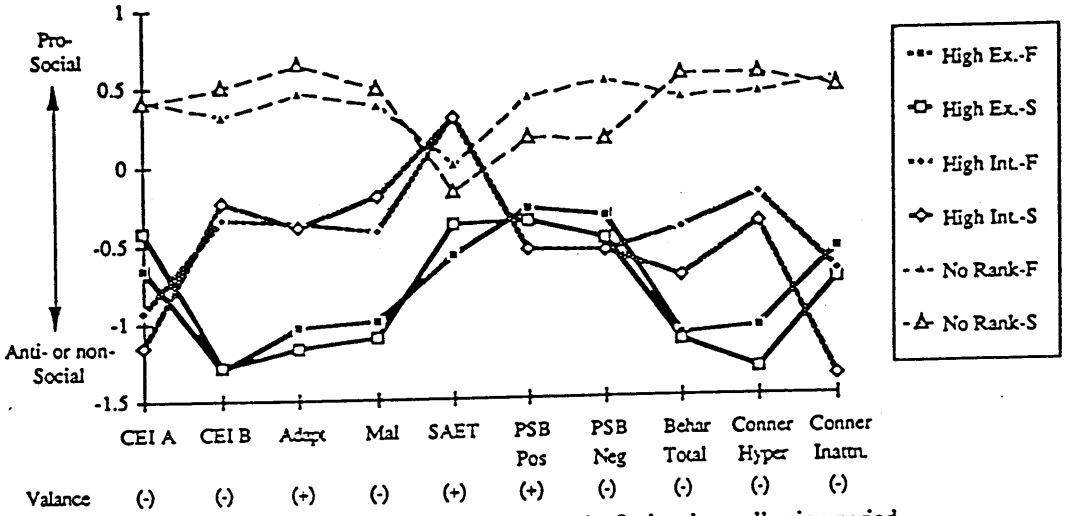
*$p < .05$
**$p < .001$

Discriminative Validity

Discriminative validity was examined first by graphing the standardized scores for the main-measures by group and through the use of a discriminant function analysis. The consistency across measures was examined by comparing the standardized scores ($M = 0$, $SD = 1$) of the children ranked highest on Stage I externalizer and internalizer dimensions, respectively, and children ranked as average (serving as a control comparison) across PSBP, Behar, and Conners measures for the Fall and Spring data collection periods (see Figure 1). The negatively oriented ranks of the CEI-PARTS A and B, Maladaptive, Behar, Conners, and peer social behavior-negative (PSB-Neg) scores were reversed to maintain continuity in valance.

The group means remained relatively equal from Fall to Spring, showing consistency over a 7-month period. The three groups were clearly delineated on the CEI-PART B, Adaptive, Maladaptive, and Conners-hyperactive scales. The externalizer group was clearly more deviant on these measures than the other two groups (see Figure 1). The CEI-PART A and Conners-inattention best differentiated internalizers from the no-rank group. While the observational measures taken individually (i.e., SAET, PSB-Pos, and PSB-Neg) did not discriminate as clearly between groups, each group had a distinctive profile. The externalizer group had low prosocial scores on all observational measures, whereas the internalizer group had the highest structured activity engaged time (SAET), yet the latter also had low prosocial peer social behavior (PSB). Both the externalizer and internalizer groups had relatively equivalent scores on the PSB measures, showing that both displayed social skill deficits.

A discriminant function using the general linear model estimates the accuracy of a set of dependent measures in predicting membership in a priori groupings. In this study, the groups were teacher recommendation of BD eligibility status (i.e., whether the teacher listed the child for further evaluation for BD status) and the dependent measures were the PSBP's Stage I (i.e., teacher rankings of externalizer and internalizer), Stage II (i.e., CEI-PARTS A and B, Adaptive, and Maladaptive), and Stage III (i.e., SAET, PSB-Pos, and PSB-Neg). The three highest ranking internalizers and externalizers, respectively, in each class were divided into groups for the discriminant analysis. The false positive error rates for both the externalizer and internalizer groups were very low, 3 and 0%, respectively, and the false negative error rates were 10 and 3%, respectively. A MANOVA test of the group means for the PSBP



F refers to the Fall data collection period and S refers to the Spring data collection period.

Number of subjects	Fall	Spring
Highest Ranked Externalizer	9	9
Highest Ranked Internalizer	9	8
No Rank (control comparison)	41	28

Measure abbreviation	Measure Name
CEI-A	Critical Events Index- Part A
CEI-B	Critical Events Index- Part B
Adapt	Adaptive Behavior
Mal	Maladaptive Behavior
SAET	Structured-Activity Engaged Time
PSB-Pos	Peer Social Behavior- Postive
PSB-Neg	Peer Social Behavior- Negative
Behar	Behar Preschool Behavior Questionnaire- All items
Conner Hyper	Conners Teacher Rating Scale- Hyperactive scale
Conner Inattn	Conners Teacher Rating Scale- Inattention scale

Figure 1. *Means of children ranked highest on the externalizing dimension, internalizing dimension, and nonranked peers on standard scores of major measures for Fall and Spring data collection periods.*

measures found a highly significant difference, $F(8,94) = 20.38$, $p < .001$). The discriminant function and MANOVA test show that the PSBP was satisfactorily accurate in predicting problem behaviors in preschoolers in this study.

Concurrent Validity

The concurrent validity of Stage II measures were examined through correlations with the Behar and Conners measures. These data, summarized in Table 5, show very good overall concurrent validity, with correlations ranging between .25 to .84. The Critical Events Index-B, Adaptive, and Maladaptive scales showed substantial concurrent validity.

DISCUSSION

These findings are promising. The interrater reliability coefficients of most PSBP measures are at least .80 which meets Salvia and Ysseldyke (1988) guidelines for a screening instrument. Good psychometric standards are attained despite the difficulties inherent in the assessment of young children (Martin, 1986). The consistency across measures illustrates the potential utility of the screening system, giving evidence that behavior problems may be identified accurately among preschool children. The discriminant and MANOVA functions show the high accuracy of the PSBP with low false positive and false negative error rates. The Sinclair et al. (1993) study shows the feasibility of using the SSBD (with minor changes) with preschool children. The present study suggests that a systematic revision of the SSBD for preschoolers can be more effective.

TABLE 5

Concurrent Validity — Pearson Correlations Between PSBP Stage II Measures and Published Rating Scale Measures

PSBP Stage II Measures	Published Rating Scale Measures		
	Behar Total	Conners Hyperactive	Conners Inattention
Critical Events Index-Part A	.39**	.25*	.55**
Critical Events Index-Part B	.82**	.77**	.51**
Adaptive	-.77**	-.77**	-.54**
Maladaptive	.80**	.84**	.52**

The number of subjects ranges from 98 to 100.
*$p < .01$
**$p < .001$

These issues need to be discussed: (a) the low correlations for observational measures, (b) the use of the PSBP finding in planning intervention, (c) cost effectiveness, (d) developmentally appropriate practice, and (e) next steps in research.

The reader should note that while the observational data have lower reliability and validity correlations than the rating measures, the data should not be discounted as invalid. Rating and observational measures both focus on the target child, yet the ratings also include the bias of the rater and ratee's relationship to the rater (Cairns & Green, 1989; Martin, 1986). Observational measures record the behavior directly with a lower amount of bias and filtering of information. However, observational measures are sensitive to ecological variables such as situation-dependent interactions and physical settings. Both kinds of data and analysis are probably important to understand behavioral disorders with its socially dependent nature.

The PSBP seems to be useful not only in the screening process but also as a guide for further assessments and interventions. The screening results taken together with other assessment instruments can confirm a preschoolage child's eligibility for special education services under the emotional/behavioral disorders category. *Treatment utility* is the degree to which assessment activities are shown to contribute to beneficial intervention outcomes (Hayes, Nelson, & Jarrett, 1987). The observational categories show the quality and degree of a child's interacting with peers (i.e., social engagement positive and negative, solitary play, alone). The observational results, both aggregate scores (i.e., PSB-Pos and PSB-Neg) and code scores (i.e., parallel play, social engagement), can be used with other observations (both quantitative and qualitative), interview, and rating form to help plan intervention strategies. The structured-activity engaged time gives a description of a child's attention to appropriate norms in structured settings. The PSBP observational measures can lead to appropriate and effective individualized intervention plans.

Presently, there are few effective and low-cost methods of screening for behavioral disorders in natural early childhood settings. Young children with mild-to-moderate learning or behavioral disorders are at the greatest risk of being overlooked with traditional developmental screening tests (Beare & Lynch, 1986). Martin's (1986) review of the literature on the assessment of the social and emotional functioning of preschool children shows that some technically and financially feasible screening instruments are available, but all require a lengthy administration if all children are screened. For most scales, there was an inverse relationship between the ease of administration and psychometric characteristics. Cost effective measures are clearly needed.

Developmentally appropriate practice (DAP) guidelines (Bredekamp, 1987) are gaining wide acceptance as the litmus test for the evaluation of assessment and curriculum procedures in many early childhood settings. DAP guidelines state that major decisions, such as assessment, should not be made on the basis of one developmental assessment or screening device but also should include observations by teachers. Assessment information should be used to plan curriculum, identify children with special needs, communicate with

parents, and evaluate program effectiveness (Bredekamp, 1987). The PSBP conforms to DAP standards by emphasizing teacher observations and recognizing developmental differences between preschool and schoolage children. These differences were deduced from three data sources: empirical findings from past studies, the judgments of a panel of experts, and preschool teacher feedback.

Further longitudinal study is needed to set decision rules for the gating criteria. Currently, plans for future studies with a greater and more diverse number of subjects are being made including a longitudinal follow-up on the subjects as well as broadening the sample to other locations. The PSBP has properties which could make it a very useful and cost effective screening and assessment instrument.

REFERENCES

Achenbach, T. M. (1982). Assessment and taxonomy of children's behavior disorders. In E. B. Lahey & A. E. Kazdin (Eds.), *Advances in clinical child psychology.* New York: Plenum.

Beare, P. L. & Lynch, E. C. (1986). Underidentification of preschool children at risk for behavioral disorders. *Behavioral Disorders, 11,* 177-183.

Behar, L., & Stringfield, S. (1974). *Manual for the preschool behavior questionnaire.* Durham, NC: Behar.

Bredekamp, S. (Ed.) (1987). *Developmentally appropriate practice in early childhood programs serving children from birth to age 8.* Washington, DC: National Association for the Education of Young Children.

Cairns, R. B. & Green, J. A., (1989). How to assess personality and social patterns: Observations or ratings. In R. B. Cairns (Ed.), *The analysis of social interaction: Methods, issues, and illustrations.* Hillsdale, NJ: Erlbaum.

Campbell, S. B. (1990). *Behavior problems in preschool children.* New York: Guilford.

Conners, C. K. (1989). *Manual for the Conners rating scales.* North Tonawanda, NY: Multi-Health Systems.

Goyette, C. H., Conners, C. K., & Ulrich, R. F. (1978). Normative data on revised Conners parent and teacher rating scales. *Journal of Abnormal Child Psychology, 6,* 221-236.

Hartman, D. P. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis, 10,* 103-116.

Hayes, S. C., Nelson, R. O., and Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist, 42,* 963-974.

Kratochwill, T. R., & Wetzel, R. J., (1977). Observer agreement, credibility, and judgment: Some considerations in presenting observer agreement data. *Journal of Applied Behavior Analysis, 10,* 133-139.

Martin, R. P. (1986). Assessment of the social and emotional functioning of preschool children. *School Psychology Review, 15,* 216-232.

Paget, K. D. (1990). Best practices in the assessment of competence in preschoolage children. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology-II.* Washington, DC: National Association of School Psychologists.

Salvia, J., & Ysseldyke, J. E. (1988). *Assessment in special and remedial education.* Boston: Houghton Mifflin.

Sinclair, E., Del'Homme, M., Gonzalez, M. (1993). Systematic screening for preschool behavioral disorders. *Behavioral Disorders, 18,*.177-188.

Skiba, R. J. (1989). The importance of construct validity: Alternative models for the assessment of behavioral disorders. *Behavioral Disorders, 14,* 175-185.

Walker, H. M. & Severson, H. H. (1990). *Systematic screening for behavior disorders user's guide and administration manual.* Longmont, CO: Sopris West.

Wolery, M., Bailey, D. B., & Sugai, G. M., (1988). *Effective teaching: Principles and procedures of applied behavior analysis with exceptional students.* Boston: Allyn & Bacon.

Edward G. Feil, Doctoral Student, School Psychology Program, Department of Special Education and Rehabilitation, College of Education, University of Oregon, Eugene, OR 97403

Wesley C. Becker, Professor, School Psychology Program, Department of Special Education and Rehabilitation, College of Education, University of Oregon, Eugene, OR 97403