Using Different Measures, Informants, and Clinical Cut-Off Points to Estimate Prevalence of Emotional or Behavioral Disorders in Preschoolers: Effects on Age, Gender, and Ethnicity

Edward G. Feil and Jason W. Small
Oregon Research Institute

Steven R. Forness
University of California–Los Angeles

Loretta A. Serna
University of New Mexico

Ann P. Kaiser and Terry B. Hancock
Vanderbilt University

Jeanne Brooks-Gunn
Columbia University

Donna Bryant, Janis Kuperschmidt, and Margaret R. Burchinal
University of North Carolina

Cheryl A. Boyce
National Institute of Mental Health

Michael L. Lopez
National Center for Latino Child & Family Research

ABSTRACT: The early identification and remediation of emotional or behavior disorders are high priorities for early-childhood researchers and are based on the assumption that problems such as school failure can be averted with early screening, prevention, and intervention. Presently, prevalence, severity, and topography of mental health needs among low-income preschoolers and their families have not been well documented. Tools for screening and intervention for behavior problems in preschool children are few and many of those tools have not been studied within diverse Head Start systems.

In this study, five instruments of symptoms and functional impairment, completed by teachers and two completed by parents, were obtained on a sample of 1,781 Head Start children from diverse racial and ethnic backgrounds from the Head Start Mental Health Research Consortium. Clinical cut-off scores were used to identify children who could be considered at relatively serious risk for emotional or behavioral disorders. At-risk classifications using clinical cut-offs at both 1.0 and 1.5 standard deviations for each measure were examined singly and in combination and then compared to the overall sample for age, gender, and ethnicity. Identification of children considered at risk ranged from a low of 1% to a high of 38%, with evidence of differential effects on age, gender, or ethnicity for some individual measures, but these tended to diminish when combinations of symptom and impairment measures were used. Implications for choosing instruments to establish eligibility for emotional or behavioral disorders in preschoolers are discussed.

■ The prevalence of emotional or behavioral disorders (EBD) in preschool children has not been well established. Roberts, Attkisson, and Rosenblatt (1998) reviewed 50 epidemiologic studies in children's mental health, of which only 10 involved preschool samples. Mean preschool prevalence was found to be 10.2% with a median prevalence of 8.3%. Qi and Kaiser (2003) reviewed 30 recent studies on behavior problems of preschool children from

low-income families and reported a mean prevalence of approximately 30%. Differences between these estimates may be at least partially accounted for by generally higher risk for psychopathology in children from low-income families, as in the Qi and Kaiser review, and by somewhat stricter or more conservative diagnostic criteria used in mental health epidemiology, as in the Roberts et al., review. Of the 40 studies in both reviews, there was no overlap, that is, none of the studies in the former review appeared in the latter or vice versa. Range of prevalence across studies in the two reviews was from 3.6% to 57%. Only 8 of the 40 studies had samples larger than 500 preschoolers.

An emotional or behavioral disorder is usually considered to be present in a preschooler when he or she displays signs of psychopathology that lead to problems in everyday life. *Caseness*, the determination that a preschooler actually has a disorder, is usually ascertained through a process of obtaining assessments from significant adults in the child's environment. Ascertainment of EBD is determined by a number of measures, but there is some consensus that symptoms of emotional or behavioral disorders must be present and that impairment in functioning must also accompany these symptoms (Kauffman, 2005; Pennington, 2002). If a child is reported to have only symptoms but remains relatively unimpaired in his or her social or personal relations, then caseness is not necessarily evident. If impairment seems present but no symptoms are reported, then a case of EBD cannot necessarily be determined in this instance either, since the child's impairment may be due to some other disorder such as mental retardation.

Presence of *symptoms* in preschoolers is usually determined by rating scales completed by parents and/or teachers, structured diagnostic interviews, developmental histories, child observations, or the like (Campbell, 2002). Functional *impairment* has not been as well studied in regard to its assessment, but measures have included social skill ratings, adaptive behavior scales, developmental testing, direct observation of social interactions, or the like (Walker, Ramsey, & Gresham 2004). In actual practice, many professionals focus only on symptom ratings assuming that, since a referral has been made, a *de facto* functional impairment has thereby been established. With almost all of the above measures, however, questions arise about diagnostic validity and reliability based on source of informant such as parent or teacher, severity criteria such as 1.0 versus 1.5 standard deviation cut-off points for rating scales, breadth of assessment instruments such as those for single disorders versus clusters or broad bands of disorders, and particular combinations of symptom and impairment measures.

Developing an algorithm to determine whether or not a child is at risk for a disorder has always been problematic, but most of this research has been done on school-age children (Forness & Kavale 1997, 2000; Witmer, Doll, & Strain, 1996). Current research, using epidemiological samples of children over 6 years of age, has focused on a number of components of the standard diagnostic procedure. Offord and his colleagues (Boyle et al., 1996, Offord et al., 1996) examined symptom checklists for EBD diagnosis and found that relatively minor increases in threshold criteria for individual items reduced teacher identification rates for some disorders (e.g., ADHD) by more than two thirds. They also confirmed that identification rates determined by teacher checklists not only could differ by a factor of 2 or 3 from those determined by parent checklists, but that even when rates were similar within a specific diagnosis, such as conduct disorder (CD), different types of children appeared in groups scored by teachers than in groups scored by parents. Shaffer et al. (1996) demonstrated that, when only checklist scores were used for these two disorders, a combined identification rate of ADHD or CD of around 6% resulted, but when a measure of functional impairment was added to the diagnostic criteria, the rate dropped to less than 2%. The rate of depression in this same study was reduced from approximately 4% to 1% when a measure of impairment was required in addition to symptom checklist scores. Significant reductions were also found for more subtle disorders, such as obsessive compulsive disorder (OCD), with a median rate of about 30% of adolescents reporting selected symptoms; however, fewer than 4% identified if a measure of impairment was added (Apter et al., 1996). These findings are similar to those of several other studies on both cross informant ambiguity and use of impairment levels in establishing a diagnosis (Achenbach, McConaughy, & Howell, 1987; Biederman, Mick, Faraone, & Burbank, 2001; Bird, Gould, & Staghezza, 1992; Bird et al., 1990; Culp, Howell, Culp, & Blankemeyer, 2001; Fergusson

& Horwood, 1995; Kramer, Phillips, Hargis, & Miller, 2004; Mesman & Koot, 2000; Mitsis, McKay, Schulz, Newcorn, & Halperin, 2000; Piacentini, Cohen, & Cohen, 1992; Razzino, New, Lewin, & Joseph, 2004).

Cluett and her colleagues (Cluett et al., 1998) attempted to determine the impact of different diagnostic criteria on identification of children at risk for EBD by systematically varying the respondent for the symptom checklist (parent and/or teacher) and the type of functional impairment (poor academic achievement and/or poor social skills) required for diagnosis. The sample included 3,694 children in second grade, who were part of a longitudinal sample of Head Start graduates. Clinical cut-off points on behavior problem checklists from parents and teachers, on social skills ratings from parents and teachers, and on individual measures of academic achievement were serially applied to the sample to determine the rate of EBD identification with various combinations of instruments. The 13 different diagnostic combinations identified from 1.1% to 27.5% of the sample. Four of these combinations resulted in significant differential effects on gender and ethnicity. Two relatively unbiased combinations were selected for further study and produced very different rates of children actually identified for special education by school professionals (Forness et al., 1998).

Careful studies have recently examined emotional or behavioral disorders with preschool onset but have not systematically varied measures of symptoms or impairment to determine variations in gender or ethnicity (Campbell, 2002; Lavigne et al., 1998a, 1998b; Stormont, 2000). Only Serna and her colleagues (Serna, Nielsen, Mattern, Schau, & Forness, 2002) have examined these issues with preschool children. Using a methodology similar to that of Cluett et al. (1998), they found a similarly wide range of EBD identification among various measures, but their findings were somewhat compromised by the fact that their sample was relatively small and not as ethnically diverse.

The present study attempts to replicate, at least partially, both the Cluett et al. (1998) and the Serna et al. (2002) studies with a large sample of preschoolers and addresses a number of questions. One such question concerns current practice in preschool diagnosis and what impact the introduction of different measures has on potential identification of preschool children at risk for EBD. Does prevalence change substantially if the teacher, the parent, or both complete the symptom checklist for EBD? Does the requirement for functional impairment change the rate substantially depending on whether general social interaction deficits or specific social skills deficits are required? Does the identification rate change substantially if a functional deficit is present as rated by the teacher in a classroom setting and if a social skills deficit must also be present as rated by the parent in a home setting? Do various measures differentially impact age, gender, or racial/ethnic characteristics of children identified as EBD? These and related questions are addressed herein. While we have used particular sets of measures in this study, such symptom or impairment ratings may nonetheless serve as analogs of other measures used to determine caseness in preschoolers and thus possibly extend our findings to the general process of ascertaining emotional or behavioral disorders in preschoolers.

Methods

Participant Selection

Participants were selected from four of the five sites of the Head Start Mental Health Research Consortium: a cross-site, multi-year study designed to examine mental health issues in children enrolled in Head Start preschool settings (Boyce, Hoagwood, Lopez, & Tarullo, 2000; Lopez, Tarullo, Forness, & Boyce, 2000). The consortium consists of sites at Vanderbilt University, University of North Carolina (UNC), University of New Mexico (UNM), University of Oregon (U of O), and Columbia University. For the current study, differences in measurement instruments for EBD precluded inclusion of participants from Columbia University. Three of the four included sites (Vanderbilt, UNC, and UNM) conducted intervention studies and thus used only pretest data for participants reported here in order to avoid any artifacts due to intervention. The fourth site (U of O) employed a multiple cohort longitudinal design with no intervention. In order to accommodate design differences across sites, a cross-sectional dataset was constructed, consisting of pre-intervention data from Vanderbilt, UNC, and UNM and assessment data from the University of Oregon. The sample (n = 1,759) includes data for all children, age 3 to 5 years, from five major

racial/ethnic categories (African American, Hispanic, Caucasian, Native American, and Asian), for whom at least one parent or teacher measure was completed.

Head Start classroom selection procedures differed across sites. The Vanderbilt site selected only children who were not yet 4-years old at the beginning of the study. From this site, 1,057 participants were selected over a 3-year period from 31 classrooms. Approximately 95% of the 38 teachers involved were from African American backgrounds. The UNC site assessed 361 3- to 5-year-old children over a 2-year period. The study involved 37 teachers, 92% of whom were African American from 37 Head Start classrooms. The UNM site assessed 349 children from 17 classrooms over a 3-year period. The 17 teachers were all of Hispanic origin. The Oregon site involved 200 classrooms studied over a 3-year period. For the first cohort, six children were selected from target classrooms: two nominated by the teacher as having externalizing problems, two nominated for internalizing problems, and two nominated to be in the normal range. About one third of the 260 children were chosen this way. For the second and third cohorts, all children in the Head Start classrooms were selected for participation to increase the recruitment rate. The teachers in this site were approximately 8% Hispanic, 8% Native American, and 76% Caucasian.

Table 1 provides the demographic breakdown of the Head Start participants for the total sample and each site by age, gender, and ethnic or racial background. Three of the four sites differ on age. No site differences were present for gender. In terms of racial or ethnic background, there are no statistically significant differences between the Vanderbilt and UNC samples, but some differences occur between these two samples and the UNM and Oregon samples.

Symptom Measures

The current study includes four symptom measures. The first two, taken from the *Social Skills Rating System* (SSRS; Gresham & Elliott, 1990), are parent-reported problem behavior (SSRS-PB-P) and teacher-reported problem behavior (SSRS-PB-T). Both scales have 10 items and rate behaviors on a frequency measure of 0 (never), 1 (sometimes), or 2 (very often). Clinical cut-off raw scores can be found in *Table 2* for both 1.0 and 1.5 standard deviations. For example, a score of 8 or above (for girls) and 11 or above (for boys) for the teacher form (SSRS-PB-T) and 10 and above for both girls and boys for the parent form (SSRS-PB-P) were considered significant for symptoms at the 1.0 standard deviation cut-off. For the SSRS system, scores within 1.0 standard deviation are considered to be in the "average range." In reference to the frequency of behaviors, children above the 1.0 cut-off on the problem behavior scale are described as having "more than average behavior problems" while those below the 1.0 cut-off on the social skills scale are described as possessing "fewer than average social skills" (Gresham & Elliott, 1990). We added a cut-off of 1.5 standard deviations for purposes of this study in order to examine children with more severe symptoms or impairments consistent with similar cut-offs for the other measure used, as described below. For the teacher form, coefficient alpha reliability was reported as 0.82 and, for the parent form, coefficient alpha was 0.73. The SSRS was standardized on a nationally representative

TABLE 1
Demographic Variables by Site

Variable	*Sample (n = 1,759)*	*Vanderbilt (n = 1,057)*	*UNC (n = 154)*	*UNM (n = 288)*	*U of O (n = 260)*
Age *M* (*SD*)	3.98 (0.51)	3.78 (0.43)$_a$	4.49 (0.30)$_b$	4.24 (0.50)$_c$	4.19 (0.51)$_c$
n (%) Male	944 (53.7)	557 (52.7)$_a$	78 (50.6)$_a$	155 (53.8)$_a$	154 (59.2)$_a$
n (%) African American	1109 (63.0)	938 (88.7)$_a$	137 (89.0)$_a$	22 (7.6)$_b$	12 (4.6)$_b$
n (%) Hispanic	330 (18.8)	19 (1.8)$_a$	3 (1.9)$_a$	234 (81.3)$_b$	74 (28.5)$_c$
n (%) Caucasian	223 (12.7)	98 (9.3)$_a$	14 (9.1)$_a$	19 (6.6)$_a$	92 (35.4)$_b$
n (%) Native American	85 (4.8)	0 (0.0)$_a$	0 (0.0)$_a$	11 (3.8)$_a$	74 (28.5)$_b$
n (%) Asian	12 (0.7)	2 (0.2)$_a$	0 (0.0)$_a$	2 (0.7)$_a$	8 (3.1)$_b$

Note. Means or percentages with the same subscript indicate that the sites did not differ significantly at the 0.05 level.

sample of 4,170 children, including children from low-income and racially and ethnically diverse families.

The third and fourth symptom instruments came from the *Early Screening Project*, developed by Hill Walker and colleagues (Feil & Becker, 1993; Walker, Severson, & Feil, 1995). It was designed for teacher use in preschool environments with children ages 3 to 5 years. The ESP consists of three stages. It is referred to as a multiple gating screening tool in that it includes a series of more intensive and precise assessments, or gates; however, only the second stage of gates involves symptom or functional impairment measures. This stage includes two symptom measures: The Critical Events Index (ESP-CEI) and the Aggressive Behavior Scale (ESP-ABS). The Critical Events Index is a 16-item teacher checklist of behavioral events having high intensity and salience but relatively low frequency (e.g., fire setting, physical assault of another). Scores on the ESP-CEI range from 0 to 16; they reflect the total number of critical events exhibited by the child within the past 6 months. At 1.0 standard deviation, a score of 2 or above on the ESP-CEI is the clinical cut off for both boys and girls. The Aggressive Behavior Scale includes nine items such as "disturbs classroom activities," scored on a 5-point Likert scale. Scores of 15 (for boys) or 14 (for girls) on the ESP-ABS are the clinical cut-off points at 1.0 standard deviation, with scores above 17 or 15, respectively, for the 1.5 clinical cut-off. Standardized scores on the ESP scales are described according to increasing levels of risk. Children scoring above the clinical cut-off of 1.0 standard deviation are identified as "at risk" with higher levels of deviation from the norm resulting in more severe levels of risk. Individuals above the 1.5 standard deviation cut-off are described as "high risk." The ESP was normed on a nationally representative sample of 2,853 children age 3 to 6 years, including a proportionately representative group of Head Start children. The interrater reliability coefficients for all of the ESP stage 2 measures were 0.80 or above.

Functional Impairment Measures

Three functional impairment measures are included in the current study. The first was also from the ESP, described previously. The Social Interaction Scale (ESP-SIS) consists of eight items, such as "verbally responds to peer's initiation," scored on a 7-point Likert scale. At 1.0 standard deviation, boys with scores of 26 or less and girls with scores of 31 or less are considered at risk, with scores of 19 and 24, respectively, at the 1.5 clinical cut-off. As noted above, reliability coefficients were above 0.80 for all ESP stage 2 measures.

The remaining impairment measures are from the SSRS. The parent-completed Social Skills subscale (SSRS-SS-P), consisting of 39 items scored on a 3-point scale (never, sometimes, and very often), assesses social functioning with respect to the home environment. Items scored on the same 3-point scale described previously assess social functioning with respect to the home environment. The Social Skills subscale, teacher form (SSRS-SS-T), consists of 30 items measuring social functioning in the classroom environment. For the parent subscale, the 1.0 standard deviation clinical cut off (indicating "fewer" of the desired social skills) is any score below 44 for girls and any score less than 40 for boys. Girls scoring below 32 and boys scoring below 24 on the teacher subscale meet clinical cut-off criteria at 1.0 standard deviation. Scores at the 1.5 clinical cut-off are provided in *Table 2*.

Statistical Analyses

Dichotomous outcome variables differentiating at risk from normative children were computed for each measure at both cut-offs. Separate logistic regression analyses were conducted to determine whether the demographic predictor variables (age, gender, and ethnicity) were significantly associated with at-risk status. All models were adjusted for site differences as well as differences in sampling methodology. Since numerous tests were performed, statistical significance was based on $p < 0.001$ to offset the inflated type I error rate. Note that separate analyses were also done for each measure and combination of measures for clinical cut-offs at 1.0 and 1.5 standard deviations, based on the literature review already cited.

Results

Table 2 presents means and standard deviations for boys and girls as well as the clinical cut-offs at 1.0 and 1.5 standard deviations for each symptom and functional impairment measure. Note that, although only teacher and not parent forms exist for the ESP, we have added "T" to each of the

ESP measures to denote that these are teacher measures, for ease of interpretation in this and all subsequent tables. For all three functional impairment measures (SSRS-SS-P, SSRS-SS-T, and ESP-SIS-T) and three of the four symptom measures (SSRS-PB-P, SSRS-PB-T, ESP-CEI-T), mean scores for both girls and boys were in the normative range. The one exception was the ESP aggressive behavior scale (ESP-ABS-T), for which girls' mean scores were in the normative range, but boys' scores were above the 1.0 standard deviation clinical cut-off.

Table 3 depicts pair-wise correlations for the 4 symptom and 3 functional impairment measures studied. Correlations among most of the measures are statistically significant; but only a few are at a level in which they would account for more than 10% of the variance between measures. This handful of correlations is primarily between selected symptom measures or between selected functional impairment measures, although a few are inverse correlations between symptom and impairment measures.

Tables 4 and *5* present prevalence rates for each functional impairment and symptom measure based on the clinical cut-offs of 1.0 and 1.5 standard deviations, respectively, and describe (according to mean age, percent male, and ethnicity) the types of children identified as at risk. Clinical cut-offs of 1.0 standard deviation produced prevalence rates ranging from 14.8% to 39.2%, and a cut-off of 1.5 standard deviations produced rates ranging from 7.8% to 34.8%. At both cut-offs, the SSRS parent social skills rating produced rates (38.8% and 19.0%, respectively) that were more than double those of the SSRS parent problem behavior rating (14.8% and 7.8%, respectively). The SSRS teacher ratings produced relatively similar patterns of differences with higher prevalence rates for impairment as opposed to symptom measures. The ESP measures clustered together somewhat at the 1.0 cut-off (23.9% to 39.2%), but produced dissimilar rates at the 1.5 cut-off with the two symptom measures, ESP-ABS-T and ESP-CEI-T, identifying rates of 34.8% and 13.0%, respectively. Overall, the SSRS problem behavior scale for both parents and teachers produced the lowest rates across both cut-offs. The measure producing the highest at-risk rates differed somewhat for each cut-off. At the 1.0 cut-off, the ESP aggressive behavior scale, the SSRS parent social skills rating, and

TABLE 2
Preschoolers' Mean Raw Scores and Clinical Cut-Offs (Age 3–5)

Measure	*N*	*Mean*	*SD*	*Clinical Cut-Off* 1 SD	1.5 SD
Symptom Measures					
SSRS Problem Behavior (SSRS-PB-P)					
Girls	660	5.64	3.37	≥ 10	≥ 11
Boys	752	6.13	3.44	≥ 10	≥ 12
SSRS Problem Behavior (SSRS-PB-T)					
Girls	753	4.04	3.90	≥ 8	≥ 10
Boys	870	5.55	4.45	≥ 11	≥ 13
ESP Aggression (ESP-ABS-T)					
Girls	271	13.22	6.19	≥ 14	≥ 15
Boys	408	16.72	8.22	≥ 15	≥ 17
ESP Critical Events Index (ESP-CEI-T)					
Girls	306	0.91	1.84	≥ 2	≥ 3
Boys	443	1.12	1.90	≥ 2	≥ 3
Functional Impairment Measures					
SSRS Social Skills (SSRS-SS-P)					
Girls	655	46.31	11.30	≤ 43	≤ 36
Boys	748	43.69	11.90	≤ 39	≤ 33
SSRS Social Skills (SSRS-SS-T)					
Girls	642	37.77	11.90	≤ 31	≤ 25
Boys	735	32.74	11.76	≤ 24	≤ 19
ESP Social Interaction (ESP-SIS-T)					
Girls	291	35.19	13.73	≤ 31	≤ 24
Boys	372	33.00	13.53	≤ 26	≤ 19

Note. For all scales, with the exception of Teacher SSRS Social Skills, data are available from the following sites: UNC, UNM, U of O, and Vanderbilt. For Teacher SSRS Social Skills, data are available from UNC, U of O, and Vanderbilt.

TABLE 3
Correlations Among Symptom and Functional Impairment Measures

	Symptom (Sx) Measures				*Functional Impairment (Fx) Measures*	
	SSRS-PB-P	***SSRS-PB-T***	***ESP-ABS-T***	***ESP-CEI-T***	***SSRS-SS-P***	***SSRS-SS-T***
Sx Measures						
SSRS-PB-T	0.24***					
ESP-ABS-T	0.20***	0.75***				
ESP-CEI-T	0.10*	0.32***	0.43***			
Fx Measures						
SSRS-SS-P	-0.20***	-0.13***	-0.20***	-0.13**		
SSRS-SS-T	-0.24***	-0.47***	-0.44***	-0.14**	0.18***	
ESP-SIS-T	-0.06	-0.14***	-0.10**	-0.28***	0.32***	0.44***

Note. Sx = Measure of symptoms; Fx = Measure of functional impairment; SSRS-PB-T = SSRS Problem Behavior scale; SSRS-SS-T = SSRS Social Skills scale; ESP-CEI-T = ESP Critical Events Index; ESP-SIS-T = ESP Social Interaction Scale; ESP-ABS-T = ESP Aggressive Behavior Scale. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

TABLE 4
Risk Status as Determined on Individual Measures and Impact on Gender, Race, and Ethnicity (1.0 SD Cut-Off)

Measure	*Total Ss With Data*	*Ss At Risk N (%)*	*Mean Age*	*Male %*	*Ethnicity %*				
					AA	*H*	*C*	*NA*	*A*
Symptom Measures (Sx)									
SSRS-PB-P	1,412	209 (14.8)	4.09	56.9	59.8	19.6	16.3	3.3	1.0
SSRS-PB-T	1,623	267 (16.5)	3.92***	48.3	67.0	9.0	18.7	4.9	0.4
ESP-ABS-T	679	266 (39.2)	4.02	68.8	39.8	22.9	25.9***	10.5	0.8
ESP-CEI-T	749	179 (23.9)	3.98	64.2	52.5	16.2	24.0	4.5	2.8
Functional Impairment Measures (Fx)									
SSRS-SS-P	1,403	544 (38.8)	3.87***	50.7	61.4***	17.5	17.1	2.9	1.1
SSRS-SS-T	1,377	388 (28.2)	3.78***	42.8***	73.5	4.9	14.9	6.2	0.5
EPS-SS-T	633	254 (38.3)	3.96***	53.9	35.0	36.2	15.4	13.0	0.4
SSRS Total Sample Demographics	1,745	—	3.98	53.5	63.4	18.6	12.7	4.6	0.7
ESP Total Sample Demographics	762	—	4.09	59.3	29.9	40.4	17.2	11.0	1.4

Sx = Measure of symptoms; Fx = Measure of functional impairment; SSRS-PB-T = SSRS Problem Behavior scale; SSRS-SS-T = SSRS Social Skills scale; ESP-CEI-T = ESP Critical Events Index; ESP-SIS-T = ESP Social Interaction Scale; ESP-ABS-T = ESP Aggressive Behavior Scale.
Race or Ethnicity: AA = African American; H = Hispanic; C = Caucasian; NA = Native American; A = Asian.
*** Adjusted model (controlling for site and sampling methodology) significant at $p < 0.001$.

the ESP social interaction scale all produced similarly high rates (39.2%, 38.8%, 38.3%, respectively). At the 1.5 cut-off, only the ESP aggressive behavior scale identified a much higher rate of at-risk children than the other measures (34.8%).

Analysis of the demographic characteristics of children identified as at risk by each instrument and cut-off revealed a number of potential differential effects. Note that in this and subsequent tables we have included demographic data on the sample at large, depicted separately for children with SSRS versus ESP data, in order to assist the reader in interpreting significant differences. In particular, a differential effect on age was tied to all three social skills scales at 1.0 and 1.5 standard deviations as well as the SSRS problem behavior

TABLE 5
Risk Status as Determined on Individual Measures and Impact on Gender and Ethnicity (1.5 SD Cut-Off)

Measure	Total Ss With Data	Ss At Risk N (%)	Mean Age	Male %	Ethnicity %				
					AA	H	C	NA	A
Symptom Measures (Sx)									
SSRS-PB-P	1,412	110 (7.8)	4.07	50.0	66.4	15.5	14.5	3.6	0.0
SSRS-PB-T	1,623	151 (9.3)	3.95	44.4	64.9	7.9	20.5	6.0	0.7
ESP-ABS-T	679	236 (34.8)	4.03	68.2	42.4	21.2	25.8	10.2	0.4
ESP-CEI-T	749	97 (13.0)	3.93	68.0	56.7	13.4	23.7	3.1	3.1
Functional Impairment Measures (Fx)									
SSRS-SS-P	1,403	266 (19.0)	3.80***	54.1	57.9***	15.4	22.6	3.0	1.1
SSRS-SS-T	1,377	176 (12.8)	3.74***	47.2	71.6	6.3	14.8	6.8	0.6
EPS-SS-T	633	154 (23.2)	3.90***	47.7	42.9	37.0	11.0	9.1	0.0
SSRS Total Sample Demographics	1,745	—	3.98	53.5	63.4	18.6	12.7	4.6	0.7
ESP Total Sample Demographics	762	—	4.09	59.3	29.9	40.4	17.2	11.0	1.4

Sx = Measure of symptoms; Fx = Measure of functional impairment; SSRS-PB-T = SSRS Problem Behavior scale; SSRS-SS-T = SSRS Social Skills scale; ESP-CEI-T = ESP Critical Events Index; ESP-SIS-T = ESP Social Interaction Scale; ESP-ABS-T = ESP Aggressive Behavior Scale.
Race or Ethnicity: AA = African American; H = Hispanic; C = Caucasian; NA = Native American; A = Asian.
*** Adjusted model (controlling for site and sampling methodology) significant at $p < 0.001$.

teacher scale at 1.0 standard deviation. All four measures identified a disproportionate amount of younger children relative to the sample at large. At the 1.0 cut-off, a differential effect on gender was evident on the SSRS teacher social skills rating with teachers identifying a disproportionate amount of girls. At the 1.5 cut-off, the gender effect was no longer evident. As shown in *Tables 4* and *5*, the SSRS parent social skills rating underidentified African American children relative to the sample at large at both cut-offs, and the teacher report on the ESP aggressive behavior scale overidentified Caucasian children (25.9%).

In *Tables 6* and *7*, we begin to address issues of "caseness" with respect to *teacher* identification of emotional or behavioral disorders. The three behavior scales are paired with the two social scales to produce six combinations of symptomatology and functional impairment. As expected, pairing a symptom measure with an impairment measure resulted in a dramatic reduction in at-risk identification rates. Application of the 1.0 standard deviation cut-off to the six combinations of teacher measures produced at-risk rates ranging from 5.0%, for the SSRS problem behavior and ESP social interaction combination, to 24.5%, for the ESP aggressive behavior and SSRS social interaction pairing. The remaining combinations ranged from 10.5% to 15.3%. As shown previously in *Tables 4* and *5*, the SSRS social skills teacher rating—an individual measure—produced lower rates (28.2% and 12.8%, respectively) than the ESP social interaction scale (38.3% and 23.2%). Here, however, the three combinations using the SSRS social skills scale as a measure of functional impairment produced generally higher rates (from 11.3% to 24.5%) than the three combinations using the ESP social interaction scale (from 5.0% to 12.8%).

Use of the 1.5 standard deviations cut-off resulted in further reductions in EBD prevalence rates. Although the two combinations that produced the highest and lowest rates at the 1.0 cut-off also produced the highest and lowest rates at the 1.5 cut-off, rates were reduced by one-half to two-thirds. At-risk rates ranged from 1.7%, for the SSRS-PB-T and ESP-SIS-T pairing to 13.0% for the ESP-ABS-T and SSRS-SS-T combination. The remaining combinations clustered closely together, ranging from 3.3% to 5.8%.

The combination of a teacher symptom measure with a teacher rating of functional

TABLE 6
Risk Status as Determined by Combinations of Different Teacher Measures of Symptoms (Sx) and Functional Impairment (Fx) (1.0 SD Cut-Off)

Diagnostic Measure		*Total Ss With Data*	*Ss At Risk N (%)*	*Mean Age*	*Male %*	*Ethnicity %*				
Sx	*Fx*					*AA*	*H*	*C*	*NA*	*A*
SSRS-PB-T	SSRS-SS-T	1,322	149 (11.3)	3.81***	40.9***	69.8	5.4	15.4	8.7	0.7
SSRS-PB-T	ESP-SIS-T	634	32 (5.0)	4.10	50.0	43.8	25.0	21.9	9.4	0.0
ESP-CEI-T	SSRS-SS-T	417	64 (15.3)	3.78***	51.6	64.1	7.8	21.9	4.7	1.6
ESP-CEI-T	ESP-SIS-T	658	84 (12.8)	3.89	57.1	59.5	20.2	14.3	4.8	1.2
ESP-ABS-T	SSRS-SS-T	347	85 (24.5)	3.88***	50.6	42.4	10.6	28.2	17.6	1.2
ESP-ABS-T	ESP-SIS-T	583	61 (10.5)	4.12	67.2	18.0	41.0	26.2	14.8	0.0
SSRS Total Sample Demographics		1,745	—	3.98	53.5	63.4	18.6	12.7	4.6	0.7
ESP Total Sample Demographics		762	—	4.09	59.3	29.9	40.4	17.2	11.0	1.4

Sx = Measure of symptoms; Fx = Measure of functional impairment; SSRS-PB-T = SSRS Problem Behavior scale; SSRS-SS-T = SSRS Social Skills scale; ESP-CEI-T = ESP Critical Events Index; ESP-SIS-T = ESP Social Interaction Scale; ESP-ABS-T = ESP Aggressive Behavior Scale.
Race or Ethnicity: AA = African American; H = Hispanic; C = Caucasian; NA = Native American; A = Asian.
*** Adjusted model (controlling for site and sampling methodology) significant at $p < 0.001$.

TABLE 7
Risk Status as Determined by Combinations of Different Teacher Measures of Symptoms (Sx) and Functional Impairment (Fx) (1.5 SD Cut-Off)

Diagnostic Measure		*Total Ss With Data*	*Ss At Risk N (%)*	*Mean Age*	*Male %*	*Ethnicity %*				
Sx	*Fx*					*AA*	*H*	*C*	*NA*	*A*
SSRS-PB-T	SSRS-SS-T	1,322	45 (3.3)	3.71***	39.5	62.8	9.3	14.0	11.6	2.3
SSRS-PB-T	ESP-SIS-T	634	11 (1.7)	4.14	27.3	45.5	36.4	18.2	0.0	0.0
ESP-CEI-T	SSRS-SS-T	417	20 (4.8)	3.59***	55.0	75.0	10.0	10.0	0.0	5.0
ESP-CEI-T	ESP-SIS-T	658	38 (5.8)	3.91	52.6	65.8	13.2	15.8	5.3	0.0
ESP-ABS-T	SSRS-SS-T	347	45 (13.0)	3.89	57.8	44.4	13.3	22.2	17.8	2.2
ESP-ABS-T	ESP-SIS-T	583	28 (4.8)	4.27	50.5	28.6	42.9	21.4	7.1	0.0
SSRS Total Sample Demographics		1,745	—	3.98	53.5	63.4	18.6	12.7	4.6	0.7
ESP Total Sample Demographics		762	—	4.09	59.3	29.9	40.4	17.2	11.0	1.4

Sx = Measure of symptoms; Fx = Measure of functional impairment; SSRS-PB-T = SSRS Problem Behavior scale; SSRS-SS-T = SSRS Social Skills scale; ESP-CEI-T = ESP Critical Events Index; ESP-SIS-T = ESP Social Interaction Scale; ESP-ABS-T = ESP Aggressive Behavior Scale.
Race or Ethnicity: AA = African American; H = Hispanic; C = Caucasian; NA = Native American; A = Asian.
*** Adjusted model (controlling for site and sampling methodology) significant at $p < 0.001$.

impairment also resulted in a notable reduction in significant differential effects on age, gender, and ethnicity relative to the sample at large. Only the combinations using the SSRS-SS-T as a functional impairment measure retained evidence of differential effects associated with age and gender, thus mirroring these same effects when this teacher social skills rating was used as a sole rating. At the 1.0 cut-off, all three SSRS-SS-T combinations overidentified younger children and, at the 1.5 cut-off, two of the three combinations (SSRS-PB-T and SSRS-SS-T, ESP-CEI-T and SSRS-SS-T) overidentified younger children relative to the sample at large. In addition, the SSRS-PB-T and SSRS-SS-T combination underidentified males with

respect to the sample at the 1.0 cut-off. At both 1.0 and 1.5 standard deviations, however, evidence of significant differential effects on ethnicity was no longer present when *combinations* of symptom and impairment measures were used.

Tables 8 and *9* further refine the assignment of caseness through the exploration of six combinations of parent and teacher measures. A parent symptom measure, SSRS-PB-P, was sequentially combined with one of the three teacher-reported symptom scales and one of the two teacher-reported functional impairment measures. Combinations of both teacher and parent symptom and impairment measures are presented first in this table, for comparison purposes. Again, a dramatic reduction in identification rates is evident. For the 1.0 standard deviation cut-off, rates ranged from 1.4% to 5.2%, and the 1.5 standard deviations cut-off produced at-risk rates ranging from 0.3% to 2.0%. It should be noted that statistically significant differential effects on age, gender, or ethnicity are no longer evident at either cut-off for the six parent and teacher combinations depicted in *Tables 8* and *9*.

Finally, in order to contrast the depiction of data in *Tables 4–9*, which include percentages of all children identified as EBD by racial or ethnic group, we have provided *Table 10* as one example of how data would appear if depicted as percentages of children from different racial and ethnic groups who are identified as EBD. Thus, *Table 10* presents exactly the same ethnicity data as contained in Table 8 but with the exception that these data are now depicted as *participants identified from racial and ethnic groups* rather than *identified groups composed of racial and ethnic participants*. Such a contrasting depiction may have somewhat different implications, as will be discussed below in reference to individual assessment decisions versus epidemiologic outcome.

Discussion

Possible limitations of this study include: (1) use of screening measures rather than extensive diagnostic or structured interviews, direct observation, and the like; (2) significant, cross-site differences in age and ethnicity; (3) lack of full data on significant numbers of children on some measures; (4) some irregularities in how subjects were selected among sites, such that an oversampling of children at potential risk for EBD may have occurred; and (5) the

TABLE 8
Risk Status as Determined by Combinations of Different Teacher and Parent Measures of Symptoms (Sx) and Functional Impairment (Fx) (1.0 SD Cut-Off)

Diagnostic Measure		*Total Ss With Data*	*Ss At Risk N (%)*	*Mean Age*	*Male %*	*Ethnicity %*				
Sx	*Fx*					*AA*	*H*	*C*	*NA*	*A*
SSRS-PB-P	SSRS-SS-P	1,402	101 (7.2)	3.99	53.5	50.5	18.8	25.7	4.0	1.0
SSRS-PB-T	SSRS-SS-T	1,322	149 (11.3)	3.81***	40.9***	69.8	5.4	15.4	8.7	0.7
SSRS-PB-P & SSRS-PB-T	SSRS-SS-T	1,015	27 (2.7)	4.01	37.0	74.1	3.7	14.8	3.7	3.7
SSRS-PB-P & ESP-CEI-T	SSRS-SS-T	301	12 (4.0)	4.17	58.3	58.3	8.3	16.7	8.3	8.3
SSRS-PB-P & ESP-ABS-T	SSRS-SS-T	251	13 (5.2)	4.28	53.8	61.5	7.7	15.4	7.7	7.7
SSRS-PB-P & SSRS-PB-T	ESP-SIS-T	515	7 (1.4)	4.39	42.9	42.9	42.9	14.3	0.0	0.0
SSRS-PB-P & ESP-CEI-T	ESP-SIS-T	527	11 (2.1)	4.20	72.7	63.6	9.1	27.3	0.0	0.0
SSRS-PB-P & ESP-ABS-T	ESP-SIS-T	473	11 (2.3)	4.42	63.6	27.3	36.4	27.3	9.1	0.0
SSRS Total Sample Demographics		1,745	—	3.98	53.5	63.4	18.6	12.7	4.6	0.7
ESP Total Sample Demographics		762	—	4.09	59.3	29.9	40.4	17.2	11.0	1.4

Sx = Measure of symptoms; Fx = Measure of functional impairment; SSRS-PB-P = SSRS Problem Behavior scale, Parent report; SSRS-PB-T = SSRS Problem Behavior scale, Teacher report; SSRS-SS-P = SSRS Social Skills scale, Parent report; SSRS-SS-T = SSRS Social Skills scale, Teacher report; ESP-CEI-T = ESP Critical Events Index; ESP-SIS-T = ESP Social Interaction Scale; ESP-ABS-T = ESP Aggressive Behavior Scale.
Race or Ethnicity: AA = African American; H = Hispanic; C = Caucasian; NA = Native American; A = Asian.
*** Adjusted model (controlling for site and sampling methodology) significant at $p < 0.001$.

TABLE 9
Risk Status as Determined by Combinations of Different Teacher and Parent Measures of Symptoms (Sx) and Functional Impairment (Fx) (1.5 SD Cut-Off)

Diagnostic Measure		*Total Ss With Data*	*Ss At Risk N (%)*	*Mean Age*	*Male %*	*Ethnicity %*				
Sx	*Fx*					*AA*	*H*	*C*	*NA*	*A*
SSRS-PB-P	SSRS-SS-P	1,402	33 (2.4)	4.00	57.6	48.5	12.1	33.3	6.1	0.0
SSRS-PB-T	SSRS-SS-T	1,322	43 (3.3)	3.71***	39.5	62.8	9.3	14.0	11.6	2.3
SSRS-PB-P & SSRS-PB-T	SSRS-SS-T	1,015	4 (0.4)	4.48	0.0	50.0	25.0	0.0	25.0	0.0
SSRS-PB-P & ESP-CEI-T	SSRS-SS-T	301	1 (0.3)	4.48	100.0	100.0	0.0	0.0	0.0	0.0
SSRS-PB-P & ESP-ABS-T	SSRS-SS-T	251	5 (2.0)	4.61	60.0	40.0	20.0	20.0	20.0	0.0
SSRS-PB-P & SSRS-PB-T	ESP-SIS-T	515	3 (0.6)	4.46	0.0	0.0	66.7	33.3	0.0	0.0
SSRS-PB-P & ESP-CEI-T	ESP-SIS-T	527	4 (0.8)	4.37	75.0	75.0	0.0	25.0	0.0	0.0
SSRS-PB-P & ESP-ABS-T	ESP-SIS-T	473	6 (1.3)	4.41	50.0	33.3	50.0	16.7	0.0	0.0
SSRS Total Sample Demographics		1,745	—	3.98	53.5	63.4	18.6	12.7	4.6	0.7
ESP Total Sample Demographics		762	—	4.09	59.3	29.9	40.4	17.2	11.0	1.4

Sx = Measure of symptoms; Fx = Measure of functional impairment; SSRS-PB-P = SSRS Problem Behavior scale, Parent report; SSRS-PB-T = SSRS Problem Behavior scale, Teacher report; SSRS-SS-P = SSRS Social Skills scale, Parent report; SSRS-SS-T = SSRS Social Skills scale, Teacher report; ESP-CEI-T = ESP Critical Events Index; ESP-SIS-T = ESP Social Interaction Scale; ESP-ABS-T = ESP Aggressive Behavior Scale.
Race or Ethnicity: AA = African American; H = Hispanic; C = Caucasian; NA = Native American; A = Asian.
*** Adjusted model (controlling for site and sampling methodology) significant at $p < 0.001$.

TABLE 10
Replication of Data contained in Table 8 With Ethnicity Percentages Depicted as Participants Identified by Ethnic Group

Diagnostic Measure		*Total Ss With Data*	*Ethnicity %*				
Sx	*Fx*		*AA*	*H*	*C*	*NA*	*A*
SSRS-PB-P	SSRS-SS-P	1,402	5.8	7.0	14.3	7.4	10.0
SSRS-PB-T	SSRS-SS-T	1,322	9.4	2.3	10.3	15.3	7.7
SSRS-PB-P & SSRS-PB-T	SSRS-SS-T	1,015	2.6	2.4	2.6	2.6	12.5
SSRS-PB-P & ESP-CEI-T	SSRS-SS-T	301	4.8	3.4	2.5	2.6	14.3
SSRS-PB-P & ESP-ABS-T	SSRS-SS-T	251	7.8	3.4	2.7	2.6	16.7
SSRS-PB-P & SSRS-PB-T	ESP-SIS-T	515	2.6	1.2	1.1	0.0	0.0
SSRS-PB-P & ESP-CEI-T	ESP-SIS-T	527	6.2	3.1	0.4	0.0	0.0
SSRS-PB-P & ESP-ABS-T	ESP-SIS-T	473	4.4	1.6	3.3	1.8	0.0

Sx = Measure of symptoms; Fx = Measure of functional impairment; SSRS-PB-P = SSRS Problem Behavior scale, Parent report; SSRS-PB-T = SSRS Problem Behavior scale, Teacher report; SSRS-SS-P = SSRS Social Skills scale, Parent report; SSRS-SS-T = SSRS Social Skills scale, Teacher report; ESP-CEI-T = ESP Critical Events Index; ESP-SIS-T = ESP Social Interaction Scale; ESP-ABS-T = ESP Aggressive Behavior Scale.
Race or Ethnicity: AA = African American; H = Hispanic; C = Caucasian; NA = Native American; A = Asian.
*** Adjusted model (controlling for site and sampling methodology) significant at $p < 0.001$.

fact that most symptom and impairment measures in this sample correlate highly with one another, thus limiting the impact of combinations of measures. In regard to the first and last limitations, it should be pointed out that both the SSRS and the ESP are screening systems that are highly reliable and valid with relatively diverse standardization samples. Unlike many diagnostic instruments in current use, they contain scales specifically designed for preschoolers. Such systems have, however, been subject to some criticism in reference to

their use with low-income preschool children. Fantuzzo, Manz, & McDermott (1998), for example, point out that the problem behavior scale of the SSRS may be inversely related to its social skills scale, which we also found in our correlations, particularly for the teacher version. Such systems are nonetheless arguably among the most widely used currently for screening, detection, and identification of children at risk for a range of emotional and behavioral disorders in preschool settings (Walker et al., 2004). As noted at the beginning of this paper, these measures in effect may also serve as analogs to other similar measures used for preschoolers. In regard to lack of participants on ESP measures in particular, it should be noted that the number of subjects, even in these reduced subsamples, nonetheless places their total among the top 20% among 40 recently available epidemiologic studies in this area, as noted above (Qi & Kaiser, 2003; Roberts et al., 1998).

Readers should be cautioned, however, that our sample demographics were indeed more heavily weighted toward certain ethnic or racial backgrounds, along with a slightly higher possibility of selection of subjects at risk for EBD, in at least one or two of the samples. Although we have attempted to control statistically for such differences, such limitations may not allow generalization to preschool populations in general. There may also remain some unexplained variance in our findings due to cross-site differences. Both the Vanderbilt and UNC subsamples were predominantly African American; the New Mexico site predominantly Hispanic; and the Oregon site contained a significant number of children from Native American backgrounds. In this respect note that, in most instances, teachers tended to generally reflect at least some of the similar racial or ethnic backgrounds as children at their respective sites. This study nonetheless represents a sample of considerable opportunity to analyze at least relative variations in prevalence, potential instrument bias, and related issues in emotional or behavioral disorders that have not been previously addressed in most preschool samples. The reader is further cautioned, however, to interpret data in the various clinical cut-off tables in reference to data that we have provided at the bottom of each table for the sample at large.

Our primary findings include striking differences in prevalence dependent on choice of measures and combinations of measures; some consistencies in significant age or ethnic bias, even when different clinical cut-off points were used; relatively little significant gender bias; some diminishment of bias when combinations of symptom and impairment measures were used; and a possibility that best practice in identification (that is, using a variety of measures and informants) may lead to an unacceptable level of false negatives. Further discussion of these findings follows.

That this sample of low-income children is at higher risk for emotional or behavioral disorders is evident, even in mean scores for certain measures depicted in *Table 2*. Note that the *means* for SSRS social skills for both boys and girls are within 3 or 4 points of reaching clinical cut-off points at 1.0 standard deviation. A similar phenomenon is evident for the mean ESP aggression scores in that the average score for boys is actually within the clinical range. A number of other mean scores in this table are also elevated. As noted in *Table 3*, the most significant correlations are among measures of symptoms and among most measures of impairment, along with inverse correlations between some symptom and impairment measures.

As noted in *Table 4*, prevalence as determined on individual measures at the 1.0 standard deviation level of risk are, for the most part, relatively close to the overall level of risk found in the Qi and Kaiser (2003) review. As a matter of fact, the overall mean for all seven measures in this table is approximately 28%, as compared to the 30% found by Qi and Kaiser. This same mean for *Table 5*, the 1.5 standard deviation cut-off, is approximately 17%, almost half of the previous mean and thus somewhat closer to the mean preschool epidemiologic finding of 10.2% reported by Roberts et al. (1998).

It is interesting that the pattern of significant age differences at both levels of clinical cut-off are quite similar across *Tables 4* and *5*. All three social skill or interaction measures from the SSRS and ESP, including both parent and teacher informants, appear to somewhat overidentify younger children relative to the sample at large, suggesting that such measures may tend to mistakenly target nondeviant social behavior as problematic in younger children. The same appears true for teacher-reported symptoms of problem behavior but only at the 1.0 clinical cut-off. This bias may prove to be particularly significant because more and more

children are being recruited into Head Start programs at age 3 or younger (Boyce et al., 2000; Peterson et al., 2004). Campbell (2002) has cautioned that behaviors of preschool children are often characterized as abnormal due to a lack of caregiver knowledge of normal developmental trajectories.

Teacher ratings of Caucasian children appear significantly more likely to overidentify aggressive symptoms in children relative to the sample at large, at least at the 1.0 clinical cut-off, and parent ratings of social skills tend to significantly underidentify African American children at the 1.0 clinical cut-off. The latter finding also held true at the 1.5 clinical cut-off. Whether this reflects cultural-familial differences or measurement bias is not clear. These findings nonetheless contrast with previous findings of presumably elevated psychopathology in African American children, especially boys, and relative underidentification of Hispanics and Caucasians in large-scale studies of school-age children (Coutinho, Oswald, Best, & Forness, 2002; Garland et al., 2001; Kataoka, Zhang, & Wells, 2002; Yeh, Hough, McCabe, Lau, & Garland, 2004). These findings also contrast with findings by Yeh, Forness, Ho, McCabe, and Hough (in press) in which overidentification of African American children, relative to children of other ethnic or racial backgrounds, ceased being statistically significant when one controlled for *severity* of disorder. Aggression and related disruptive behaviors in preschoolers are somewhat difficult to characterize because they are characteristically comorbid with a variety of other disorders (Bennett & Offord, 2001; Gadow, Sprafkin, & Nolan, 2001; Konold, Hamre, & Pianta, 2003; Thomas & Guskin, 2001). It is unfortunate that no measures in the present study isolated internalizing disorders *per se*, since studies have recently begun to focus more directly on phenomenology of these disorders in preschoolers (Luby et al., 2002, 2003).

It is somewhat surprising that the only evidence of gender bias tended to be a significant *underidentification* of males by teachers in social skills at the 1.0 clinical cut-off, which contrasts with normative school-age findings of male overidentification for EBD (Bussing et al., in press; Coutinho et al., 2002). Whether such male vulnerability has not yet manifested itself in preschool years is open to question, but receives some support in this instance. This underidentification even appears to carry over into subsequent tables when teacher-rated social skills are paired with teacher-rated problem behavior symptoms.

In *Tables 6* and *7*, teacher measures are combined to reflect best practice of using both a measure of EBD symptoms *and* a measure of impairment presumably resulting from such symptoms. Note that, with one or two exceptions, prevalence is relatively close to that of mental health epidemiologic data reviewed by Roberts and his colleagues (1998). The mean prevalence across all six combinations is approximately 13% for 1.0 standard deviation cut-offs and almost 6% for 1.5 standard deviations, as compared to the median of 8.3% and the mean of 10.2% found in the Roberts' study (Roberts et al., 1998). As noted in both tables, there is very little evidence of bias, except for some age and gender bias that might possibly be a residual of the original bias for teacher-rated social skills depicted in *Tables 4* and *5*. The social skills measure on the SSRS tends to reflect *skill* development, whereas the social measure on the ESP tends to tap actual *interaction* and may thus, in fact, be a more appropriate consideration for developing caseness in preschool psychopathology.

Combinations in *Tables 8* and *9* could be considered to reflect the ultimate in best practice for school identification in that symptoms are present across both home and school settings, yet impairment is particular to the school setting where eligibility for school services must be established. Parent combinations of symptoms and impairment and teacher combinations of these same two measures, which were also reported in *Tables 6* and *7*, are provided at the top of these tables for comparison purposes. Note that these two combinations of symptoms and impairment measures produce different prevalence rates and differing patterns of bias in age and ethnicity, as in previous studies on informant bias in school-age children noted at the beginning of this paper. Cai, Kaiser, & Lipsey (2004) produced a meta-analysis of 35 studies on cross-informant reports on child behavior problems from parents and teachers, but only five studies involved preschool children and only one of those involved a sample of 500 or more preschoolers. For those five studies, correlations among parent and teacher measures were relatively low, as we found in this study—at least compared to correlations among teacher measures. In more recent studies done by Cai and her colleagues on low-income African American

children using the Child Behavior Checklist for preschoolers (Achenbach & Rescorla, 2000), low correlations were also found (Cai, Kaiser, & Hancock, 2004; Cai, Kaiser, Hancock, & Lipsey 2004). They likewise found that parents generally rated more children as having emotional or behavior problems than did teachers, somewhat the opposite of findings in *Tables 8* and *9* in our study, suggesting possible different environmental demands and caregiver expectations.

In regard to the other six combinations in *Tables 8* and *9*, they produce prevalence rates that seem, with one or two exceptions, *unacceptably low* particularly for low-income children who would be expected to be at higher risk for EBD. Children identified at the 1.5 clinical cut-off level are, in fact, even well below restrictive levels of identification typical of that for school-age children identified in the special education category of emotional disturbance under the Individuals with Disabilities Education Act (U.S. Department of Education, 2003). The almost complete underidentification, on some of these measures, of children from Asian and Native American backgrounds is further troubling.

Thus, the best that can be said for exemplary practice, given the data in *Tables 8* and *9*, is that it would appear to produce virtually no false positives. These data indeed represent a controversy in psychiatric diagnosis on the "or rule" versus the "and rule" (Comer & Kendall, 2004). Some clinicians use the more conservative "and rule" in which a disorder is considered present only if the reports of all informants meet criteria for that disorder. Others consider a disorder as present if only one of the informants report clinical cut-off criteria for that disorder. It would thus be interesting to follow children identified by various measures and combinations longitudinally to see if they indeed retain their diagnostic levels of classification over time, because follow-up studies of children first identified in preschool have produced somewhat mixed results in this regard (Bennett & Offord, 2001; Kroes et al., 2002; Lavigne et al., 2001).

Table 10, a different depiction of ethnicity data originally presented in *Table 8*, was included as a reminder to readers that overrepresentation of children from different ethnic or racial groups in disability categories is a complex and highly charged issue. Since the emphasis in this study is primarily on prevalence, we have chosen to display all of our data in *Tables 4–9* as a percentage of all children identified as EBD by ethnic or racial groups. MacMillan and Reschly (1998) have pointed out, however, that this approach may tend to somewhat overemphasize possible ethnic or racial bias compared to the approach in *Table 10*, which depicts the percentage of ethnic or racial groups identified as EBD. Individual clinicians may be more interested in the latter approach because it focuses on relative odds of an individual child being diagnosed from among his or her peers of the same ethnic or racial background rather than the composition of the category ultimately identified, which is a primary purpose of most prevalence studies.

The question ultimately arises as to which measure or which combination of measures appears to produce a reasonable prevalence estimate, without significant bias, based on findings in this dataset. The only single measure that appears to do so is the parent SSRS problem behavior subscale. This measure factors into both internalizing and externalizing symptoms (Gresham & Elliot, 1990); and, in both *Tables 4* and *5*, appears to have no significant differences from the sample adjusted-model means across age, gender, and all five ethnic or racial groups. It also produced prevalence figures of approximately 15% at the 1.0 standard deviation cut-off and approximately 8% at the 1.5 standard deviation level. Both match reasonably well with the relatively conservative mean prevalence level found by Roberts et al. (1998). As to *combinations* of measures, the ESP critical events index and the ESP social interaction scale might be among the best of choices. This combination probably comes closest to a reasonable prevalence estimate, at least at the 1.0 standard deviation cut-off, and is almost exactly at the average for all six combinations in prevalence percentage, both at the 1.0 and 1.5 standard deviation levels; not only is there no significant bias across both *Tables 6* and *7*, but, on both tables, it probably comes about as close as any combination overall to the sample means for age, gender, and the five racial or ethnic categories. This combination also has a certain logical fit in that the critical events index screens for serious symptoms across a wide range of psychiatric disorders, and the social interaction scale measures a range of interpersonal behaviors.

It is, of course, hazardous to recommend, or to dismiss as biased or invalid, *any* of these

measures, alone or in combination, because each measure is part of its own integral *system* of screening or identification, and the results are strictly limited to the dataset used in this study. To the extent, however, that such measures represent analogs of other diagnostic instruments used with preschool children, such an exercise may nonetheless be instructive and may possibly serve as a model for examining other widely-used instruments used to determine psychopathology in preschoolers.

REFERENCES

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavior and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213–232.

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington: University of Vermont, Department of Psychiatry.

Apter, A., Fallon, T. J., King, R. A., Ratzoni, G., Zohar, A. H., Binder, M., et al. (1996). Obsessive-compulsive characteristics: From symptoms to syndrome. *Journal of the American Academy of Child and Adolescent Psychiatry, 35*, 907–912.

Bennett, K. J., & Offord, D. R. (2001). Screening for conduct problems: Does the predictive accuracy of conduct disorder symptoms improve with age? *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 1418–1425.

Biederman, J., Mick, E., Faraone, S. V., & Burbank, M. (2001). Patterns of remission and symptom decline in conduct disorder: A four-year prospective study of an ADHD Sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 290–298.

Bird, H. R., Gould, M. S., & Staghezza, B. (1992). Aggregating data from multiple informants in child psychiatry epidemiological research. *Journal of the American Academy of Child and Adolescent Psychiatry, 31*, 78–85.

Bird, H. R., Yager, T. J., Staghezza, B., Gould, M. S., Canino, G., & Rubio-Stipec, M. (1990). Impairment in the epidemiological measurement of childhood psycho-pathology in the community. *Journal of the American Academy of Child and Adolescent Psychiatry, 29*, 796–803.

Boyce, C. A., Hoagwood, K., Lopez, M. L., & Tarullo, L. B. (2000). The Head Start mental health research consortium: New directions for research partnerships. *Behavioral Disorders, 26*, 7–12.

Boyle, M.H., Offord, D.R., Racine, Y., Szatmari, P., Fleming, J.E., & Sanford, M. (1996). Identifying thresholds for classifying childhood psychiatric disorders: Issues and Prospects. *Journal of the American Academy of Child and Adolescent Psychiatry, 35*, 1440–1448.

Bussing, R., Zima, B. T., Mason, D., Hou, W., Garvan, C. W., & Forness, S. R. (in press). Use and persistence of pharmacotherapy for elementary school students with attention deficit hyperactivity disorder. *Journal of Child and Adolescent Psychopharmacology*.

Cai, X., Kaiser, A. P., & Hancock, T. B. (2004). Parent and teacher agreement on Child Behavior Checklist items in a sample of preschoolers from low-income and predominantly African American families. *Journal of Clinical Child and Adolescent Psychology, 33*, 303–312.

Cai, X., Kaiser, A. P., Hancock, T. B., & Lipsey, M. W. (2004). *Parent- and teacher-classified problem behaviors in preschool children from low-income and predominantly African American families: Cross-informant agreement and disagreement.* Unpublished manuscript, Department of Special Education, Peabody College of Vanderbilt University, TN, 37203.

Cai, X., Kaiser, A. P., & Lipsey, M. W. (2004). *Parent and teacher reports of problem behaviors in children: Cross-informant agreement, correlates, and information integration.* Unpublished manuscript, Department of Special Education, Peabody College of Vanderbilt University, TN, 37203.

Campbell, S. B. (2002). *Behavior problems in preschool children: Clinical and developmental issues* (2nd ed.). Pittsburgh, PA: University of Pittsburgh.

Cluett, S. E., Forness, S. R., Ramey, S. L., Ramey, C. E., Hsu, C., Kavale, K. A., et al. (1998). Consequence of differential diagnostic criteria on identification rites of children with emotional and behavioral disorders. *Journal of Emotional and Behavioral Disorders, 6*, 130–140.

Comer, J. S., & Kendall, P. C. (2004). A symptom-level examination of parent-child agreement in the diagnosis of anxious youths. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*, 878–886.

Coutinho, M. V., Oswald, D. P., Best, A. M., & Forness, S. R. (2002). Gender and sociodemographic factors in the disproportionate identification of minority students as emotionally disturbed. *Behavioral Disorders, 27*, 100–112.

Culp, R. E., Howell, C. S., Culp, A. M., & Blankemeyer, M. (2001). Maltreated children's emotional and behavioral problems: Do teachers and parents see the same thing? *Journal of Child and Family Studies, 10*, 39–50.

Fantuzzo, J., Manz, P.H., & McDermott, P. (1998). Preschool version of the Social Skills Rating System: An empirical analysis of its use with low-income children. *Journal of School Psychology, 36*, 199–214.

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

Fergusson, D. M., & Horwood, L. J. (1995). Predictive validity of categorically and dimensionally scored measures of disruptive childhood behaviors. *Journal of the American Academy of Child and Adolescent Psychiatry, 34*, 477–485.

Forness, S. R., & Kavale, L. A. (1997). Defining emotional and behavioral disorders in school and related services. In J.W. Lloyd, E.J. Kameenui, & D. Chard (Eds.), *Issues in educating students with disabilities* (pp. 45–61). Hillsdale, NJ: Erlbaum.

Forness, S. R., & Kavale, L. A. (2000). Emotional or behavioral disorders: Background and current status of the EBD terminology and definition. *Behavioral Disorders, 25*, 264–269.

Forness, S. R., Ramey, S. L., Ramey, C. T., Hsu, C., Brezausek, C. M., MacMillan, D. L., et al. (1998). Head Start children finishing first grade: Preliminary data on school identification of children at risk for special education. *Behavioral Disorders, 23*, 111–124.

Gadow, K. D., Sprafkin, J., & Nolan, E. E. (2001). DSM-IV symptoms in community and clinic preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 1383–1392.

Garland, F., Hough, R. L., McCabe, K. M., Yey, M., Wood, P. A., & Arrons, G. A. (2001). Prevalence of psychiatric disorders in youths across five sectors of care. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 409–418.

Gresham, F. M., & Elliott, S. N. (1990). *Social skills rating system*. Circle Pines. MN: American Guidance Service.

Kataoka, S. H., Zhang, L., & Wells, K. B. (2002). Unmet need for mental health care among U.S. children: Variation by ethnicity and insurance status. *American Journal of Psychiatry, 159*, 1548–1555.

Kauffman, J. M. (2005). *Characteristics of emotional and behavioral disorders of children and youth*. Upper Saddle River, NJ: Merrill Prentice Hall.

Konold, T. R., Hamre, B. K., & Pianta, R. C. (2003). Measuring problem behaviors in young children. *Behavioral Disorders, 28*, 111–123.

Kramer, T., Phillips, S., Hargis, M., & Miller, T. (2004). Disagreement between parent and adolescent reports of functional impairment. *Journal of Child Psychology and Psychiatry, 45*, 248–259.

Kroes, M., Kalff, A. C., Steyaert, J., Kessels, A. G. H., Feron, F. J. M., Hendriksen, J. G. M., et al. (2002). A longitudinal community study: Do psychosocial risk factors and child behavior checklist scores at 5 years of age predict psychiatric diagnoses at a later age? *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 955–963.

Lavigne, J. V., Arend, R., Rosenbaum, O., Binns, H. J., Christoffel, K. K., & Gibbons, R. D. (1998a). Psychiatric disorders with onset in the preschool years: I. Stability of diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry, 37*, 1246–1254.

Lavigne, J. V., Arend, R., Rosenbaum, R., Binns, H. J., Christoffel, K. K., & Gibbons, R. D. (1998b). Psychiatric disorders with onset in the preschool years: II. Correlates and predictors of stable case status. *Journal of the American Academy of Child and Adolescent Psychiatry, 37*, 1255–1261.

Lavigne, J. V., Cicchetti, C., Gibbons, R. D., Binns, H. J., Larsen, L. & DeVito, C. (2001). Oppositional defiant disorder with onset in preschool years: Longitudinal stability and pathways to other disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 1393–1400.

Lopez, M. L., Tarullo, L. B., Forness, S. R., & Boyce, C. (2000). Early identification and intervention: Head Start's response to mental health challenges. *Early Education and Development, 11*, 265–282.

Luby, J. L., Heffelfinger, A. K., Mrakotsky C., Brown, K. M., Hessler, M. J., Wallis, J. M., et al., (2003). The clinical picture of depression in preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry, 42*, 340–348.

Luby, J. L., Heffelfinger, A. K., Mrakotsky C., Hessler, M. J., Brown, K. M., & Hildebrand, T. (2002). Preschool major depressive disorder: Preliminary validation for developmentally modified DSM-IV criteria. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 928–937.

MacMillan, D. L., & Reschly, D. J. (1998). Overrepresentation of minority students: The case for gender specificity or reconsideration of the variables examined. *Journal of Special Education, 32*, 15–24.

Mesman, J., & Koot, W. M. (2000). Child-reported depression and anxiety in preadolescence: I. Associations with parent- and teacher-reported problems. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 1371–1378.

Mitsis, E. M., McKay, K. E., Schulz, K. P., Newcorn, J. H., & Halperin, J. M. (2000). Parent-teacher concordance for DSM-IV attention-deficit/hyperactivity disorder in a clinic-referred sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 308–313.

Offord, D. R., Boyle, M. H., Racine, Y., Szatmari, P., Fleming, J. E., Sanford, M., & Lipman, E. L. (1996). Integrating assessment data from multiple informants. *Journal of the American Academy of Child and Adolescent Psychiatry, 35*, 1078–1085.

Pennington, B. F. (2002). *The development of psychopathology: Nature and nurture*. New York: Guilford.

Peterson, C. A., Wall, S., Raikes, H. A., Kisker, E. E., Swanson, M. E., Jerald, J., et al. (2004). Early Head Start: Identifying and serving children with

disabilities. *Topics in Early Childhood Education, 24*, 76–88.

Piacentini, J. D., Cohen, P., & Cohen, M. (1992). Combining discrepant diagnosis information from multiple sources: Are complex algorithms better than simple ones? *Journal of Abnormal Child Psychology, 20*, 51–63.

Qi, C. H., & Kaiser, A. P. (2003). Behavior problems of preschool children from low-income families: Review of the literature. *Teaching Early Childhood Special Education, 23*, 188–216.

Razzino, B. E., New, M., Lewin, A., & Joseph, J. (2004). Need for and use of mental health services among parents of children in the Head Start Program. *Psychiatric Services, 55*, 583–586.

Roberts, R. E., Attkisson, C. C., & Rosenblatt, A. (1998). Prevalence of psychopathology among children and adolescents. *American Journal of Psychiatry, 155*, 715–725.

Serna, L., Nielsen, E., Mattern, N., Schau, C., & Forness, S. R. (2002). Use of different measures to identify preschoolers at risk for emotional or behavioral disorders: Impact on gender and ethnicity. *Education and Treatment of Children, 25*, 415–437.

Shaffer, D., Fisher, P., Dulcan, M. K., Davies, M., Piacentini, J., Schwab-Stone, M. E., et al. (1996). The NIMH diagnostic interview schedule for children version 2.3 (DISC-2.3): Description, acceptability, prevalence rates, and performance in the MECA study. *Journal of the American Academy of Child and Adolescent Psychiatry, 35*, 865–877.

Stormont, M. (2000). Early child risk factors for externalizing and internalizing behaviors: A 5-year follow-forward assessment. *Journal of Early Intervention, 23*, 180–190.

Thomas, J.M., & Guskin, K.A. (2001). Disruptive behavior in young children: What does it mean? *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 44–51.

U.S. Department of Education (2003). *Twenty-fifth annual report to Congress on the implementation of Individuals with Disabilities Education Act.* Washington, DC: Author.

Walker, H. M., Ramsey, E., & Gresham, F. M. (2004). *Antisocial behavior in school: Evidence-based practices* (2nd edition). Belmont, CA: Wadsworth/Thomson.

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

Witmer, D., Doll, B., & Strain, P. (1996). Social and emotional development in early childhood: The identification of competence and disabilities. *Journal of Early Intervention, 20*, 299–318.

Yeh, M., Forness, S. R., Ho, J., McCabe, K., & Hough, R. L. (in press). Parental etiological explanations and disproportionate racial/ethnic representation in special education services for youth with emotional disturbance. *Behavioral Disorders*.

Yeh, M., Hough, R. L., McCabe, K., Lau, A., & Garland, A. (2004). Parental beliefs about the causes of child problems: Exploring racial/ethnic patterns. *Journal of the American Academy of Child and Adolescent Psychiatry, 43*, 605–612.

AUTHORS' NOTES

Please address all correspondence and reprint requests to Edward Feil, Ph.D., Oregon Research Institute, 1715 Franklin Blvd., Eugene, OR 97403-1983. Phone: (541) 484-2123; Fax: (541) 484-1108; E-mail: edf@ori.org.

The views expressed are solely those of the authors and do not necessarily represent the views or policy of the U.S. Department of Health and Human Services.