



The Early Screening Project for Young Children with Behavior Problems

EDWARD G. FEIL, HILL M. WALKER, AND HERBERT H. SEVERSON

Several investigators have shown that behavior problems often have their origins in early childhood (Fischer, Rolf, Hasazi, & Cummings, 1984; Lerner, Inui, Trupin, & Douglas, 1985; Patterson, Reid, & Dishion, 1992; Waxler, 1993). In addition, it has been shown that problem behaviors tend to be stable over time for preschoolers (Kohn, 1977) and are predictive of future learning problems in third grade (Fischer et al., 1984; Patterson, DeBaryshe, & Ramsey, 1992). Without early intervention, children with behavior disorders are at risk for increasing levels of long-term social maladjustment (Patterson et al., 1989), and remediation becomes increasingly more difficult with age (Bower, 1981). Compensatory programs, such as Head Start, are based in part upon the notion that systematic early intervention can reduce later problems (Lerner et al., 1985; Rolf & Haazi, 1977).

The early identification and remediation of emotional and behavioral disorders is a high priority for most educators and is based on the assumption that behavioral and academic problems can be averted through early detection, prevention, and intervention. Presently, there are few assessment instruments for screening behavior problems in preschool children, and, for the most part, they are inefficient (Martin, 1986). This research seeks to meet this need through development of a preschool adaptation

Behavioral problems among preschool-age children have become a priority for early childhood educators. Factors such as the exponential rise in childcare utilization and increasing poverty and incidences of child abuse have made the establishment of procedures for the early identification and remediation of problem behaviors among preschool children imperative. Screening is an initial step in identifying those who exhibit a certain disorder, problem, or disability. The goal of the current research was to study the effectiveness of a functional screening and identification system for behavior problems among preschool children ages 3 to 5 years. This research focused on the development of the Early Screening Project (ESP), a multiple-gated screening system adapted from the Systematic Screening for Behavior Disorders. Data were collected over 32 months, from September 1991 through June 1994. Participants were 2,853 children, ages 3 to 6 years, enrolled in typical and special education preschool and kindergarten classrooms in Oregon, California, Texas, Utah, Kentucky, New Hampshire, Nebraska, and Louisiana. Concurrent validity was assessed by administering the Behar Preschool Behavior Questionnaire, the Conners' Teacher Rating Scale, and Achenbach's Child Behavior Checklist measures. Results demonstrated significant reliability and validity coefficients. The ESP procedure provides reliable, cost-effective, and accurate screening of preschool-age children to facilitate early remediation of behavior problems.

of the Systematic Screening for Behavior Disorders (SSBD; Walker & Severson, 1990). The adaptation is called the Early Screening Project (ESP; Walker, Severson, & Feil, 1995).

Stability of Behavior Problems Over Time

Despite developmental changes, longitudinal follow-up studies show that the core of antisocial behavior patterns changes very little over time (Patterson et al., 1992). For many children, behavior problems follow a progression, such as the following: (a) disobedience in the home, (b) temper tantrums, and (c) teacher reports of fighting and stealing. Frequent, relatively trivial antisocial acts often lead to acts that are infre-

quent and much more serious (Patterson et al., 1992). Trivial examples of coercive acts observed in the home and at school are early prototypes of adolescent delinquent behaviors. Each step in this escalating process puts the child at ever increasing levels of risk for long-term social maladjustment and its negative correlates (Patterson et al., 1992).

Although not all at-risk children seem to pass through this progression from trivial to more serious antisocial acts, approximately 75% of children with serious behavior problems apparently do (Reid, 1993). It has been shown that the further a youth moves along this progression, the greater the risk for later delinquency (Patterson et al., 1992). Patterson and Reid's research indicate that extreme antisocial scores

are actually more stable than scores in the middle of the distribution. It has been estimated that youth with early childhood onset of antisocial behavior constitute 3% to 5% of the population, but they account for half of all crimes committed by children and youth (Hinshaw, Lahey, & Hart, 1993). Children who are antisocial are at more serious risk than any other subgroup (with the possible exception of children with autism and children with schizophrenia) for a wide spectrum of adult adjustment problems such as substance abuse, institutionalization for crimes and mental disorders, disrupted marriages, and marginal employment records (Patterson et al., 1992).

Children who are withdrawn have been viewed as "at risk" for peer rejection, and the combination of rejection and isolation has been found to predict pathological and dysfunctional behavior in adolescence and adulthood (Rubin, Daniels-Bierness, & Bream, 1984). Research has shown that children who are socially withdrawn (i.e., having internalizing characteristics) are at high risk for future problems, such as leaving school (Hops, Walker, & Greenwood, 1988). Teachers have rated children who are withdrawn as having fewer social skills, less verbal ability, and more behavioral deficits (Greenwood, Todd, Hops, & Walker, 1982). Passive-dependent behavior was found to remain stable over time for young girls (Herbert, 1991). Preschoolers and kindergartners who are socially withdrawn display less cognitively mature play patterns, are less assertive than their more sociable peers, and provide fewer problem-solving strategies during hypothetical dilemmas posed by an adult interviewer (Rubin et al., 1984). A child's ability to initiate and maintain positive social interactions with others must be seen as an essential developmental achievement (Greenwood, Walker, Todd, & Hops, 1976).

P.L. 99-457 and Child-Find Activities

Under the Education of the Handicapped Act Amendments of 1986 (P.L. 99-457), each state has begun to imple-

ment "child-find" systems for identifying young children needing special education and related services, albeit with limited tools for completing this task (Beare & Lynch, 1986; Martin, 1986). According to Beare and Lynch, young children with mild to moderate learning or behavioral problems are at the greatest risk for being overlooked by traditional developmental screening tests. However, Martin noted that few technically and financially feasible screening instruments are available to assess the social and emotional functioning of preschool children, and all require lengthy administration time if every child is to be screened. Traditional methods of assessment, such as interviewing, are unreliable due to the preschool child's restricted ability to understand assessment cues and demands of the testing situation. Most of the available instruments involve behavior rating scales describing the child's behavior, completed by either the teacher or the parent. Martin (1986) rated preschool age-appropriate rating scales on their ease of administration and scoring and their psychometric reliability, as well as validity characteristics, and reported an inverse relationship between the ease of administration and psychometric characteristics.

The Systematic Screening for Behavior Disorders Procedure

The Systematic Screening for Behavior Disorders (SSBD) procedure, developed by Walker and Severson (1990), has been characterized as having acceptable levels of accuracy, cost efficiency, and consumer satisfaction (Phillips, Nelson, & McLaughlin, 1993). The SSBD differs from traditional screening instruments in that it incorporates three increasingly discriminative gates, or stages, ranging from teacher rankings to teacher ratings to direct observation of behavior. This screening system was designed to identify students at the elementary-grade level who have elevated risk status for either externalizing (e.g., aggressive, hyperactive, and antisocial) or internalizing (e.g., shy, timid, and isolated) behavioral disorders. In many cases, the SSBD proce-

dures may be a suitable alternative to traditional teacher-referral methods commonly used in schools. Following the introduction of the SSBD in 1985, other researchers confirmed that it could be successfully adapted for preschool use (Eisert, Walker, Severson, & Block, 1989; Sinclair, Del'Homme, & Gonzalez, 1993). However, although these investigators demonstrated the potential applicability of the SSBD with preschool populations, they also identified modifications necessary to make it appropriate for screening such children.

In revising the SSBD for preschoolers, it was necessary to consider the developmental appropriateness of each screening stage for younger children. Sinclair et al. (1993) applied the SSBD intact, except that direct observations of Academic Engaged Time were eliminated and the direct observation of Peer Social Behavior during free play in the classroom and on the playground was doubled from two to four 10-minute sessions. Further, in Stage 1 the teachers were asked to nominate and rank only seven externalizers and seven internalizers (out of classes of 15), rather than 10 each. The three top-ranked externalizers and internalizers were followed by Stage 2 rating scales and Stage 3 observations. Also, the cut-off criteria for defining at-risk children were adjusted to take into account the behavior levels of younger versus older children (e.g., younger children show more parallel play). The scores of the 36 selected children in the Sinclair et al. study were then compared with the normative data for elementary-level externalizers and internalizers reported by Walker and Severson (1990).

The authors of this article engaged in a 3-year process of adapting procedures for preschool populations. The results of this program of research and development to validate the preschool adaptation with preschool children, now referred to as the Early Screening Project (ESP), are described next.

METHOD

Data collected on the ESP from September 1991 through June 1994 are

presented in this article. These activities were conducted as a program of research involving a series of separate but interrelated samples. The purpose of the different samples was to replicate and extend empirical findings on the reliability and validity of the ESP instrument. The assignment of participants to samples was based on the location of the participants' homes and schools, time of data collection, and the specific data collection procedures used.

Participants

Participants were 2,853 children, ages 3 to 6 years, who were enrolled in general education and specialized programs during the 1991–1994 period. These children were from preschool and kindergarten classrooms in the following states (the number corresponds to participating subjects): California (517), Kentucky (687), Louisiana (386), Nebraska (65), New Hampshire (25), Oregon (221), Texas (612), and Utah (341). The specialized preschools included programs for children identified as having serious emotional or behavioral disorders, having developmental and language delays, and living in families with low incomes (Head Start). The total sample was composed of 46% girls and 54% boys, with most of the children (78%) not eligible for special education services. Of those who did qualify for special education services, only 2% were eligible under the behavioral disorder category. Sixty-nine percent of the children were White (as reported by their teachers), 16% were Hispanic, 12% were African American, and 3% were Native American or Asian, respectively. Family income (as reported by teachers) was 39% middle income (\$15,000–\$75,000/year); yet a substantial portion (58%) was reported to be low income (less than \$15,000/year or Head Start eligible). Of the 1,304 families with low incomes, 974 had children enrolled in Head Start. Community size was 10% urban (over 1 million), 6% semi-urban (between 250,000 and 1 million), 21% suburban, and 63% rural (less than 100,000). A decreasing number of children participated from Stages

1 to 3. Of the total 2,853 children, 1,401 (49%) moved to Stage 2 and 541 (19%) were assessed in Stage 3 (see Figure 1).

The Early Screening Project

Stage 1. Stage 1 was based on teachers' nominations and rankings of their students on externalizing and internalizing behavior dimensions. Teachers were asked to list the five children who best exemplified a behavioral description of externalizing characteristics and the five children who best exemplified a description of internalizing characteristics. The two lists were mutually exclusive. The teachers then rank ordered the nominated children on each list according to how closely each child matched the behavioral profile. Stage 1 procedures were modified from the SSBD (5 children on each list instead of 10) because of the smaller class sizes typically found in preschools.

Stage 2. ESP Stage 2, which relies upon teacher ratings, consists of four measures: the Critical Events Index, Aggressive Behavior Scale, Adaptive Behavior Scale, and Maladaptive Behavior Scale. Stage 2 differs substantially from the SSBD. For example, the SSBD Critical Events Index consists of 33 occurrence/nonoccurrence items (e.g., fire setting, reporting abuse). The ESP Critical Events Index was reduced to 16 occurrence/nonoccurrence items. Nine SSBD Critical Events items were converted to 5-point frequency ratings and included in the new Aggressive Behavior Scale. These were

1. Has tantrums;
2. Physically assaults an adult;
3. Is physically aggressive with other children;
4. Damages property;
5. Suddenly cries;
6. Ignores teacher warnings;
7. Makes lewd gestures;
8. Swears;
9. Is rejected by peers.

The Aggressive Behavior Scale contains nine items and is used only with children ranked high on the externalizing

dimension. The Adaptive Behavior Scale contains eight items representing overall prosocial forms of behavior (e.g., cooperation, positive social interactions). Four items relating to cognitive skills or academic work domains were omitted from the SSBD Adaptive Behavior Scale in the ESP version due to their developmental inappropriateness for preschool children. The Maladaptive Behavior Scale consists of nine items representing either overall antisocial or nonsocial forms of behavior (e.g., defies teacher requests, creates disturbances) and was retained intact from the SSBD. Stage 2 items were carefully worded in order to facilitate completion of the ESP by a diverse group of preschool teachers with limited experience in assessment. All SSBD items regarding academics were omitted because of their inapplicability to most preschool curricula.

Stage 3. Stage 3 behavioral observations use a modified Peer Social Behavior code. As recommended by Sinclair et al. (1993) and Eisert et al. (1989), observations of Academic Engaged Time were omitted in the ESP. The Social Behavior observations were further changed to develop greater ease of use while maintaining accuracy in identifying at-risk preschool children. Thus, the ESP observation procedure was changed from an interval to a duration recording method. A highly significant correlation of .96 ($p < .01$) between the two types of recording systems (i.e., interval's duration) showed that they provided generally equivalent information.

The duration method of recording both antisocial and nonsocial forms of behavior allows the observer to track externalizing (antisocial) and internalizing (nonsocial) behavior patterns using a single observation criterion. In addition, the observer is able to record important qualitative notes about the nature of the behavior being observed. Stage 3 ESP involves direct observations of a child's social behavior in the classroom and on the playground. The Social Behavior Observation provides a record of the quality, level, and dis-

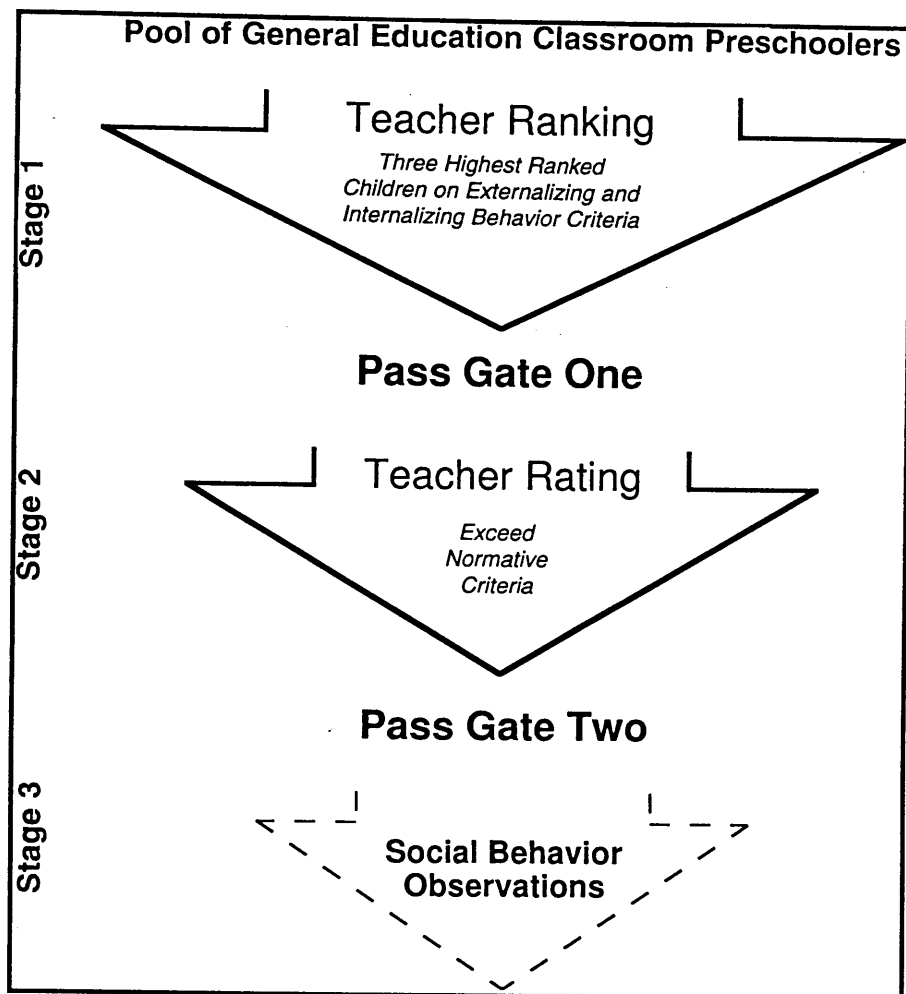


FIGURE 1. Multiple gating procedure used in the Early Screening Project (ESP).

tribution of a child's social behavior during free-play settings. Antisocial or nonsocial behavior was defined as: (a) a *negative* exchange of either verbal or physical interaction, (b) disobedience regarding established classroom rules, (c) tantrums, and (d) solitary play. Children are observed for 20 minutes, 10 minutes each on two occasions. If the total time is less than 20 minutes, a third observation is scheduled to bring the observed time up to a minimum of 20 minutes. In the procedure, the stopwatch runs when the child exhibits antisocial or nonsocial behavior and is turned off when the child displays prosocial behavior. The stopwatch is then restarted when the child again exhibits antisocial or nonsocial behavior. This procedure is repeated throughout the recording observation period to

record the total time the child is involved in either antisocial or nonsocial behavior.

Concurrent Validity Measures

In order to test the accuracy of concurrent criterion-related validity, a series of concurrent measures were used as part of the research on the technical adequacy of the ESP. As noted previously, there is a paucity of teacher rating scales for preschool-age children (Martin, 1986). For this reason, a mix of concurrent measures were used in order to obtain a broad assessment with multiple measures. The sites (i.e., Oregon, Kentucky, and Louisiana) that participated in reliability and validity testing of the ESP are listed in Table 1.

Behar Preschool Behavior Questionnaire. The Behar Preschool Behavior Questionnaire (PBQ; Behar & Stringfield, 1974) was developed as a screening instrument for identifying preschoolers showing potential behavior disorders. Behar and Stringfield found that the scale significantly differentiated groups of children with ($n = 102$) and without ($n = 496$) behavior problems. The PBQ is one of the few teacher rating scales developed specifically for screening preschool children. In our initial ESP study conducted in Oregon ($n = 121$), the participating teachers and assistant teachers completed the PBQ at the same session in which they completed the ESP Stage 2 scales.

Conners' Teacher Rating Scale. The Conners Teacher Rating Scale (Conners, 1990) is widely used to assess children ages 3 to 17 years for behavioral problems, especially for attention-deficit hyperactive disorder (ADHD). The Conners has seven subscales: Hyperactivity, Conduct Problems, Emotional Indulgence, Anxious-Passive, Asocial, Daydream-Inattention, and Hyperactivity index. The scale has been found to have acceptable levels of reliability and validity (Conners, 1990; Goyette, Conners, & Ulrich, 1978). In our first two studies in Oregon ($n = 121$ and $n = 65$) the teachers and assistant teachers completed the Conners in the same session as they completed the Stage 2 ESP scales.

Teacher Report Form. The Teacher Report Form (TRF; Achenbach & Edelbrock, 1986) was constructed to obtain teachers' reports of their students' behavior problems and adaptive functioning. The TRF, a teacher-completed behavior checklist for school-age children, includes behavior problems that teachers rate on a 3-point scale (0 = *not true*, 1 = *somewhat or sometimes true*, 2 = *very often or often true*). The high technical quality of the TRF is well documented (Gresham, 1985). Teachers in Kentucky ($n = 339$), Louisiana ($n = 135$), and in one Oregon school ($n = 37$) completed the TRF at the same session as the ESP Stage 2 scales.

TABLE 1
Reliability and Concurrent Validity Samples

Site	Subjects in ESP Stage 1 (n)	Reliability studies		Concurrent validity measures		
		Interrater	Test-Retest	Behar	Conners	TRF
Oregon 1	121	√	√	√	√	
Oregon 2	65	√	√		√	
Oregon 3	35	√	√			√
Kentucky	687					√
Louisiana	386					√

Note. ESP = Early Screening Project; Behar = Behar Preschool Behavior Questionnaire; Conners = Conners Teacher Rating Scale; TRF = Teacher Report Form.

Data Collection Procedures and Schedule

In our initial research on the ESP conducted in Oregon, 121 preschoolers were evaluated on all three stages of the ESP in the fall (November and December) and spring (March to May) of 1991-1992. An additional 63 Head Start students were included as a longitudinal sample in the spring of 1993. No gating criteria (i.e., exclusion of normal subjects) were used in the administration of the Stage 3 observational data, as data on all children were needed for reliability and validity procedures. Stage 3 observations were completed by a group of professionally trained observers. Training sessions were conducted until each observer could code videotaped scenes at a 90% accuracy level (average training time was 6 hours). The observers were unaware as to subject ranking or history, and observer accuracy was measured by the interval agreement method (Hartman, 1977; see Interrater Reliability below). A random sample of 20% ($n = 74$) of observations was completed in observer pairs to monitor interrater reliability.

As part of the national program of research on the ESP at sites in all the states included in the study, the ESP was used as part of normal ongoing screening procedures. Local professionals in these sites attended a daylong inservice training on ESP administration procedures. Gating criteria were used in these sites, so a decreasing num-

ber of children participated in the subsequent stages. Observers were professionals in the community (e.g., special education consultants) who attended an ESP training workshop; these observers were not blind as to subject ranking or history.

RESULTS

Item Aggregation into Scales

In our analysis of the ESP database, the first task was to aggregate the individual items and observational categories into functional scales. The ESP Stage 2 items were compared to aggregated scale total scores using Pearson product-moment correlations. The item-total correlations for all Stage 2 measures were highly significant and ranged from .16 to .89, with a median of .63. The median item-total correlations for the Critical Events Index and the Aggressive, Adaptive, and Maladaptive Behavior scales were .39, .70, .82, and .81, respectively. Mean levels of the Stage 3 observational data were compared by session in order to decide whether to aggregate the observation sessions into totals per code category. There were a low number of observations for Session 3 subjects, who were only observed for a third session if their total time (per observational coding system) was less than 20 minutes. Because the percentages of the coded behaviors did not change appreciably over sessions, aggregated totals were used

for further analyses of the ESP observation data.

Interrater Reliability of Teacher-Assistant Teacher Ratings

Pearson product-moment correlations and kappa coefficients between raters (i.e., teacher/assistant teacher and teacher/parent pairs) for ESP Stages 1 and 2 and the concurrent measures (i.e., PBQ, Conners', and TRF, when applicable) were completed to obtain interrater reliability coefficients (see Table 1).

ESP Stage 1. In order to determine whether a child was nominated to be among the three highest ranked externalizers and internalizers, respectively, in each classroom by the teacher and assistant teacher, kappa coefficients were computed, resulting in coefficients of .70 and .48. These coefficients indicated that ESP Stage 1 has adequate interrater reliability for screening purposes.

ESP Stage 2 and Concurrent Measures. Comparing the Oregon teachers' and assistant teachers' scale scores resulted in highly significant Pearson product-moment correlations ranging from .58 to .74 (see Table 2). The scales, which contain more negatively valenced items (i.e., Critical Events, Aggressive Behavior, and Maladaptive Behavior), seemed to be more salient to the teacher. That is, noxious behavior seemed to be more accurately reported, and, therefore, the interrater reliability was greater than for the positively valenced scale (i.e., Adaptive). Although the coefficients are not as large as in the Conners Hyperactive scale, these correlations are of sufficient magnitude to be useful because they are approximately equal to those for the other concurrent measures (i.e., PBQ and Conners Inattention). The correlations are well within their acceptable range for concurrent validity coefficients as determined by reviews of past studies (Goyette et al., 1978).

ESP Stage 3 Observations—Social Behavior. The interrater reliabilities of the ESP observational measure were

calculated from a random sample of 20% of the remaining observations. In our initial research in Oregon ($n = 186$), the Social Behavior (SB) coding interrater reliability was analyzed by interval agreement. For an agreement to occur, both observers would have to mark the same category and same valence, if appropriate (i.e., positive or negative), in the identical time interval (Hartman, 1977). This procedure resulted in an average percentage agree-

ment of .87 across observers, which is within acceptable limits for a screening device of this type (Salvia & Ysseldyke, 1988). For the subsequent samples, the SB interrater reliability was derived by dividing the sum of the smaller duration scores by the sum of the larger duration score (Hartman, 1977). This provides a percentage indicator of observer differences, weighted for length of observation, and resulted in an average coefficient of .93.

Test-Retest Reliability

For establishing test-retest reliability in samples conducted in Oregon, teachers and assistant teachers were asked to nominate, rank order, and rate the children again in the spring following a 6-month interim period (fall to spring). A second cross-tabulation table was constructed for each participating classroom (see Table 3). Kappa coefficients computed between the teachers in the fall and spring data collections resulted in coefficients of .63 for externalizers and .35 for internalizers. These coefficients show a drop in magnitude, but this is to be expected with a 6-month period between data collection occasions.

Scores obtained in fall and spring administration of the ESP Stage 2 (i.e., Critical Events, Aggressive Behavior, Adaptive Behavior, and Maladaptive Behavior) and concurrent measures (i.e., PBQ and Conners) were compared and resulted in highly significant correlations ranging between .74 and .90 (see Table 4). The ESP Stage 2 measures' correlations are comparable to the concurrent scales in regard to test-retest reliabilities. Correlations of fall and spring teacher ratings on the Behar and Conners scales were statistically significant, with coefficients between .61 and .82 (see Table 4).

Discriminative Validity

The ESP's discriminative validity was examined by graphing the standardized scores for the measures by subject group and through initial use of a discriminative function analysis procedure. Children were included in this analysis if: (a) complete data were available on them, and (b) they were either highly ranked (externalizer or internalizer) or part of the comparison group (non-ranked). The consistency across ESP measures was examined by comparing the standardized t scores ($M = 50$, $SD = 10$) of the children ranked highest on Stage 1 externalizer and internalizer dimensions, respectively, as well as children ranked as average (serving as a control comparison) across ESP measures (see Figure 2). The positively ori-

TABLE 2
ESP Stage 2, Behar, and Conners Scales' Interrater Reliability Coefficients

Scale	n	Teacher A		Teacher B		r
		M	SD	M	SD	
ESP Stage 2 teacher ratings						
Critical Events Index	175	1.13	1.69	1.04	1.44	.66***
Aggressive Behavior Scale	175	13.79	6.64	12.40	5.17	.74***
Adaptive Behavior Scale	175	30.33	7.33	30.04	7.23	.58***
Maladaptive Behavior Scale	175	18.23	8.05	17.87	7.35	.64***
Concurrent teacher ratings						
Behar	114	11.06	8.91	10.90	10.00	.54***
Conners Hyperactive	175	11.86	12.00	9.69	10.56	.80***
Conners Inattention	175	2.34	2.30	1.95	2.06	.63***

Note. ESP = Early Screening Project; Behar = Behar Preschool Behavior Questionnaire; Conners = Conners Teacher Rating Scale.

*** $p < .001$.

TABLE 3
ESP Stage 1 Cross-Tabulations and Test-Retest Reliability Kappa Coefficients for Studies in Oregon

Externalizers	Fall	Spring		
		Top 3	Not Top 3	
	Top 3	5 (5%)	18 (17%)	23
	Not Top 3	73 (69%)	9 (9%)	82
		78	27	105
Kappa	.63			
Internalizers	Fall	Spring		
		Top 3	Not Top 3	
	Top 3	10 (10%)	12 (11%)	22
	Not Top 3	69 (66%)	14 (13%)	83
		79	26	105
Kappa	.35			

ented Adaptive Behavior Scale's scores were reversed to maintain continuity in valance (Figure 2).

The externalizer group was clearly more deviant (less well-adjusted) on these measures than the other two groups (see Figure 2). The Critical Events, Adaptive Behavior, and Social Behavior measures best differentiate internalizers from the nonranked group. The externalizer group had high scores

(i.e., antisocial and nonsocial) on the Aggressive and Maladaptive scales, respectively. Both the externalizer and internalizer groups had relatively equivalent scores on the Stage 3 Social Behavior observational measure. The *t*-score means of ESP measures by high externalizer, high internalizer, and nonranked groups are shown in Figure 2.

A discriminative function analysis, using the general linear model, estimates

the accuracy of a set of dependent measures in predicting a priori groupings of subjects (i.e., boys to girls, pass and fail, etc.). In this study, the groupings were based on teacher recommendations of behavioral disorders (BD) eligibility status (whether the teacher listed the child as needing further evaluation for BD status), and the dependent measures were those of ESP Stage 1 (teacher rankings of externalizer and internalizer), Stage 2 (Critical Events, Aggressive Behavior, Adaptive Behavior, and Maladaptive Behavior scales), and Stage 3 (Social Behavior observation). The three highest ranking internalizers and externalizers, respectively, in each class were divided into groups for the discriminative analysis.

The false positive and negative error rates were very low, 4% and 9%, respectively. A MANOVA test of the group means for the ESP measures found a highly significant difference, $F(8,94) = 20.38, p < .001$. The discriminative function and MANOVA test show that the ESP is acceptably accurate in predicting problem behaviors among preschoolers in this study.

TABLE 4
ESP Stage 2, Behar, and Conners Scales' Test-Retest Reliability Coefficients

Scale	n	Fall		Spring		r
		M	SD	M	SD	
ESP Stage 2 teacher ratings						
Critical Events Index	105	1.04	1.55	1.27	1.81	.74***
Aggressive Behavior Scale	105	13.44	6.45	13.74	6.17	.90***
Adaptive Behavior Scale	105	31.68	7.25	31.69	6.84	.75***
Maladaptive Behavior Scale	105	17.60	7.41	18.12	7.66	.80***
Concurrent teacher ratings						
Behar	105	10.92	8.74	12.17	9.06	.79***
Conners Hyperactive	105	10.95	10.53	11.29	10.22	.82***
Conners Inattention	105	2.24	2.12	1.87	1.91	.61***

Note. ESP = Early Screening Project; Behar = Behar Preschool Behavior Questionnaire; Conners = Conners Teacher Rating Scale.

*** $p < .001$.

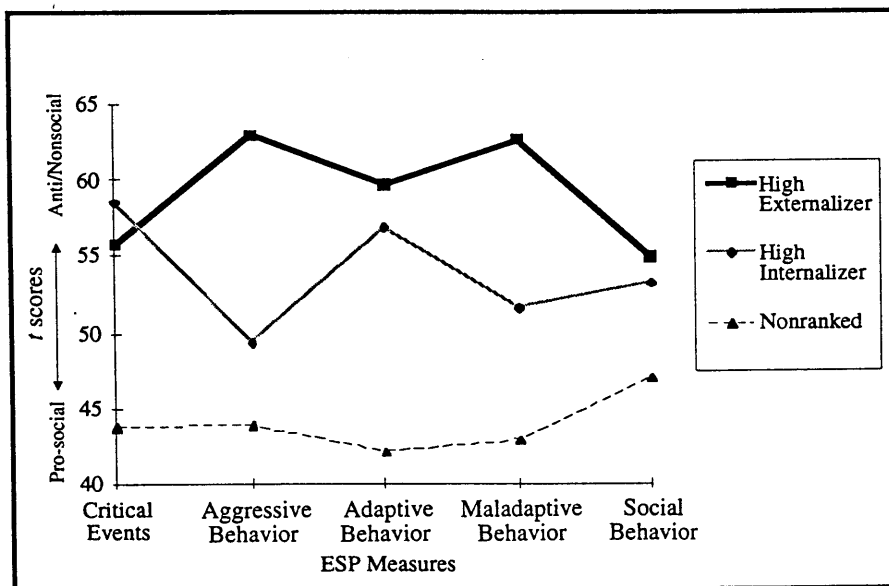


FIGURE 2. *T*-score means of ESP measures by high externalizer, high internalizer, and nonranked groups.

Concurrent Validity

The concurrent validity of the ESP was examined through correlations between the ESP Stage 2 measures and the PBQ, Conners, and TRF instruments. These results showed very good ESP concurrent validity, with correlations ranging between .18 to .89 (see Table 5). The Aggressive Behavior, Adaptive Behavior, and Maladaptive Behavior scales showed substantial concurrent validity. The lower correlation for the Critical Events Index could be due to the nature of the very low-frequency behavior-specific items comprising this index. The low correlation between teacher ratings and observations is not atypical, due to the nature of the differing methods (Cairns & Green, 1979).

DISCUSSION

The ESP conforms to the developmental standards for assessment methodology among preschool-age children

TABLE 5
Concurrent Validity with Pearson Product-Moment Correlations Between ESP Measures and Concurrent Measures

Concurrent measures	ESP measures				
	Critical events	Aggressive Behavior	Adaptive Behavior	Maladaptive Behavior	Social behavior obs.
Behar	.37***	.68***	-.77***	.79***	.18*
Conners Hyperactive	.46***	.80***	-.74***	.89***	.23**
Conners Inattention	.59***	.60***	-.69***	.70***	.33***
TRF-Externalizer	.44***	.83***	-.68***	.88***	.24***
TRF-Internalizing	.54***	.19***	-.34***	.22***	.18**
Median	.46	.68	-.69	.79	.23

Note. The number of subjects ranges from 121 to 541 from Oregon, Louisiana, and Kentucky. ESP = Early Screening Project; Behar = Behar Preschool Behavior Questionnaire; Conners = Conners Teacher Rating Scale; TRF = Teacher Report Form.
* $p < .05$. ** $p < .01$. *** $p < .001$.

suggested by Bredekamp (1987). It assesses preschool-age children's social and emotional behavior with multi-method techniques and with an emphasis on teacher judgment (i.e., Stages 1 and 2), and developmental differences between preschool- and school-age children have been accommodated and taken into account. These differences were derived from three sources: empirical findings from past samples on child development, the judgments of a panel of experts, and preschool teachers' feedback. In the formulation phase of this research, all three sources were combined to structure the initial form of the ESP. ESP item selection and adaptations were also based on a search of the relevant literature. The Stage 2 observational measures were redesigned to reflect early childhood social behavior patterns. This attention to early childhood development, and the extensive care taken to accurately characterize typical behaviors occurring within early childhood settings, have been instrumental in moving the ESP toward meeting those developmental standards.

Reliability correlation coefficients appear to be quite adequate for screening purposes. Interrater reliability coefficients for ESP Stage 2 measures (i.e., Critical Events, Aggressive Behavior, Adaptive Behavior, and Maladaptive Behavior) were comparable to the concurrent measures' interrater correlation coefficients. The consistency across ESP measures for the total sample demonstrates the ESP's ability to differentiate

between groups (see Figure 2). However, the argument could be made that the consistency across measures could be due, in part, to common method variance (Cairns & Green, 1979; Patterson et al., 1992). The high degree of relationship between group membership (i.e., high externalizer, high internalizer, and nonranked) and teacher ratings is not unusual because the teachers provided the choices for group membership and the ratings.

Validity coefficients for the ESP were found to be adequate for a screening measure of this type. The discriminant function results showed that the ESP has a very low chance of overidentifying children with behavior problems. Usually it would be better for a screening instrument to slightly overidentify potentially at-risk children, because later assessment could separate the false-positives from the true-positives group. Because the issue of labeling young children as behavior disordered or seriously emotionally disturbed can be fraught with personal feelings of stigmatization, the ESP's small chance of obtaining a false-positive result is a potential asset (i.e., practitioners can be confident that a child who is identified with the ESP is actually different from his or her peers in terms of risk status). In sum, the ESP has more than acceptable psychometric characteristics and provides cost-effective procedures, which makes its use appropriate according to current special education standards for decision making.

The use of a multiple gating system within a large-group setting minimizes teacher and special services consultant time. Further, it allows proactive and universal screening of all children and affords each preschooler an equal chance to be identified for externalizing or internalizing problems. Longitudinal research is increasingly showing that we cannot wait for at-risk children to be reactively identified by exceeding the tolerance limits of the school system (Zigler, Taussig, & Black, 1992). If we allow problems to escalate to this point, we are often unable to turn them around. Further, it is extremely difficult to engage in effective early intervention in the absence of proactive screening efforts that are timely, sensitive, accurate, and cost-effective with educators' time. Stages 1 and 2 ESP screening procedures can be completed for an entire preschool in about 1½ hours. The gating procedure was found to reduce assessment time up to 16% over other procedures for elementary school children (Walker, Severson, Nicholson, Kehle, Jenson, & Clark, 1994). Walker et al. (1994) also found a very strong relationship between time and cost of assessment in the context of evaluating behavior disorders at the elementary level.

As noted, longitudinal studies on the ESP are needed to confirm its predictive validity. In this regard, it is important to show that preschool behavior problems identified early on are stable over the long term and that the ESP

detects preschoolers whose problems are likely to become worse if left untreated.

Current plans for future ESP studies include application with a greater and more diverse number of participants and a longitudinal follow-up on a subsample of current subjects. The proposed studies also will broaden the ESP normative sample to include other geographical locations. Consumer satisfaction with the ESP has not been tested directly. A study to ask field coordinators about problems in ESP administration procedures and benefits of use would be very helpful in guiding further modifications and establishing usage guidelines. For example, during the field testing of the ESP, several site coordinators noted that the screening procedures also functioned as an educational tool for teachers. Because the training and educational background of preschool teachers varies widely, many of them benefit in their teaching and management from exposure to the explicit behavioral descriptions contained in the ESP. The specific behaviors of externalizing and internalizing characterizations and Stage 2 measures, such as the Critical Events Index, give preschool teachers salient indicators for behavior patterns that could lead to adjustment problems later in a child's development.

About the Authors

EDWARD G. FEIL, PhD, is a research associate at Oregon Research Institute in Eugene. His research interests include early childhood development and early assessment and intervention for young children with emotional and behavioral problems. HILL M. WALKER, PhD, is the associate dean and director of the Center on Human Development in the College of Education at the University of Oregon in Eugene. His research interests include behavioral assessment, social skills intervention, and antisocial behavior. HERBERT H. SEVERSON, PhD, is a research scientist at Oregon Research Institute and associate professor of counseling psychology at the University of Oregon in Eugene. His research interests are early identification of behavior disorders, smoking and substance use prevention, smokeless tobacco cessation, and risk perception. Address: Edward G. Feil, Oregon Research Institute, 1715 Franklin Boulevard, Eugene, OR 97403; e-mail: edf@ori.org

Authors' Notes

1. We would like to express our thanks to Drs. Wes Becker and Roland Good for their contributions and assistance, without which this research could not have been conducted.
2. This research was supported in part through grants from the following two sources: U.S. Department of Education, Office of Special Education and Rehabilitative Services, Research in Education of the Handicapped Programs: Student-Initiated and Field-Initiated Research; and the U.S. Department of Health and Human Services, Administration for Children and Families: Head Start Research Fellows Program. The views expressed herein are not necessarily those of the USDOE or USHHS.

References

- Achenbach, T. M., & Edelbrock, C. (1986). *Manual for the Teacher's Report Form and teacher version of the Child Behavioral Profile*. Burlington: University of Vermont Department of Psychiatry.
- Beare, P. L., & Lynch, E. C. (1986). Under identification of preschool children at risk for behavioral disorders. *Behavioral Disorders, 11*, 177-183.
- Behar, L., & Stringfield, S. (1974). *Manual for the Preschool Behavior Questionnaire*. Raleigh-Durham, NC: Behar.
- Bower, E. M. (1981). *Early identification of emotionally handicapped children in school* (3rd ed.). Springfield, IL: Thomas.
- Bredenkamp, 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. A., & Green, J. A. (1979). How to assess personality and social patterns: Observations or rating. In R. B. Cairns (Ed.), *The analysis of social interaction: Methods, issues, and illustrations* (pp. 209-229). Hillsdale, NJ: Erlbaum.
- Conners, C. K. (1990). *Manual for the Conners Rating Scales*. North Tonawanda, NY: Multi-Health Systems.
- Education of the Handicapped Act Amendments of 1986, 20 U.S.C. § 1400 et seq.
- Eisert, D. C., Walker, H. M., Severson, H., & Block, A. (1989). Patterns of social-behavioral competence in behavior disordered preschoolers. *Early Childhood Development and Care, 41*, 139-152.
- Fischer, M., Rolf, J. E., Hasazi, J. E., & Cummings, L. (1984). Follow-up of a preschool epidemiological sample: Cross-age continuities and predictions of later adjustment with internalizing and externalizing dimensions of behavior. *Child Development, 55*, 137-150.
- 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.
- Greenwood, C. R., Todd, N. M., Hops, H., & Walker, H. M. (1982). Behavior change targets in the assessment and treatment of socially withdrawn preschool children. *Behavioral Assessment, 4*, 273-297.
- Greenwood, C. R., Walker, H. M., Todd, N. M., & Hops, H. (1976). *Preschool teachers' assessments of social interaction: Predictive success and normative data* (CORBEH Report No. 26.). Eugene: University of Oregon, Center on Human Development, College of Education.
- Gresham, F. M. (1985). Behavior disorder assessment: Conceptual, definitional, and practical considerations. *School Psychology Review, 4*, 495-509.
- Hartman, D. P. (1977). Considerations in the choice of interobserver reliability estimates. *Journal of Applied Behavior Analysis, 10*(1), 103-116.
- Herbert, M. (1991). *Clinical child psychology*. Chichester, UK: Wiley.
- Hinshaw, S. P., Lahey, B. B., & Hart, E. L. (1993). Issues of taxonomy and comorbidity in the development of conduct disorders. *Development and Psychopathology, 5*(1 & 2), 31-49.
- Hops, H., Walker, H. M., & Greenwood, C. R. (1988). *Procedures for establishing effective relationship skills*. Seattle: Educational Achievement Systems.
- Kohn, M. (1977). *Social competence, symptoms, and underachievement in childhood: A longitudinal perspective*. New York: Wiley.
- Lerner, J. A., Inui, T. S., Trupin, E. W., & Douglas, E. (1985). Preschool behavior can predict psychiatric disorders. *Journal of the American Academy of Child Psychiatry, 24*(1), 42-48.
- Martin, R. P. (1986). Assessment of the social and emotional functioning of preschool children. *School Psychology Review, 15*, 216-232.
- Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. *American Psychologist, 44*, 329-335.
- Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial boys*. Eugene, OR: Castalia.

(continued on p. 213)

