Proactive Screening for Emotional/ Behavioral Concerns in Head Start Preschools: Promising Practices and Challenges in Applied Research

Edward G. Feil
University of Oregon and Oregon Research Institute

Hill Walker
University of Oregon

Herbert Severson
University of Oregon and Oregon Research Institute

Alison Ball
University of Oregon

ABSTRACT: There is a growing need for cross-cultural research on screening instruments appropriate for use with young children from culturally diverse backgrounds who are at risk for emotional or behavioral problems. Head Start classrooms provide an ideal environment for (a) conducting applied research and (b) encouraging best practices in screening, assessment, and early detection among diverse multicultural low-income populations. This research assessed the cross-cultural psychometric characteristics and validity of a multiple-gating screening procedure used by the Early Screening Project (ESP) to screen and identify children at risk for behavioral problems in Head Start centers in rural and urban sites in Oregon. The ESP procedure relies on teacher judgments, in vivo behavioral observations, and normative criteria to identify preschool children exhibiting serious behavioral problems. This research provides initial evidence that the ESP can be used appropriately within the context of multicultural Head Start populations.

The early screening, identification, and treatment of emotional and behavioral disorders has become a high priority for early childhood educators, Head Start policymakers, staff, and parents. Effective methods of detecting children who are at risk are an essential prerequisite for intervening early in the trajectory of emotional or behavioral problems. For early intervention to be effective and socially valid, it is essential that (a) the right children are targeted for intervention and services, (b) the screening and identification procedures used select true positives and hold false negatives to an absolute minimum, and (c) the child's status is assessed on key risk and protective factors associated with long-term outcomes, both negative and positive (Hawkins, Von Cleve, & Catalano, 1991; Walker et al., 1996).

The federal Head Start program has undergone numerous policy and programmatic changes in standards, now including the early identification and assessment of disabling conditions. This mandate is based upon the belief that early identification, assessment, and subsequent intervention(s) can reduce and ameliorate the future adjustment and performance problems of vulnerable young children (Lerner, Inui, Trupin, & Douglas, 1985; Zigler, Taussig, & Black, 1992). Head Start programs provide early access to low-income, diverse populations who may have elevated risk status

for school failure and emotional or behavioral problems. The early screening and identification procedure, that is the focus of this research, the Early Screening Project (ESP; Walker, Severson, & Feil, 1995), incorporates the proactive, universal screening standard for emotional and behavioral problems required by Head Start Performance Standards and ensures that all children are screened for the presence of these problems.

Cultural factors can powerfully mediate the social behavior of children from diverse backgrounds within the context of schooling (Koot, Van Den Oord, & Boomsma, 1997). Consequently, screening and identification procedures also must be normed and constructed appropriately so that the impact of cultural background factors does not cause Head Start children to be misjudged in the screening process and identified for the wrong reasons. For example, children from minority backgrounds who have been exposed to long-term risk conditions (e.g., poverty) may have a higher likelihood of being identified and referred by their teachers simply because of the referral agent's knowledge of these factors.

Research is needed that ensures that technically adequate and user-friendly screening and identification procedures can be used appropriately and effectively within the context of cultural diversity represented by Head Start child populations. Achievement of this goal requires that universal screening procedures be used in which every child is screened in the same manner according to objective criteria and is given an equal chance to be identified. Additionally, the child's level and pattern of scores should be compared to an appropriate normative profile of children whose gender and cultural background factors are similar (Sugai, Maheady, & Skouge, 1989).

The purpose of our research was to (a) test the feasibility of the ESP's use with each of these child populations and (b) determine whether separate decision-making cutoff points and criteria are necessary to ensure objective use of the ESP procedures with these subgroups of the Head Start population. This article presents the results from the first cohort's initial year of a longitudinal study. The practical lessons learned are discussed in the hope of assisting other researchers working with diverse Head Start populations.

Prevalence of Emotional or Behavioral Problems

The actual prevalence of children with behavioral problems is difficult to ascertain with any certainty; figures and statistics vary greatly in the literature. For example, Bower (1982) argued that approximately 10% of all students have moderate to serious emotional problems, while Brandenburg and associates (Brandenburg, Friedman, & Silver, 1990) suggested that at least 7% of all students may have emotional problems serious enough to warrant treatment. Others have said that about 2% to 3% of all children should be served under the aegis of Public Law 94-142's Seriously Emotionally Disturbed category (Forness & Knitzer, 1990; Kazdin, 1987; Sinclair, Del'Homme, & Gonzalez, 1993). Most recently, a U.S. government report for the nation as a whole stated that fewer than 1% of all school-age children are currently served under the category of Seriously Emotionally Disturbed (U.S. Department of Education, 1999). For Head Start, only about 2% of children were reported to have an emotional or behavioral disorder (U.S. Department of Health and Human Services, 1998). The disparity between such prevalence estimates and data on the actual percentage of children served indicates a significant deficiency in the early identification of children exhibiting serious emotional or behavioral problems such as antisocial behavior patterns and conduct disorder.

Screening for Behavior Problems

We believe that a careful structuring of the classroom teacher's evaluation of all children in the classroom, in relation to objective criteria that define behavioral at-risk status, can yield long overdue improvements in the referral practices of most school systems. At best, current practices appear to be reactive and highly idiosyncratic to the behavioral standards of individual referring teachers (Gerber & Semmel, 1984). At worst, they are extremely biased in the direction of securing the removal of referred students from the educational mainstream with the accompanying goals of increasing classroom homogeneity, reducing classroom management pressures, and improving overall teachability (Ysseldyke, Algozzine, & Epps, 1983; Ysseldyke, Christen-

son, Pianta, & Algozzine, 1983). We also believe that these practices can be improved significantly via the following methods:

1. Adoption of more objective definitional criteria for school-related behavioral problems and disorders.
2. Structured involvement of teacher appraisal procedures in the initial screening and assessment process.
3. Use of multiple-gating assessment procedures (Loeber, Dishion, & Patterson, 1984) to provide integrated and multiple sources of data in a cost-efficient screening and identification process.

Multiple gating is a procedure that contains a series of progressively more expensive and precise assessments, or "gates," that (a) provide for the sequential assessment and cross-validation of multimethod forms of child assessment and (b) establish a decision-making structure for the aggregation of information produced by different assessment sources. It appears that the climate for adoption of such a model is timely, given the widespread dissatisfaction that parents and educators have expressed regarding current behavioral assessment practices at both preschool and elementary levels (see Huntze & CCBD Subcommittee on Terminology, 1985; Jenson, 1984; Kauffman, 2001; Wood, Smith, & Grimes, 1985). When combined with professionals' advocacy for the adoption of more objective and standardized assessment procedures (see Council for Children with Behavioral Disorders [CCBD], 1987), the case for more generically effective practices is highly persuasive.

Walker and Severson (1990) designed a three-stage multiple-gating assessment model for the screening and identification of potentially at-risk elementary-age children that addresses many of the problems in assessment practices alluded to above. The Systematic Screening for Behavior Disorders (SSBD) procedure has undergone extensive evaluation, and it has been recommended by the U.S. Office of Education as an exemplary best practice. Walker and colleagues (1995) adapted the SSBD for use with preschool populations via development of the ESP. The ESP required changes in the item definitions and instrument formats in Stages Two and Three of the original version of the SSBD. Items that referred to classroom academic issues were deleted in the ESP, and observational measures were redesigned to reflect early childhood development.

The ESP universal screening procedure provides for cost-effective, mass screening of all young children who are enrolled in regular preschool and kindergarten classrooms; it links (a) definitional criteria, (b) screening and assessment procedures, and (c) normative-based eligibility decision making into one self-contained system. This model relies heavily on structured teacher judgment of child behavioral characteristics in the first two assessment stages and uses normatively referenced behavioral observation data to provide independent in vivo assessments of the child's functioning within instructional and free-play settings in Stage Three. The results of assessments and decision making in initial screening stages are cross-validated by increasingly more intensive assessments in subsequent screening stages.

The system is patterned after models developed and validated by Greenwood, Walker, Todd, and Hops (1979) for preschool screening of children at risk for social withdrawal and by Loeber and colleagues (1984) for screening of children at risk for adoption of a delinquent life-style. It also provides each child in a regular classroom setting with an equal chance to be identified for both externalizing and internalizing behavioral disorders (Achenbach & Edelbrock, 1978; Ross, 1980). These two dimensions cover the broad range of school behavioral disorders that occur in both the preschool and elementary age range. Achenbach and Edelbrock (1978) and Ross (1980) have argued persuasively for the adoption of this bipolar classification system to govern school-based assessment practices.

Multicultural Issues and Screening for Behavioral Problems: Culture and Assessment

One assessment goal that has received much attention in the United States has been culture-free measurement. Overrepresentation of students from ethnic minority groups within special education has been a persistent finding (Artiles & Trent, 1994; Sugai & Maheady, 1988). In regard to intellectual assessments, nondiscriminatory testing has been at the center of several court decisions (e.g., *Larry P.*, 1979 [Bersoff, 1980]) with one of the results

TABLE 1
Demographics

	Teacher-Reported Ethnicity						
	Asian American	*African American*	*Hispanic*	*Native American*	*White*	*Other*	*Total*
Male	59	55	109	35	226	1	485
Female	43	57	106	23	197	-	426
Total	102	112	215	58	423	1	911

being the discontinuation of all IQ testing of African-American children in California. Craig, Kaskowitz, and Malgoire (1978) found that African-American children were recommended for placement in settings for children with educable mental retardation and emotional disorders in greater relative numbers than White children. To make matters worse, assessment and identification of children with behavioral problems are plagued by the lack of objective criteria, unacceptable reliability, poorly structured and biased referral procedures, and the want of direct observational data (Sugai & Maheady, 1988). However, while earlier ESP norming procedures and psychometric studies did sample from a range of ethnic groups and low-income populations, ethnicity had never been a primary research focus.

Participants

Children enrolled in six Head Start programs in Oregon were recruited to be involved in this study. These Head Start programs operate approximately 200 classrooms serving a total of approximately 3,000 children aged 3 and 4 years. Head Start teachers in these sites were invited to participate in the universal screening and identification phases of the project. After teachers in 40 classrooms agreed to participate, all 954 children in their classes were screened with ESP Stage One ranking procedures. Of the children screened, 47% were females. Eighteen percent of the children were eligible for special education services and an additional 4% were under evaluation, totaling 210 children (22%). Of these children, 2% were categorized as having behavioral disorders, 1% attention deficit with hyperactivity disorder (ADHD), 7% developmental delays, 7% speech/language delays, 4% noncategorical delays, and 1% other categories (e.g., other health impaired). As reported by their teachers, 45% of the children were White and 25%, 12%, 6%, and 11% were Hispanic, African American, Native American and Asian American, respectively (see Table 1).

ESP Stage One nomination and rank-ordering procedures, using behavioral descriptions of externalizing and internalizing dimensions, resulted in six children being selected for further assessment using teacher and parent questionnaires and direct observations (i.e., the three highest-ranked externalizers and the three highest-ranked internalizers from screening Stage One). A randomly chosen comparison boy and girl (neither of whom was nominated or ranked on either of the externalizing and internalizing dimensions in Stage One by his or her classroom teacher) were also included. Teachers sent home letters inviting families to participate, and 39% of eligible families returned signed consent forms (i.e., 126 of a possible 320). Dividing the sample into internalizer, externalizer, and nonranked groups showed fairly even participation across identified groups and slightly higher participation for the nonranked comparison group. By groups, participation rates were 38% of possible internalizers (i.e., 46 of 120), 37% of possible externalizers (i.e., 44 of 120), and 45% of nonranked comparisons (i.e., 36 of 80). Teachers completed ESP Stage Two measures and criterion measures (e.g., Social Skills Rating Scale) on these children, who were also observed using the Peer Social Behavior observation code (Walker et al., 1995). Parents completed a short questionnaire including the criterion measures returned by mail. Trained personnel recruited from within the Head Start programs and local community completed the observations. Training sessions were conducted until each observer could code scenes at 80% accuracy (average training time was 6 hours).

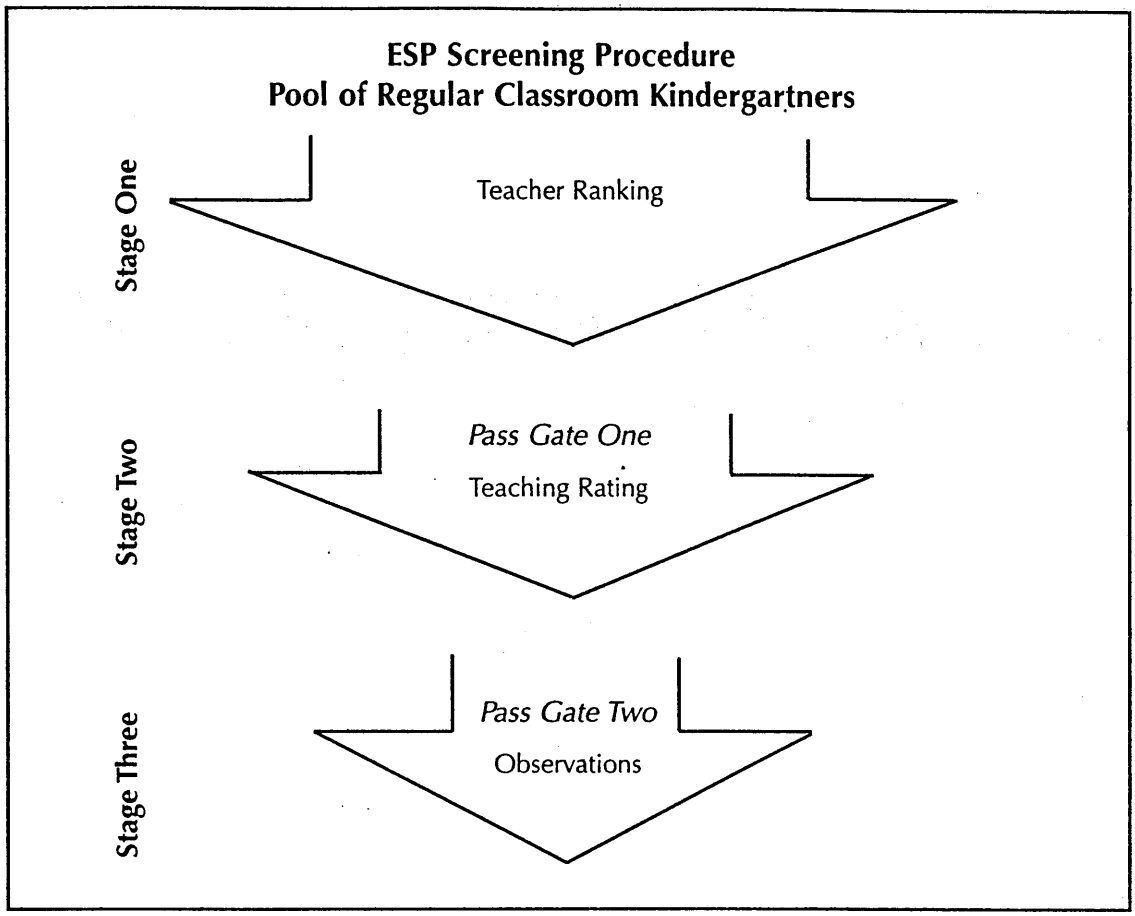

FIGURE 1. Screening and Student Identification Processes

Twenty-four parents (19%) were interviewed to obtain a better sense of neighborhood contextual factors, with specific attention given to violence since it is an important factor in the development of antisocial behavior. The sample was distributed across three of the sites, two in rural settings and one urban. Of the mothers interviewed, 21% reported they had seen a nonviolent crime in their neighborhood more than once, 12% reported they had seen a violent crime in their neighborhood more than once, and 8% reported having been a victim of a violent crime in their home.

The Early Screening Project

Efficient and low-cost mass screening procedures are implemented in Stage One of the ESP to identify preschool and kindergarten children who may be at risk for behavioral disorders or problems (Walker et al., 1995). Figure 1 illustrates the screening and student identification processes involved in the multiple-gating procedure investigated in this study. The six highest-ranked children in this study on the externalizing and internalizing dimensions, respectively, as well as the two comparison (nonranked) children, were assessed by the teacher on both a Critical Events index and behavioral rating frequency indices.

Stage Two ESP assessments are more complex and intensive, and also expensive in terms of teacher time, but they are conducted on only a small subset of the total number of pupils screened in each classroom. In addition to their screening functions, Stage Two assessments define the specific content of each rated child's behavioral problems. Normative criteria on the Stage Two instruments are used to determine whether any of the rated children qualify for Stage Three ESP behavioral observations and parent rating assessments (Walker et al., 1995). Qualifying children are then observed directly in academic and playground settings, and their performance levels are compared to an ESP normative observation database for same-age and -sex peers.

Beginning in 1991, a series of studies was conducted on the ESP to assess its reliability and validity. The findings have been promising to date (Feil & Becker, 1993; Feil, Severson, & Walker, 1998; Feil, Walker, & Severson, 1995). Study participants (from 1991–1994) consisted of 2,853 children, aged 3 to 6 years, who were enrolled in typical and specialized programs. These children were from preschool and kindergarten classrooms in the following states (the number corresponds to children in the sample): California (517), Kentucky (687), Louisiana (386), Nebraska (65), New Hampshire (25), Oregon (220), Texas (612), and Utah (341).

The ESP reliability and validity data show strong results. The interrater reliability coefficients of most ESP measures are at least .80, which meets Salvia and Ysseldyke's (1988) guidelines for a screening instrument. Good psychometric standards have been attained despite the difficulties inherent in the assessment of young children (Martin, 1986). Validity studies show consistently high relationships to the following criterion measures: Conners Teacher Rating Scales (Conners, 1989), Preschool Behavior Questionnaire (Behar & Stringfield, 1974), and Child Behavior Checklist-Teacher Report (Achenbach, 1991). Correlations with these criterion measures were highly significant, ranging from .34 to .87, with most above .70.

Furthermore, a discriminant analysis provided a measure of the accuracy of the ESP with both specificity and sensitivity coefficients. Specificity and sensitivity are important criteria when choosing an assessment method (Elliot, Busse, & Gresham, 1993). Sensitivity is the percentage of true positives and specificity the true percentage of negatives (Schaughency & Rothlind, 1991). Results for the ESP show good sensitivity (62%) and excellent specificity (94%), leading to accurate assessments with a minimal risk in identifying a child who exhibits developmentally appropriate behavior.

The ESP has been found to be user friendly, and reports from staff users and reviewers have been positive regarding both its length and simplicity (Yoshikawa & Knitzer, 1997). One preschool director stated that she expects that use of the ESP increases the credibility of the staff when they make referrals to local early childhood special education programs. The ESP can make a positive difference in obtaining timely referral, diagnoses, and follow-through for preschool children showing emotional or behavioral problems (Yoshikawa & Knitzer, 1997).

Criterion-Related Measures

To compare the ESP to current assessment procedures for accuracy, teachers were asked to complete the Child Behavior Checklist (CBCL; Achenbach, 1991), and both parents and teachers completed the Social Skills Rating System (SSRS; Gresham & Elliot, 1990). The Child Behavior Checklist-Teacher Report Form was "designed to obtain teachers' reports of their students problems and adaptive functioning in a standardized format." The high psychometric quality of the CBCL is well documented (Gresham, 1985). In a review of the CBCL, Wilson and Bullock (1989) cited a .89 test–retest reliability and stated that the instrument correctly identified 76% of children referred for emotional or behavioral problems. Recently, Achenbach (1997) created a preschool version of the teacher questionnaire. Reliability estimates show strong results (e.g., mean test–retest coefficient of .84). The SSRS provides an excellent measure of peer-to-peer and parent- and teacher-related social skills as well as a comparison of parents' and teachers' perceptions of each item. Test–retest reliability and validity in studies of parents' and teachers' ratings show solid results (Gresham & Elliot, 1990). The SSRS Externalizing and CBCL Externalizing scores are highly correlated ($r = .75$).

Results

Table 2 presents ESP and concurrent measures (i.e., SSRS, CBCL), total number in sample, mean, standard deviation, and number of children exceeding cut-off criteria as defined by the authors of these instruments (Achenbach, 1997; Gresham & Elliot, 1990; Walker et al., 1995). Cut-off scores were set at 1 standard deviation from the mean for the ESP and SSRS and 1.5 standard deviations for the CBCL. ESP teacher ratings resulted in 43% to 52% of children exceeding cut-off criteria for externalizing (e.g., aggressive) measures and 25% to 34% for internalizing, with a smaller percentage exceeding the cut-off on the direct observation of peer social behavior (13%). Most of the criterion measures showed slightly smaller percentages (7%–25%) exceeding cut-off criteria, with SSRS parent reports (32%–57%)

TABLE 2
Means and Standard Deviations of Measures

Measure	N	*Mean*	SD	*Exceeding Cut-Off for Risk Status*
Early Screening Project				
Critical Events	126	1.58	1.97	54 (43%)
Aggressive	101	16.00	7.43	48 (48%)
Social Interaction	97	35.24	12.56	24 (25%)
Adaptive	103	29.08	7.42	35 (34%)
Maladaptive	103	20.98	8.74	54 (52%)
Observations of Antisocial & Nonsocial Behavior in Free Play	102	19.88%	17.26%	13 (13%)
Social Skills Rating Scale-*Teacher*				
Cooperation	79	13.34	3.71	13 (16%)
Assertion	95	10.54	4.40	23 (24%)
Self-Control	87	11.84	4.90	22 (25%)
Total Social Skills	71	36.89	12.09	13 (18%)
Social Skills Rating Scale-*Parent*				
Cooperation	76	11.62	3.02	22 (29%)
Assertion	76	13.64	3.19	24 (32%)
Responsibility	70	8.73	3.49	40 (57%)
Self-Control	78	11.47	3.16	36 (46%)
Total Social Skills	58	45.03	10.29	21 (36%)
Child Behavior Checklist-*Teacher*				
Anxious/Obsessive	91	1.90	2.33	12 (13%)
Depressed/Withdrawn	95	7.43	6.14	12 (13%)
Fears	94	.79	1.43	7 (7%)
Somatic Problems	94	.61	1.52	7 (7%)
Immature	94	3.92	3.60	8 (8%)
Attention Problems	93	7.16	7.58	14 (15%)
Aggressive Behavior	94	8.84	9.86	16 (17%)
Internalizing	89	9.75	8.45	28 (31%)
Externalizing	92	16.08	16.47	28 (30%)

and overall CBCL teacher reports of internalizing and externalizing, respectively (31% and 30%) having larger numbers exceeding cut-off criteria. Since the ESP is a screening measure, the ESP's authors used 1 standard deviation to make it more inclusive, thereby increasing the number of false positives but minimizing false negatives. The addition of the direct observations resulted in fewer children exceeding cut-off criteria, which would seem to increase sensitivity. While the ESP Stage Two teacher measures identify a larger percentage of children, the Stage Three Observations show increased specificity.

Correlation coefficients among the ESP and criterion measures showed good agreement between measures, moderate agreement across raters (parents and teachers), and an overall better relationship among externalizing subscales (see Table 3). Coefficients ranged from .91 to .83 between the ESP Aggressive and CBCL and the SSRS Externalizing (teacher measures) subscales and ranged from .53 to .44 between ESP Adaptive and Social Interaction and CBCL Internalizing subscale (teacher measures). Correlation coefficients between parents and teachers were highest with the SSRS Externalizing subscale (r = .38,

TABLE 3
Correlation Matrix of Selected Early Screening Project (ESP) and Concurrent Measures Among Parents and Teachers

	ESP Adaptive Behavior	*ESP Critical Events*	*ESP Maladaptive Behavior*	*ESP Social Interaction Scale*	*CBCL Teacher Externalizing*	*CBCL Teacher Internalizing*	*SSRS Parent Externalizing*	*SSRS Parent Internalizing*	*SSRS Teacher Externalizing*	*SSRS Teacher Internalizing*
ESP Aggressive Behavior	−.70[c]	.34[c]	.85[c]	−.12	.91[c]	.38[c]	.31[c]	.12	.83[c]	.24[a]
ESP Adaptive Behavior		−.32[b]	−.75[c]	.45[c]	−.71[c]	−.44[c]	−.28[a]	−.06	−.64[c]	−.22[a]
ESP Critical Events			.24[a]	−.18	.38[c]	.53[c]	.31[b]	.16	.23[a]	.27[b]
ESP Maladaptive Behavior				−.12	.88[c]	.41[c]	.36[b]	.07	.83[c]	.25[a]
ESP Social Interaction					−.17	−.45[c]	−.17	−.30[a]	−.06	−.24[a]
CBCL Teacher Externalizing						.44[c]	.41[c]	.07	.85[c]	.31[b]
CBCL Teacher Internalizing							.06	.21	.33[b]	.66[c]
SSRS Parent Externalizing								.45[b]	.38[c]	.05
SSRS Parent Internalizing									−.01	.03
SSRS Teacher Externalizing										.33[b]

Note. CBCL, Child Behavior Checklist; SSRS, Social Skills Rating Scale.
[a]$p < .05$
[b]$p < .01$
[c]$p < .001$

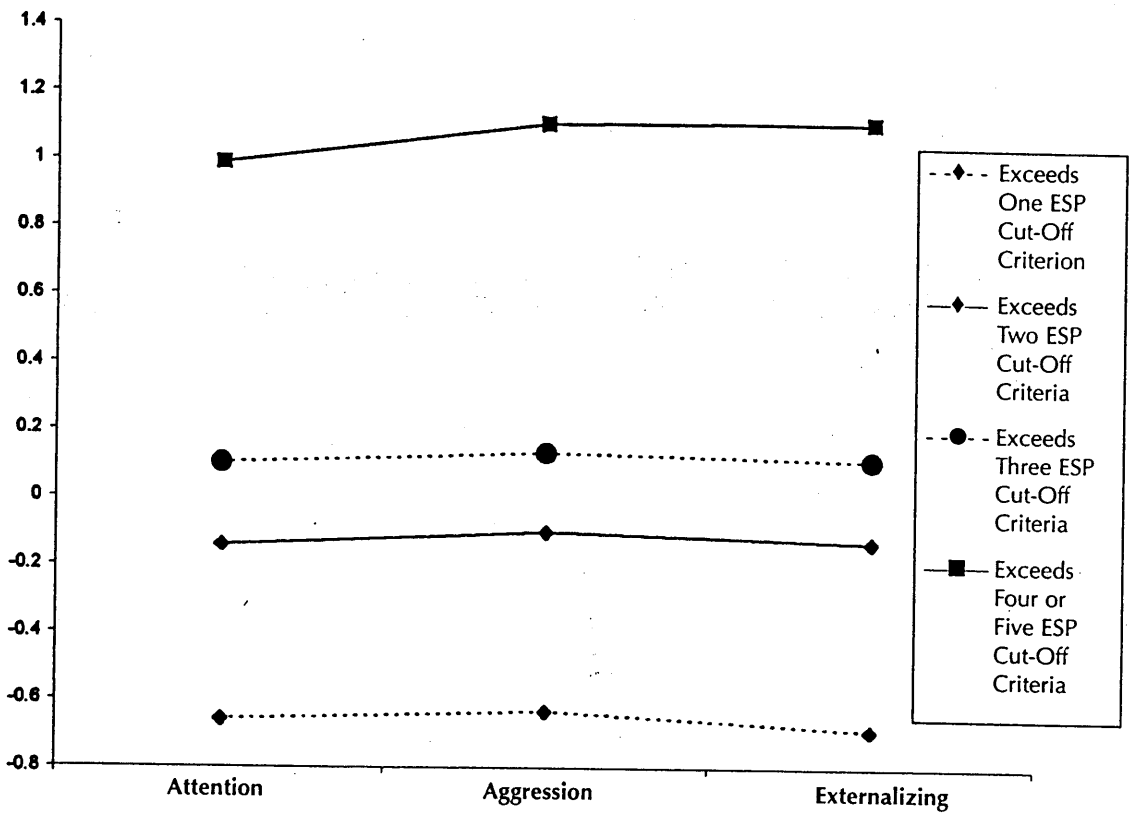

FIGURE 2. Mean Z-Scores on CBCL-TRF Externalizing Subscales by Groups Exceeding ESP Cut-Off Criteria

$p < .01$) and between the SSRS Parent Internalizing and the ESP Social Interaction Scales ($r = -.30$, $p < 05$).

When a child's score exceeds a subscale's cut-off criterion, the child could be considered at risk. If a child's scores exceed criteria on several subscales, the child could be considered more at risk. A measure of severity can be created by counting the number of subscales on which a child's scores exceed cut-off criteria. For example, scores exceeding criterion on four subscales would be considered more at risk than one score exceeding a cut-off criterion. On the ESP, 32 out of 126 children (25%) exceeded cut-off criteria on one subscale, 43 children (34%) exceeded cut-off criteria on two or three subscale measures, and 25 children (25%) exceeded cut-off criteria on four or more. These four groups' means on each CBCL externalizing measure show remarkable stability (see Figure 2). That is, each line is relatively flat with good separation, especially among the children who are most at risk (i.e., exceeding four or more ESP cut-off criteria).

Each ESP measure was analyzed by ethnic status using chi-square analysis of ethnicity by exceeding cut-off criteria. Ethnic status, as well as an aggregated White/non-White variable to maintain better statistical power, was used to test for differences among ethnic groups. These test results were encouraging, with no significant differences among ethnic groups, even among teacher ratings and observations. The Critical Events Index had the only *p*-values below .10 ($p = .06$), having *fewer* non-White students in the cut-off range.

Discussion

Culture and Screening

Over all, we were encouraged by the results obtained with respect to the cross-cultural nature of the Early Screening Project. Our results showed no significant differences in the number of referrals when using the ESP among ethnic groups. These results are preliminary,

and it is possible that the number of participants was not great enough for sufficient statistical power to detect differences. Therefore, we currently are conducting additional cross-cultural studies and analyses with greater numbers of participants to increase the chance of detecting a meaningful difference. The lack of differences could also be due to the fact that since Head Start staff members are frequently of the same ethnicity as the families they are serving, the opportunity for cross-cultural differences in the teacher rating portions do not exist. In fact, some teachers may rate their children high because of a type of halo effect due to a high specificity but lower sensitivity. Therefore, the ethnicity of the teachers and/or observer may be an important factor in the failure to detect statistical differences among ethnic groups.

Lessons Learned

Our team encountered several challenges that impacted the results of this study. In our search for diverse populations, we were required to cover a large section of Oregon, which made data collection difficult. In addition, program variations among Head Start programs made data collection individualization a necessary requirement. While Head Start has been characterized as a single program entity, each Head Start program is independent, with different schedules and procedures. Since each program site was essentially unique in this regard, much research staff time was spent tailoring our assessment practices to fit the individual programs' requirements and characteristics.

We found that targeted recruiting was not successful for a Head Start population. We received many anecdotal comments from staff and parents that the targeted recruitment, even of the nonranked comparison children, seemed to produce fears of stigmatization. We also found that, with Head Start samples, developing relationships at all program levels is an important enabling factor that contributes substantially to the success of the research study. Participating Head Start staff spend a considerable amount of time overcoming barriers in order to engage families. Being a federal program makes working with parents who are frequently distrustful of government programs quite difficult. Therefore, the targeted recruitment process often made the Head Start teachers feel reticent to promote a program that put them in the uncomfortable situation of having to explain their initial ranking and eligibility status. Most teachers did not engage parents directly about the research project; most recruitment took place via informed consent letters from the research project, which was not effective. Whether or not a teacher supports and therefore promotes the program has great impact.

For our next cohorts, to avoid any possible stigmatization and to improve the representativeness of our sample, we plan to invite all families enrolled in participating Head Start classrooms to participate. While increasing costs (i.e., staff time, incentive payments), we hope to reduce the stigma associated with an invitation to participate in the project. We do expect to recruit more comparison families using this method, but we are afraid that the families who are most at risk will decline participation due to the pressures of coping with a chaotic and stressful environment. We will continue to interview parents as to neighborhood violence and context, hoping to find more salient variables (e.g., parenting) that predict behavioral problems rather than relying on ethnicity or income.

Future Directions

We were encouraged by the results showing no differences among participating ethnic groups. The question remains, however, as to whether this lack of differences holds up in regard to long-term outcomes (e.g., accessing specialized services, retention). A longitudinal study could help to identify the salient characteristics of the ESP that could increase sensitivity. As part of this project, we are committed to longitudinal research in order to evaluate the ESP's predictive validity among representatives of different ethnic groups. Our hope is that teachers and direct observations can identify children reliably based on their behavior and not their ethnicity.

Conclusion

The ESP can be used as part of a best-practice approach in early detection and intervention programs for school adjustment problems. Preschool programs, facing increasing requirements and demands (e.g., Child Find), need to maximize their resources within a proactive and fair system. Our results suggest

the possibility that young children of all ethnic groups are being underserved or underidentified.

The proactive nature of the ESP provides for assessment of adjustment problems for all children in a classroom, not simply those children with externalizing behaviors whose behavior has become so aversive that a teacher will refer them for evaluation. The ESP screens for internalizing characteristics (e.g., socially withdrawn) as well, which frequently are overlooked because they do not disrupt classroom activities or pressure the teacher's child-management skills.

The ESP can minimize the time spent on and cost of preschool assessments while increasing accuracy over currently used screening instruments and approaches. We believe that practitioners can use the ESP with diverse populations and feel confident that cultural bias will not have a great impact on their ratings.

References

Achenbach, T. M. (1991). *Manual for the Teacher's Report Form and 1991 Profile.* Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T. M. (1997). *Guide for the Caregiver-Teacher Report Form for Ages 2–5.* Burlington: University of Vermont, Department of Psychiatry.

Achenbach, T., & Edelbrock, C. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin, 85,* 275–301.

Artiles, A. J., & Trent, S. C. (1994). Overrepresentation of minority students in special education: A continuing debate. *Journal of Special Education, 27,* 410–438

Behar, L., & Stringfield, S. (1974). *Manual for the Preschool Behavior Questionnaire.* Durham, NC: Behar.

Bersoff, D. N. (1980). P. V. Riles: *Legal perspective. School Psychology Review, 9,* 112–122.

Bower, E. M. (1982). Defining emotional disturbance: Public policy and research. *Psychology in the Schools, 19*(1), 55–60.

Brandenburg, N. A., Friedman, R. M., & Silver, S. E. (1990). The epistemology of childhood psychiatric disorders: Recent prevalence findings and methodologic issues. *Journal of the American Academy of Child & Adolescent Psychiatry, 29,* 76–83.

Carlson, P., & Stephens, T. (1986). Cultural bias and identification of behaviorally disordered children. *Behavioral Disorders, 11,* 191–198.

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

Council for Children with Behavioral Disorders (CCBD). (1987). Position paper on definition and identification of students with behavioral disorders. *Behavioral Disorders, 13,* 9–19.

Craig, P. A., Kaskowitz, D. H., & Malgoire, M. A. (1978). *Teacher identification of handicapped pupils (ages 6–11) compared with identification using the other indicators* (Vol. 2). Menlo Park, CA: Stanford Research Institute, Educational Policy Research Center.

Elliot, S. N., Busse, R. T., & Gresham, F. M. (1993). Behavior Rating Scales. *School Psychology Review, 22,* 313–321.

Feil, E. G., & Becker, W. C. (1993). Investigation of a multiple-gated screening system for preschool behavior problems. *Behavioral Disorders, 19,* 44–55.

Feil, E. G., Severson, H. H., & Walker, H. M. (1998). Innovations in the screening of young children for emotional/behavioral delays. *Journal of Early Intervention, 21,* 252–266.

Feil, E. G., Walker, H. M., & Severson, H. H. (1995). The Early Screening Project for young children with behavior problems. *Journal of Emotional and Behavioral Disorders, 3*(4), 194–202.

Forness, S. R., & Knitzer, J. (1992). A new proposed definition and terminology to replace "serious emotional disturbance" in Individuals with Disabilities Education Act. *School Psychology Review, 21*(1), 12–20.

Gerber, M. M., & Semmel, M. I. (1984). Teacher as imperfect test: Reconceptualizing the referral process. *Educational Psychologist, 19*(3), 137–148.

Greenwood, C., Walker, H. M., Todd, N., & Hops, H. (1979). Selecting a cost-effective device for the assessment of social withdrawal. *Journal of Applied Behavior Analysis, 12,* 639–652.

Gresham, F. M. (1985). Behavior disorder assessment: Conceptual, definitional, and practical considerations. *School Psychology Review, 14,* 495–509.

Gresham, F. M., & Elliott, S. (1990). The Social Skills Rating System (SSRS). Circle Pines, MN: American Guidance Service.

Hawkins, J. D., Von Cleve, E., & Catalano, R. F. (1991). Reducing early childhood aggression: Results of a primary prevention program. *Journal of the American Academy of Child and Adolescent Psychiatry, 30,* 208–217.

Huntze, S., & CCBD Subcommittee on Terminlogy. (1985). Statement to support replacing the term *seriously emotionally disturbed* with the term *behaviorally disordered* as a descriptor for children and youth who are handicapped by their behavior (CCBD position paper). *Behavioral Disorders, 10,* 167–174.

Jenson, W. (1984). *Severely emotionally disturbed versus behavior disorders: Consideration of a*

label change. Salt Lake City: University of Utah, Department of Educational Psychology.

Kauffman, J. M. (2001). *Characteristics of emotional and behavioral disorders of children and youth* (7th ed.). Upper Saddle River, NJ: Merrill/ Prentice Hall.

Kazdin, A. (1987). *Conduct disorders in childhood and adolescence.* London: Sage.

Koot, H. M., Van Den Oord, E., & Boomsma, D. I. (1997). Behavioral and emotional problems in young preschoolers: Cross-cultural testing of the validity of the Child Behavioral Checklist/ 2–3. *Journal of Abnormal Child Psychology, 25,* 183.

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,* 42–48.

Loeber, R., Dishion, T. J., & Patterson, G. R. (1984). Multiple gating: A multi-stage assessment procedure for identifying youths at risk for delinquency. *Journal of Research in Crime and Delinquency, 21*(1), 7–32.

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

Ross, A. (1980). *Psychological disorders of children: A behavioral approach to theory, research and therapy* (2nd ed.). New York: McGraw-Hill.

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

Schaughency, L. A., & Rothlind, J. (1991). Assessment and classification of attention deficit hyperactive disorders. *School Psychology Review, 20*(2), 187–202.

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

Sugai, G., & Maheady, L. (1988). Cultural diversity and individual assessment for behavioral disorders. *Teaching Exceptional Children, 21,* 28–31.

Sugai, G. M., Maheady, L., & Skouge, J. (1989). Best assessment practices for students with behavior disorders: Accommodation to cultural diversity and other individual differences. *Behavioral Disorders, 14,* 263–278.

U.S. Department of Education. (1999). *Digest of education statistics* (NCES Publication No. 2000–031). Washington, DC: National Center for Educational Statistics, Office of Educational Research and Improvement.

U.S. Department of Health and Human Services. (1998). *Head Start Program Performance Measures: Second progress report.* Washington, DC: Research, Demonstration and Evaluation Branch and the Head Start Bureau, Administration on Children, Youth and Families.

Walker, H. M., Horner, R. H., Sugai, G., Bullis, M., Sprague, J. R., Bricker, D., & Kaufman, M. J. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional and Behavioral Disorders, 4,* 193–256.

Walker, H. M., & Severson, H. H. (1990). *Systematic Screening for Behavior Disorders (SSBD): User's guide and technical manual.* Longmont, CO: Sopris West.

Walker, H. M., Severson, H. H., & Feil, E. G. (1995). *The Early Screening Project: A proven child-find process.* Longmont, CO: Sopris West.

Wilson, M. J., & Bullock, L. M. (1989). Psychometric characteristics of behavior rating scales: Definitions, problems, and solutions. *Behavioral Disorders, 14,* 186–200.

Wood, F., Smith, C., & Grimes, J. (Eds.). (1985). *The Iowa Assessment Model in Behavioral Disorders: A training manual.* Des Moines: Iowa Department of Public Instruction.

Yoshikawa, H., & Knitzer, J. (1997). *Lessons from the field: Head Start mental health strategies to meet changing needs.* New York: National Center for Children in Poverty.

Ysseldyke, J., Algozzine, B., & Epps, S. (1983). A logical and empirical analysis of current practices in classifying students as handicapped. *Exceptional Children, 50,* 160–166.

Ysseldyke, J., Christenson, S., Pianta, B., & Algozzine, B. (1983). An analysis of teachers' reasons and desired outcomes for students referred for psychoeducational assessment. *Journal of Psychoeducational Assessment, 1,* 73–83.

Zigler, E., Taussig, C., & Black, K. (1992). Early childhood intervention: A promising preventative for juvenile delinquency. *American Psychologist, 47,* 997–1006.

AUTHORS' NOTE:

The authors would like to acknowledge the support of this research by Grant No. 97-JN-FX-0022 from the Administration for Children, Youth and Families, Department of Health and Human Services. We thank the staff and parents of Warm Springs Tribal Head Start, Klamath Family Head Start, Mid Columbia Children's Council, South Coast Head Start, and Portland Public Schools Head Start program for their time, trust, and cooperation in bringing this project to fruition. Direct correspondence and reprint requests to Edward Feil, Oregon Research Institute, 1715 Franklin Boulevard, Eugene, OR 97403. Ph: 541/484-2123; Fax: 541/484-1108; E-mail: edf@ori.org

AUTHORS:

EDWARD G. FEIL, Research Associate, Institute on Violence and Destructive Behavior, and Research Scientist, Oregon Research Institute; HILL WALKER, Professor, Institute on Violence and Destructive Behavior; HERBERT SEVERSON, Associate Professor, Institute on Violence and Destructive Behavior, and Research Scientist, Oregon Research Institute; and ALISON BALL, Research Associate, Institute on Violence and Destructive Behavior, College of Education, University of Oregon, Eugene

MANUSCRIPT:

Initial Acceptance: 8/11/00
Final Acceptance: 9/19/00