American Journal of Orthopsychiatry
2007, Vol. 77, No. 3, 402–418

0002-9432/07/$12.00 DOI: 10.1037/0002-9432.77.3.402

Mental Health Screening of Preschool Children: Validity and Reliability of ABLE

Oscar A. Barbarin, PhD
University of North Carolina, Chapel Hill

Children with behavioral, emotional or language problems struggle to do well at school often with limited success. ABLE (Attention, Behavior, Language, and Emotions) a new screening tool was used to estimate the prevalence and the severity of concerns parents and teachers have about children's school adjustment and evaluate their need for services. Data obtained from the parents and teachers of children randomly selected from public Pre-K classrooms in 6 states ($N = 415$) and from a mental health screening of rural and urban children ($N = 5,577$) support the validity and reliablity of ABLE. Parents identified severe problems in 18.4% of children and Pre-K teachers identified 10.5%. By kindergarten, the proportion of children identified by their teachers with serious problems more than doubled to 23%. Inattention/overactivity and behavior problems were identified most often. These children were 3.4 times more likely to be certified later for special education services by kindergarten than children not identified with problems by ABLE. However, fewer than 14% of children in public Pre-K identified with serious problems in Pre-K had received mental health services by the end of Kindergarten.

Keywords: pre-school children, mental health, assessment, screening, early childhood

Many states have responded to concerns about the school readiness of young children by funding early childhood programs for 4 year olds. These programs, referred to as state sponsored prekindergarten (Pre-K), have targeted children who are at risk of early school difficulty because of household poverty. Typically, state-sponsored Pre-K programs have been operated in public school buildings or under the auspices of state and local education agencies for 2 to 6 hours per day and 3 to 5 days per week. Many programs have adopted structured curricula that emphasize school readiness, especially the acquisition of language and early literacy skills.

Many children arrive at school with socioemotional difficulties serious enough to compromise academic success (U.S. Department of Health & Human Services, 1999, 2001). For example, 46% of a nationally representative sample of kindergarten teachers reported that most of the children in their class lacked the kinds of self-regulatory skills that would enable them to function productively in the kindergarten classroom (Rimm-Kaufman, Pianta, & Cox, 2000). Because children with behavioral and emotional problems struggle to succeed in school, kindergarten teachers expressed as much concern for children's socioemotional readiness for school as for their academic skills per se (NCES, 1993).

Along with the pressures on Pre-K programs to enhance school readiness has come a recognition of the role socioemotional functioning plays in the development of academic competence (Raver, 2002). For this reason, identification and attenuation of socioemotional difficulties have been viewed as integral to the success of publicly supported early childhood programs. For example, Head Start has adopted standards mandating mental health assessment and intervention for socioemotional problems of children enrolled in its programs (Head Start Quality Research Consortium, 2003). Full implementation of these standards has been constrained by a dearth of valid screening tools whose content is relevant to the Pre-K program context and which capture the most common problems in that setting. A plethora of measures have been developed and used for this purpose (e.g., Achenback & Edelbrock, 1991; Behar, 1974; Berls & McEwen, 1999; Blankenmeyer, Culp, Hubbs-Tait, & Culp, 2002; Bricker, Squires, & Mounts, 1995; Brigance, 1991; Canivez & Rains, 2002; Connors, 1990; Gresham & Elliott, 1990; Harrison, 1990; High/Scope Educational Research Foundation, 1992; Kohn, 1988; Naglieri, LeBuffe, & Pfeiffer, 1994). However, agreement on a single measure for socioemotional screening has been elusive because measures lacked psychometric rigor or a link to intervention, produced results that could not be easily interpreted and used in settings that lacked access to diagnostically trained mental health professionals, or because the time required to complete the measure was a disincentive for staff and parents already heavily burdened by extensive reporting requirements (Sinclair & Gonzalez, 1993). Development of ABLE (Attention, Behavior, Language, and Emotions) has been undertaken with the goal of addressing these concerns by creating

Data reported here were collected partly by the National Center for Early Development and Learning (NCEDL). The NCEDL Multi-State Study of Pre-Kindergarten was conducted by a team of researchers including Oscar Barbarin, Donna Bryant, Margaret Burchinal, Richard Clifford, and Diane Early at the University of North Carolina at Chapel Hill, Carollee Howes at UCLA, and Robert Pianta at the University of Virginia.

This research was supported by a Foundation for Child Development grant to the author and by the Educational Research and Development Center Program, PR/Award Number R307A60004, administered by the Institute of Education Sciences, U.S. Department of Education. However, the contents do not necessarily represent the positions or policies of the Foundation or of U.S. Department of Education, and endorsement by the Federal government should not be assumed.

For reprints and correspondence: Oscar Barbarin, PhD, Frank Porter Graham Child Development Institute, 517 S. Greensboro St., CB # 8040, Carrboro, NC 27510-8040. E-mail: barbarin@unc.edu

a simple screening tool that identifies children at risk of early onset socioemotional difficulties (Barbarin, 2006). This report describes the development of ABLE and assesses its effectiveness in identifying children with problems and in predicting which children would be certified through an Individual Education Plan (IEP) as eligible for special education services by the end of Pre-K or the end of kindergarten.

Prevalence of Socioemotional Difficulties in Young Children

Young children are susceptible to a range of behavioral and emotional difficulties. Epidemiological studies have reported adjustment difficulties among preschool children that range from delays in development of receptive and expressive language to oppositional behavior and serious emotional disturbances (Campbell, 1995; Keenan & Wakschlag, 2000). Conduct problems, hyperactivity, and phobias have been the most commonly reported problems in large nationally representative samples (Barbarin & Soler, 1993) and in select community samples of young children (Kellam, Branch, Agrawal, & Ensminger, 1975). Some studies of mental health prevalence use formal diagnostic categories such as the *DSM–IV* (American Psychiatric Association, 1994) or the DC: 0–3 as a framework (Costello, Angold, Burns, Erkanli, Stangl, & Tweed, 1996). When this categorical approach was used, the overall prevalence of behavioral and emotional difficulties among preschool children was similar to the rates observed for older children, that is, about 8% to 10% for conditions that meet the *DSM–IV* criteria for diagnosable mental disorders and in the range of 14% to 22% for disorders that were impairing but did not meet all the conditions of a formal diagnosis (Mash, 2003). In a nonclinic community sample in Ontario Canada, 9.9% of school-age children were diagnosed with Overanxious Disorder, 6.6% with Attention Deficit Hyperactivity Disorder (ADHD), and 5.5% with Conduct Problems (Offord, Boyle & Racine, 1989). Other forms of behavior problems ranged from 5.7% to 9.9% and different forms of mood disturbances ranged between 1.8% and 5.9%. Although rates of serious diagnosable disorders in children fall in a range from 10% to 22%, as many as 33% of children in nonselective community samples evidence subclinical psychological symptoms, which have a detrimental impact on quality of life and compromise functioning in domains such as school, family life or social relations outside the home (McDermott & Weiss, 1995).

Although categorical diagnostic frameworks such as the DSM and DC: 0–3 are valuable tools in the delivery of mental health services, their value for socioemotional screenings in early childhood is limited. The goal of a screening tool such as ABLE is not diagnosis of DSM mental disorders but identification, in a nonclinical population of children, of those who might benefit from intervention for early onset symptoms that disrupt adjustment to the home or the early childhood setting. Screening tools should focus on problems that are precursors to serious and disruptive psychological disorders. A dimensional approach is better suited for that purpose and has been adopted in numerous efforts to assess children's mental health (e.g., Child Behavior Checklist [CBCL]; Achenbach & Edelbrock, 1991). The dimensional approach uses empirically derived domains arrayed in continuous dimensions of attentional, behavioral, and emotional problems that range from high to low rather than yes/no or case/noncase as with the categorical approach.

Dimensional measures can be transformed into categorical indicators with the use of empirically derived cut-off points. Using this approach Fantuzzo, Bulotsky, McDermott, Mosca, and Lutz (2003) found that among urban Head Start programs, 13% of children ($n \sim 900$) had a severe difficulty, of whom about 5.9% had difficulties classified as Oppositional, 5.6% as Aggressive; 5.3% as Withdrawn, 4.8% as Socially Reticent; and 4.5% as Inattentive/Hyperactive. Although no claims have been made that their cut-off score identified children with diagnosable disorders, the relevance of these dimensional scores for child well-being is hardly trivial. For example, Fantuzzo et al. (2003) found that overactive behavioral problems were prodromal to or predictive of later socially disruptive behavior in the classroom and that underactive emotional problems were associated with later disengagement in play situations and poor academic outcomes. Although the occurrence of any single problem was low, at least 11% of a community sample of children had a problem that could be considered as serious. For this reason, screening of children in early childhood is a key first step in the prevention of later adjustment difficulties.

Challenges to the Validity of Screening Measures

A conceptual challenge to mental health assessment in young children arises from questions about the validity of discrepant rates of disorders related to gender, race/ethnicity, and economic status. Specifically, boys, ethnic minority, and children from impoverished backgrounds have more often been identified as having difficulties than girls, Whites, and children from affluent backgrounds (Barbarin & Soler, 1993; Raadal, Milgrom, Cauce, & Manci, 1994). Given its consistency across age and settings, higher rates of externalizing problems in boys than in girls probably reflects valid differences in psychological functioning more than flaws in measurement. However, the validity of claims about race and ethnic group differences has been challenged on both conceptual and methodological grounds (Rogler, 1993). These differential rates have been considered artifacts of assessment procedures, culturally inappropriate conceptualization of problems, or comparison to an inappropriate normative group or standard. For example, Cauce (1995) observed that using the normative standard for the CBCL results in much higher rates of clinical case designations among African American children ages 5 to 11 in comparison to other groups of children. These researchers challenged the appropriateness of the CBCL norms for drawing conclusions about the need of ethnic minority children for mental health intervention. Rogler (1993) went further, arguing that the high rates of designating children of color as requiring clinical intervention arose from their failure to recognize the role of culture. Central to this criticism was the assertion that the behaviors and emotions mental health assessments conceptualized as dysfunctional may have been adaptive within the sociocultural and physical environments of ethnic minority children and families (Canino & Spurlock, 1994). For example, withdrawn behavior, aggression, or high anxiety might be responsive and adaptive to an unpredictable and threatening environment (Dubrow & Garbarino, 1989). In addition, many studies reporting racial and ethnic differences confounded race/ethnicity with socioeconomic status. When income and parent education were controlled, race/ethnicity did not predict children's mental health status.

Although no simple resolution is evident to these conceptual and methodological challenges, especially the issue of cultural relevance, steps can be taken to reduce possible bias in mental health screening. One step is to increase the cultural and contextual relevance of the content by using assessment dimensions and indicators which are based on descriptive reports of parents and teacher concerns about children's adjustment difficulties that reflect the quality of children's responses to situational demands at school and at home. This is the approach adopted by Fantuzzo and colleagues in what might be called contextual assessment (Fantuzzo, Coolahan, Mendez, McDermott, & Sutton-Smith, 1998). Another approach is to base the organizing framework for assessment on dimensions derived from a sample representative of the children for whom the screening is designed (Fantuzzo, Manz, & McDermott, 1998). Both steps were followed in the development of the ABLE screening.

Obtaining information directly from children is often difficult because of children's limited capacity to self-monitor and the lack of a language for describing behavior and inner states (Fallon & Schwab-Stone, 1994). Consequently, we typically rely on information provided by adult caretakers and direct observation. Reports of child behavior by parents, teachers, or peers are susceptible to bias and distortion as are the reports of independent observers. In addition, parents and teachers may hesitate to label a child as having a problem without considerable evidence and certainty about the accuracy of their judgments (Fantuzzo, Bulotsky, McDermott, Mosca, & Lutz, 2003). Alternatively, parents or teachers who are under significant stress may consider behavior maladaptive that an independent observer might evaluate as typical or normal (Barbarin & Crawford, 2006).

The inability to judge the severity of children's problems may be one of the most common reasons teachers and parents hesitate to report problems. It is often difficult to judge whether problems are serious enough to require intervention or whether they are likely to resolve on their own. Many worrisome and disruptive symptoms such as opposition, crying, tantrums, and aggression occur commonly among children in whom a significant long-term disorder fails to materialize (Mash, 2003). Because many children of preschool age are still acquiring and mastering skills such as the self-regulation of attention, behavior, and emotions, parents and school staff may be uncertain about whether to characterize their bouts of disobedience, crying, inattention or fighting, as warning signs of imminent deviance or as perturbations that are integral to a child's movement toward healthy adjustment and social competence. Symptomatic behaviors reported by parents and teachers often are not stable over time and may not be severe enough to require intervention because they remit over time without intervention. To wit, ratings of problems in Pre-K account for only 9% of the variance of problems reported in the kindergarten year (Burchinal et al., 2004). Thus, it was not always clear whether children's failure to adapt to the regimes of the early childhood setting was a transitional issue that would resolve in time with firm guidance and consistent support, or prodromal to (a precursor of) serious and chronic disorder that would impede a child's academic progress and social adjustment.

Indicators of Problem Severity

Definitive judgments about the severity of a problem may be difficult to make but they are possible. Research on developmental psychopathology (see Mash, 2003) and the diagnostic rules of the *DSM–IV* (American Psychiatric Association, 1994) offer guidelines for determining severity in terms of problem duration, pervasiveness, and the level of impairment associated with symptoms. Problems that are persistent and of long duration tend to be more serious and follow a more adverse course. Similarly, the more settings and situations in which a problem is exhibited, the more disruptive its effects and less promising its prognosis. Pervasiveness of problems across settings has been employed as an indicator of severity of Pre-K children's emotional and behavioral problems by Lutz, Fantuzzo, and McDermott (2002). They examined adaptive and maladaptive behavior in the context of routine classroom situations. Severity was determined by the pervasiveness of problem behavior across situations encountered routinely in the preschool classroom. Situations include helping teacher, learning new tasks, intentional instruction in whole class setting, playing games, waiting in line, and free play. The contextual approach assesses severity in terms of the number of settings, situations, or contexts in which problem behavior such as aggression occur. For example, if aggression occurs in 7 out of 20 contexts, the child is identified as a having a problem.

Comparison to peers (children of the same race/ethnicity, social class, gender, or age) has also been a useful gauge of the atypicality and severity of problems. This approach may also dampen the ethnic or cultural bias noted earlier. It compares the symptomatic behavior of one child to that of peers with respect to such factors as the symptom's intensity, frequency, disruptiveness, and duration (Garber, 1984). A final indicator of severity is the judgment that a problem may require professional attention, often because it has not responded to parent or teacher intervention. Although the duration, pervasiveness, and adverse peer comparisons are useful, the degree to which a problem interferes with normal functioning at home, at school, or in the community is arguably the most sensitive and widely used barometer of severity. It incorporates the extent to which the quality of life and of relationships with others is compromised and the capacity for independence and growth is diminished (American Psychiatric Association, 1994).

Research Issues and Hypotheses

This research report focuses on the nature and prevalence of socioemotional concerns parents and teaches have about preschool children and the psychometric properties of ABLE, the screening tool used to identify those concerns. The report addresses several other questions. To what extent do parents and teachers agree on their concerns about children? How do teachers and parent's reports in the early part of Pre-K compare to referrals for IEPs at the end of Pre-K and at the end of kindergarten?

Accuracy in identifying problem behaviors in young children may be affected by factors which lead to under- or overreporting. The types of problems teachers report and the frequency with which they report them may be influenced by or related to teachers' education and years of experience, or the overall quality of the Pre-K program. Teachers with more experience and who have worked with many different cohorts of children may have more realistic expectations about what children are capable of, a wider basis for comparing behavior, and perhaps more accurate views of

what is normative and what is aberrant in children's behavior. Moreover, it is possible that reports of behavior problems will be more attributable to dysfunctional programs than to difficulties inherent in the child. In other words, behavior and emotional difficulties identified on the ABLE attributed to the child may arise from features of the classroom such as nonsupportive emotional environments and low quality programs with chaotic, ineptly managed, or emotionally toxic environments.

To address the issue of reliability and validity of judgments about severity, the study also examines whether the problems parents and teachers identify are associated with important outcomes such as academic competence. The proposed severity indicators are expected to predict referral for service not only at the end of Pre-K but also at the end of kindergarten. Specifically, children whose problems are characterized by either (a) longer duration, (b) higher levels of impairment, (c) symptoms that are worse than peers, (d) symptoms that occur in diverse settings, or (e) symptoms deemed serious enough to warrant professional attention are more likely to be referred for IEP evaluation and receive services before they go to first grade.

Referral to and receipt of special services by a child at the end of Pre-K and at the end of kindergarten is a stringent standard with which to judge the predictive effectiveness of a screening tool. Teachers and school systems vary greatly in the effectiveness, available resources, and efficiency of identifying children with problems that requires service referrals. There are inevitable inconsistencies across and within school districts in who will be referred for which services. Interdistrict differences in eligibility standards and intradistrict inconsistency in the application of its standards could mean that a child eligible for and receiving services in one school or district might not be recommended for or receive such services in a neighboring school or district (Fantuzzo, Stoltzfus, Lutz, Hamlet, Balraj, Turner, & Mosca, 1999). This lack of consistency introduces error in making cross-district predictions about which children will be referred for services. This suggests that the expected level of rate of selectivity and sensitivity would probably be lower than in the case where there is an objective and consistently applied standard. Nevertheless, in the absence of an absolute or more feasible and appropriate standard to gauge accuracy of case designation, IEP status is used.

Data reported here were collected in two studies: (a) The Multi-State study of the quality of public Pre-K programs and the outcomes of the children served by these programs (Clifford et al., 2005) and (b) the Mental Health Screening Study.

Study 1: Multi-State Study of Public Preschool Programs

Method

In each of 6 participating states, 40 randomly selected classrooms were observed and assessed for program quality using the Early Childhood Environment Rating Scale—Revised (ECERS-R; Harms, Clifford, & Cryer, 1998). A random sample of four 4-year-olds was selected stratified by gender. All children were assessed for literacy, language, mathematics, and socioemotional competence in the Fall and Spring of their Pre-K and kindergarten years. Excluded from the study were children with an IEP or children who lacked sufficient skills in English or Spanish to complete the assessment protocol. All interviewers had at least a bachelor's degree; many possessed advanced degrees and had prior research experience in early childhood settings. Interviewers were trained to reliably administer the ECERS-R and the child assessment protocol.

In addition to classroom observation, families of children in five states agreed to interviews and observations of parent–child interactions in their homes. Parents participating in the home interviews responded to questions on family life and children's functioning. Although they represent a selective subsample, they are comparable to the total multistate sample with respect to family income, poverty rates, parental health, child's language, race/ethnicity, and gender. Family interviewers were trained extensively to gather information on the social and familial environments of children. Interviewers used the preferred language of the family and in most cases, the ethnicity of the interviewer was matched to the ethnicity of the family. For a fuller description of the classroom sampling design and procedures, see Clifford et al. (2005) and for more details on the methods used in gathering data on families, see Barbarin et al. (2006).

Participants

The demographics of the parents and teachers whose data were used for these analyses are presented in Table 1. Most respondents were female. Otherwise, they were a very diverse group although the teachers were not as diverse a group as the families of the children they served. A majority of teachers had at least a bachelor's degree and had higher average incomes than families. A few states such as Georgia and New York have opened their programs to children without regard to family income and consequently, serve children across the income span. Consequently, even though half of the families reported incomes that fell below 150% of the poverty line, the sample of families in this study included highly educated and affluent families as well.

Measures

Ratings of Child Psycho-Social Functioning

ABLE Level I Screen was used to ascertain parents' concerns about children's psychosocial adjustment. Oppositional behavior, aggression, language, and dysregulation of attention and emotions were included on the ABLE I screening based on a review of 5 years of teacher referrals in a large urban Head Start program and

Table 1
Multi-State Study Demographic Characteristics by Respondent

	Teachers	Parents
Number	238	415
Mean years experience teaching 4-year-olds	9.7	
Estimated annual income	$34,000	$25,000
% Female	98%	99%
Race/ethnicity		
Euro-American	62%	48%
African American	17%	25%
Latino	14%	17%
% Who speak Spanish	27%	18%
% With bachelors degree or higher	58%	22%

on interviews with experienced early childhood teachers who were asked to identify the most frequently occurring problems of children who had difficulty adjusting to the Pre-K setting. Language has not ordinarily been included as a component of socioemotional adjustment. However, it was strongly endorsed by teachers as an essential component of any screening. In interviews, teachers drew a connection between language problems and difficulties in social and emotional adjustment. Teachers noted that the transition from crying to relying on language to express needs, wants, and pains was a good indicator of socioemotional maturation. They noted that children who lacked the ability to express themselves, to follow verbal instructions, and to take part in the give-and-take verbal exchanges often had difficulty entering peer groups and maintaining friendships. Because a child's capacity for self-expression was also critical in developing relations with caregiving adults, teachers judged children with severely limited expressive and receptive linguistic skills to be at risk for social isolation and peer rejection. Moreover, language problems along with aggression and opposition were among the most common reasons for requests for services. Accordingly, the ABLE I screen consisted of dichotomous ratings (yes/no) of Language Problems in addition to Inattention/Overactivity, Disobedience, Aggression, Bad Temper, Fearfulness, Sadness, and Peer Rejection.

As a part of ABLE I, parents and teachers responded to a set of questions that were indicators of problem severity. Problem severity was rated by teachers and parents in terms of whether it occurred over a long period, whether it impaired functioning, whether it occurred in multiple settings, whether the child was worse than peers on that dimension, and whether the problem seemed to require professional help to resolve. For each indicator, responses were coded to yield a dichotomous yes/no.

ABLE II Scales served as a follow-up screening for children rated identified as having at least one serious problem on the ABLE I screening. Continuous measures of children's attention, behavior (i.e., aggression, opposition), language, and emotion self-regulation formed standardized scales with which to corroborate the results obtained on ABLE I. Items for the scales were adapted from the attention, social maturity, opposition, anxious-depressed subscales of the factor-analytically derived Problem Behavior Index (Zill, 1985) and supplemented with items based on the diagnostic criteria for Oppositional Defiant, Conduct, and Hyperactivity disorders of the *DSM–IV*, 4th Edition (American Psychiatric Association, 2000). Items were rated on a five-point scale where 1 = *not a problem*; 2 = *a mild problem*; 3 = *a moderate problem*; 4 = *a serious problem*; and 5 = *a very serious problem.*

Initial construct validity of the ABLE II scales was established in Study 1 through exploratory factor analysis with data from the Multi-State sample and confirmed in Study 2 using data from the Midwestern urban and Southern rural samples. Exploratory and Confirmatory factor analyses used principle-components analysis in SPSS version 14. The factor structure was rotated using the Varimax algorithm and Kaiser Normalization and using .40 as the minimum criterion for factor loading. The exploratory factor solution accounted for 62% of the variance. The factor structure generated through these analyses was consistent with the theoretical rationale for the items and scales. Almost all items loaded on the scale factor for which they were initially designed. One notable exception was the "quiet withdrawn" item that loaded on the language factor instead of the emotional regulation factors for which it was originally intended. Table 2 presents the results of the factor analysis.

Additional evidence of construct validity for each of the scales arises from the significant correlation of ABLE scale scores with the teacher ratings using the Hightower (TCRS) and with the BPI scored scales (Zill, 1985).

Inattention/Overactivity (*Self Regulation of Attention*) consisted of 12 items that addressed issues of restlessness, impulsivity, and inattention and had high internal consistency (Cronbach's alpha = 90). The Inattention/Overactivity scale was positively correlated with the BPI Hyperactivity (.60) and Headstrong scales (.62). It was inversely correlated with the TCRS Task Orientation (−.40) and Frustration Tolerance (−.40) scales.

Aggression assessed the extent to which a child used physical force, verbal intimidation, and social pressure to threaten or harm others. The content of this four item scale included *being mean*, *bossy*, *grumpy,* and *teasing others* (Cronbach's alpha = .77). It was positively correlated with the BPI Headstrong scale (.59) and inversely correlated with the TCRS Frustration Tolerance (−.31) and Peer Social Skills (−.17), but was unrelated to Assertiveness (.00)

Opposition (*Self-Regulation of Behavior*) consisted of a 12 item scale based on symptoms from the DSM criteria for oppositional defiant behavior as stubborn, quick tempered, and disobedient (Cronbach's alpha = .90). Opposition was positively correlated with the BPI Headstrong scale (.68) and negatively correlated with Task Orientation scale of the TCRS (−.30).

Language used a 10-item scale to assess deficits in receptive and expressive language. Language items included the ability to communicate feelings and ideas and to comprehend the communications of others. Item selection was based on recommendations of a panel of speech and language experts and the scale was validated using both the Midwest urban and the Southern rural samples of preschool children from Study 2. Analyses of these data reveal that the Language scale has strong evidence of internal consistency (Cronbach's alpha = .90). The Language scale was positively correlated with the BPI immaturity subscale (.29) and was inversely correlated with TCRS Peer Social Skills (−.21).

Emotional Dysregulation used seven items to assess emotional agitation expressed in fear and crying, timidity, excessive worry, need for reassurance, obsessive thinking, and disturbances of mood or affect expressed as sadness and withdrawal. The Emotional Dysregulation scale was positively correlated with the BPI Anxious-Depressed (.60) and Immature (.53) scales and negatively correlated with the Frustration-Tolerance scale (−.26). The estimate of internal consistency was acceptable (Cronbach's alpha = .77).

Teachers used a five-point *Teacher Child Rating Scale* (TCRS; Hightower et al., 1986) to rate problem behavior and social competence. The *Behavior Problem* subscale included ratings of difficulties in several domains of functioning: behavior (e.g., disruptive in class, overly aggressive); emotions (e.g., anxious, unhappy); and learning (e.g., poor work habits, difficulty following directions). The *Social Competence* subscale assessed social skills such as assertiveness, frustration tolerance, task orientation, and relations with peers. The TCRS has strong support for its reliability and validity. Estimates of test-retest range from .61 to .91 and estimates of internal consistency range from .85 to .95 (Cronbach's alpha) (Hightower et al., 1986).

Table 2
Factor Analysis of ABLE II Items Supporting the Construct Validity of ABLE II Scales

Items	Factor loadings				
	ADD	OPP	AGG	EMO	LAN
Acts before thinking	.445				
Destroys things	.459		.526		
Poor concentration	.721				
Easily distracted	.712				
Makes careless mistakes	.706				
Moves from one activity to another	.698				
Overly active	.683				
Constantly moving	.654				
Clumsy	.553				
Impatient	.438	.569			
Tampers with things	.428	.448			
Argues too much		.694			
Stubborn		.684			
Unable to wait for what he or she wants		.651			
Does not heed warnings		.645			
Bad temper		.643			
Does not accept correction		.561			
Not sorry about misbehaving		.415			
Blames others		.402			
Mood changes		.419		.399	
Cries too much		.521			.402
Teases others			.781		
Starts fights			.744		
Bullies			.674		
Yells			.641		
Very dependent				.608	
Psychosomatic complaints				.580	
Anxious	.450			.498	
Very fearful				.463	
Fixated on certain ideas or subjects				.392	
Needs too much attention from adults				.410	
Unable to speak in sentences					.912
Hard to understand					.874
Can't converse					.857
Can't report on a recent experience					.824
Unable to answer questions about a story					.749
Does not put words together sensibly					.713
Undecipherable speech					.652
Easily confused					.607
Quiet/withdrawn					.539

Note. ADD = Attention/Overactivity Problems; OPP = Opposition; AGG = Aggression; EMO = Emotional Dysregulation; LAN = Expressive and Receptive Language.

An *end of school questionnaire* was used at the end of the Pre-K year and kindergarten. Teachers reported whether or not the study child had been referred for a disability assessment or an evaluation for behavioral/emotional problems. In addition, they reported whether or not the child had an IEP.

Observer ratings of child behavior (ORCB) was used for 2 days in the fall and 2 days in the spring, observations were conducted of each study child by trained observers at the Pre-K program (Ladd & Profilet, 1996). Observations were conducted in the classroom and on the playground. Observations focused on the following dimensions of behavior:

1. *Inattention/Overactivity*—distractible/overactive (child had trouble focusing on tasks, paying attention, staying with an activity; may be coupled with a higher-than-average physical activity; e.g., fails to persist at play activities or classroom tasks, quits activities or tasks too soon, never seems to complete or finish tasks, looking all over instead of focusing on task at hand, didn't seem to listen or hear what peers/teachers were saying, lacked sustained concentration, body or body parts constantly in motion/can't stop moving).

2. *Aggression toward peers* (child verbally or physically aggressive toward peers; e.g., hostile talk or gestures, yelled, threatened, bribed, insulted, hit, pushed, slapped, started fights, bullied peers, and fought back if provoked was also rated as aggressive).

3. *Domineering with peers* (child tried to dominate or attempted to control peer [classroom] activities to an excessive degree; e.g., bossy/controlling to peers,

sought to control peers' behaviors, play roles, activities, or conversations).

4. *Anxious/fearful* (child exhibited tense, anxious, fearful behavior; e.g., child seemed skittish, uncomfortable in peer situations, tense look on face, rigid, tense body posture or movements, cries a lot or sad, seemed worried/fearful).

5. *Withdrawn/asocial* (child appeared inhibited, shy, or disinterested with peers or peer activities, seemed to prefer to play alone/be by self, withdraws from peers' overtures/invitations, seems shy, apprehensive, and inhibited among peers)

6. *Excluded by peers* (child was actively excluded by peers from social interactions/activities; e.g., played alone because not welcome in peers' activities; peers tended to rebuff/reject child's overtures).

On each of these dimensions the observer rated the child using a five-point scale where 5 = *very–(e.g., distractible, aggressive, domineering, anxious, withdrawn, or excluded);* 3 = *somewhat–;* and 1 = *not very–.*

Coder Reliability

Just before data collection, coder reliability was tested in the fall and retested in the spring using videotaped segments from Pre-K classrooms. The most widely recommended method for assessing coder reliability (Cohen's Kappa) was not used because the low frequency and low variability of the target behavior could lead to a misleading estimate of reliability. In the case of low frequency behaviors Kappa yields artificially deflated estimates (.39), whereas its alternative—Percent Coder Agreement—yields inflated estimates (.99). To be considered reliable, coders had to give the exactly correct answer at least 70% of the time across all codes and had to get at least 50% exactly correct on any one section. On average, on their final test, data collectors gave the exact correct answer 90.04% of the time (*SD* = 5.75).

Direct Child Assessment

Peabody Picture Vocabulary Test 3rd Edition (PPVT-III)

The PPVT-III is an achievement test of receptive vocabulary that correlates well with other measures of language, literacy, and academic achievement (Dunn & Dunn, 1997). Children were shown a set of four pictures and were asked to select the picture that best represents the meaning of a word spoken by the examiner. A standard score is computed for this scale (median α = .94 as reported by authors; in this study Cronbach's alpha = .69 [fall] and .60 [spring]). The fall to spring test–retest reliability was .75.

Oral and Written Language Scale (OWLS) Oral Expression Scale

The OWLS Oral Expression Scale is a standardized measure of the understanding and use of spoken language (Carrow-Woolfolk, 1995). The examiner read aloud a verbal stimulus while the child looked at a stimulus board containing one or more pictures. Children responded orally by answering a question, completing a sentence, or generating a new sentence or sentences (test–retest reliability = .86). Correlations between the OWLS and achievement tests have ranged from .44 to .89 (Carrow-Woolfolk, 1995).

Woodcock-Johnson III Tests of Achievement

Applied Problems Subtest (Woodcock et al., 2001). The Woodcock-Johnson III Tests of Achievement includes standardized measures of academic achievement. The Applied Problems subtest examines the ability to analyze and solve math problems (Cronbach's alpha = .92 to .94).

Identifying Numbers

This criterion-referenced measure of knowledge of the Arabic number symbols from 1 to 10 (NCEDL, 2001). Children were shown a card with the numbers 1 through 10 arrayed randomly and asked to identify as many numbers as they could. The maximum possible score is 10 (Cronbach's alpha = .94 (fall) and .93 (spring); fall to spring test-retest = .75). See chart 1 for a summary of measures.

Chart 1: Overview of Study Measures

Study 1: Measures

ABLE I

A brief checklist for screening of parent or teacher concerns about attention, behavior, language, and emotions and of symptom severity.

ABLE II

Multi-item scales with clinical norms for problems of inattention/overactivity, aggression, opposition, language, and regulation of emotions.

Teacher Child Rating Scales

Multi-items standardized ratings of child problem behavior and social competence: disruptive and aggressive behavior emotions (e.g., anxious, unhappy); and learning (e.g., poor work habits, difficulty following directions). Competence scales social assessed assertiveness, frustration tolerance, task orientation, and relations with peers (Hightower et al., 1986).

Observer Ratings of Child Behavior (ORCB)

Observations of child in school by trained observer of problems of inattention/overactivity, aggression, anxiety, social withdrawal, and peer relations (Ladd & Profilet, 1996).

End of School Questionnaire

Questions on referral for a disability assessment and development of individualized educational plan (IEP) for child by end of the school year.

Peabody Picture Vocabulary Test 3rd Edition (PPVT-III)

A normed measure of receptive vocabulary (Dunn & Dunn, 1997).

Oral and Written Language Scale (OWLS) Oral Expression Scale

A normed measure of expressive language competence (Carrow-Woolfolk, 1995).

Woodcock-Johnson III Tests of Achievement: Applied Problems Subtest

A normed measure of early math skills (Woodcock et al., 2001).

Identifying Numbers

A criterion referenced measure of knowledge of numbers and counting (NCEDL, 2001).

Study 2: Measures

ABLE I

A brief checklist for screening of parent or teacher concerns about attention, behavior, language, and emotions and of symptom severity.

ABLE II

Multi-item scales with clinical norms for problems of inattention/overactivity, aggression, opposition, language, and regulation of emotions.

Follow-up Observation of Child

Children screened on ABLE as having serious concerns were observed by mental health professionals to corroborate the ABLE screening.

Follow-up Report by Teacher

Teacher.

Power Analyses

Power is broadly defined as the probability that a statistical significance test will reject the null hypothesis or the ability of a statistical test to detect an effect, given that the effect actually exists. Using the program Gpower 3.0, we find that the sample size for Study 1 was sufficient to detect significant correlations with a power of .70 and to detect mean differences with a power of .90. This post hoc power analysis demonstrated that the sample size gave the study sufficient power to detect even small effects. Finally, a chart of the measures for Studies 1 and 2 is included to provide an overview of the measures used in these validation studies (see Chart 1).

Results

Prevalence

Problem prevalence was assessed using weighted data. Pre-K teachers identified fewer children with problems than did parents (10.5% vs. 18.4%) but the proportion identified by kindergarten teachers was more than double the percent (23.1%) identified by Pre-K teachers. Table 3 presents the percentages of children rated by parents, Pre-K teachers, and kindergarten teachers for specific problems. The proportion of children in the sample with mental health concerns was remarkably close to prevalence rates reported in early studies of children's mental health. Inattention/overactivity, disobedience, and aggression are the most prevalent concerns identified across the entire sample of Pre-K children. Except for inattention/overactivity (13%), the kindergarten teachers reported problems at about the same rate as parents but the rates reported by Pre-K teachers were lower on most problems. When percentages were calculated only for children with a concern that was severe, the proportion who had bad temper was quite high (53.8%). Also, common within this group were language difficulties (30.0%), disobedience (29.2%), aggression (27.8%), and inattention/overactivity (25.7%).

Comorbidity of Problems

Comorbidity of problems was high. Children who had one problem were very likely to have at least one more. Of the children for whom parents had at least one serious concern; 31% had two and 15% had three or more. When we examined which concerns were more likely to co-occur with a second concern, the comor-

Table 3
Weighted Frequency of Mental Health Concerns Expressed by Parents and Teachers About Preschool Children

	Parents at Pre-K	Pre-K teacher	Kindergarten teacher
Bad temper	11.1%	NA	NA
Inattention/overactivity	5.0%	5.6%	13.0%
Disobedience	5.2%	5.4%	6.8%
Language	5.8%	NA	NA
Aggression	5.6%	3.5%	4.5%
Fearfulness	2.4%	1.0%	2.0%
Sadness	2.6%	1.4%	2.4%
Peer rejection/social withdrawal	1.4%	1.4%	1.7%

Table 4
Correlation of ABLE I Ratings of Parental Concerns With Ratings of Problems by a Teacher (TCRS, Teacher Child Rating Scales) and External Observer (ORCB, Observer Rating of Child Behavior)

ABLE Level I	TCRS Scales	ORCB
Bad temper	.13*	.07
Disobedience	.20***	.12*
Inattention/overactivity	.26***	.06
Language	.05	.05
Fearfulness	.13*	.12*
Aggression	.25***	.17***
Sadness	.01	.11*
Peer rejection	.12*	.01

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

bidity rate ranged from 50% for language to 97% for bad temper. Bad temper, aggression, and disobedience were a frequently observed combination.

To determine if teachers or program characteristics were associated with problems attributed to children, several statistical tests were conducted. Chi-square Tests of association were conducted to tests associations between teacher education, experience, race/ethnicity, and ECERS program quality rating with 10 possible concerns. None of these associations was significant. One-way ANOVAs were computed with the problem scales as dependent variables and teacher education, years of experience, and program quality each serving as the independent variable. None yielded significant results. Pearson product–moment correlations were used to test the relationship between program quality, years of education, years of experience, and program quality as measured by ECERS with ratings of problem severity. Only one correlation was significant: A modest correlation between the ECERS language/interaction factor and rating of problem severity ($r = -.08$, $p < .02$).

Parent-Teacher and Parent-Observer Agreement

Next we analyzed data relevant to the validity of ABLE Level I concerns and severity index by correlating parental reports with teacher report of problems on the Teacher Child Rating Scales (TCRS) and observer ratings of child behavior at school (ORCB). Table 4 presents the correlations of Level I problem designations by parents with the corresponding ratings from the TCRS and the ORCB. With the exception of Language, the correlations of ABLE 1 concerns with the corresponding teacher rating or observational score were significant. With the exceptions of Language and Sadness, the correlations between parent and teacher ratings were significant. The effect sizes were very small (.01 to .04).

A second set of analyses related to interrater agreement was conducted. Table 5 contains a correlation matrix between parent-rated and corresponding teacher-rated scales performed in the Pre-K year. The constructs in the teacher ratings (TCRS) were similar to but not identical in every instance to the constructs rated by parents. Correlations between scales that measure comparable constructs for parents and teacher are located on the diagonal of the correlation matrix. Parent-rated scales correlated significantly with the comparable teacher-rated scales and in the expected direction. For example, Inattention/Overactivity from the ABLE was inversely correlated with Task Orientation from the TCRS. The effect sizes (*ES*) were small (*ES* = .01 to .09). This correlation matrix provided modest support for the construct validity of the scales.

ABLE I ratings of aggression scale from Pre-K were significantly correlated with observer ratings of aggression in the fall ($r = .18$, $p < .01$, $ES = .05$) and in the spring ($r = .14$, $p < .01$, $ES = .02$) of the Pre-K year. Similarly, observations of child hyperactivity and attention problems were correlated with parents rating of child attention scale both in the fall ($r = .16$, $p < .01$, $ES = .03$) and in the spring ($r = .12$, $p < .05$, $ES = .01$). Observed fearfulness was related to the rating of emotional regulation in the fall ($r = .10$, $p < .05$, $ES = .01$) but not the spring. Teacher rating of language skill was correlated with parent rating of language deficits in both the fall ($r = .09$, $p < .05$, $ES = .01$) and in the spring ($r = .13$, $p < .01$, $ES = .02$).

Relation to Child Academic Skills

Table 6 presents correlations between ABLE II scales for Inattention/Overactivity, Aggression, Opposition, Language, and Emotional dysregulation and academic skills in the fall and spring of kindergarten. These skills include receptive vocabulary, expressive language, math, and counting. The Language scale was significantly correlated with PPVT and OWLS in the fall (*ES* = .04, .07, respectively) and in the spring (*ES* = .06, .11, respectively).

Table 5
Correlations Between Parent-Rated ABLE II Scales and Teacher-Rated Hightower TCRS

	Parent rated scales					
Teacher-rated Scales	Total Problems	Inattention/ overactivity	Oppositional	Aggression	Language	Emotional dysregulation
Total problems	.28***	.33***	.21***	.17**	.18***	.18***
Task orientation	−.25***	−.31***	−.21***	−.12*	−.19***	−.10
Conduct problems	.29***	.34***	.24***	.16**	.12*	.20***
Frustration tolerance	−.21***	−.23***	−.18***	−.12*	−.11	−.12*
Language competence	−.14**	−.17**	−.11*	−.12*	−.14**	−.05
Internalizing problems	.13*	.13*	.07	.12*	.14**	.12*
Social competence	−.23***	−.25***	−.17***	−.17**	−.20***	−.13*

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 6
Correlations of Pre-K Cognitive Measures With ABLE II Scales

	Inattention/overactivity	Aggression	Opposition	Language	Emotional dysregulation
PPVT: Fall	−.18**	−.18**	−.11**	−.21**	−.03
PPVT: Spring	−.13**	−.16**	−.07	−.25**	.06
OWLS: Fall	−.23**	−.15**	−.13**	−.27**	−.13**
OWLS: Spring	−.16**	−.14**	−.10*	−.30**	−.01
WJ math fall	−.26**	−.16**	−.15**	−.22**	−.06
WJ math spring	−.26**	−.16**	−.13**	−.25**	−.01
Numbers - fall	−.14**	−.04	−.09*	−.06	.01
Numbers - spring	−.10*	−.04	−.04	−.06	.10*

* $p < .05$. ** $p < .01$.

These results provide modest support for the link between Inattention/Overactivity, Aggression, and Opposition to academic functioning. The more problems a child had, the lower the academic functioning in both fall and spring.

Severity Indicators

Table 7 presents the correlation among the severity items that made up the severity screening criteria for parents and teachers. The relationships among indicators appeared to be similar across all raters. Problem severity was strongly related to the other indicators. This suggests that these criteria may capture or represent constructs that are at the core of the severity indicators.

Of the children for whom parents expressed concern, 7% did not meet criterion on any of the severity indicators; 37% met the criterion on 1; 40% on 2 to 3 indicators; and 16% on 4 to 6 indicators. Table 8 shows how many children met each of the severity screening thresholds rated by parents in the Pre-K year, Pre-K teachers, and kindergarten teachers. Most children had problems that lasted 6 months or more. The duration criterion was not rated by teachers. Of the remaining criteria, adverse peer comparison was passed by the largest number of children in the Pre-K teacher ratings and impairment was most commonly passed by children in kindergarten teacher rating. All other screens were passed by moderate proportions of children (20–40%). For Pre-K teachers, problem pervasiveness and adverse peer comparisons were most often endorsed. For kindergarten teachers, adverse peer comparisons were endorsed most often and at twice the rate of the other severity indicators.

The next set of analyses compared children who were screened as cases using the severity indicators to children who did not meet the threshold of severity based on the severity ratings of the parent. Using the criterion of endorsing two or more severity indicators, 18.4% of the children were identified as cases by parents, 10.1% by Pre-K teachers, and 18.2% by kindergarten teachers. A demographic comparison of children identified as cases to children not screened reveals no significant differences with respect to poverty status, race/ethnicity, language, mother's education, or state of residence. Moreover, the case and noncase groups did not differ in

Table 7
Correlations Among Severity Indicators for Parents, Pre-K Teachers, and Kindergarten Teachers

	Long duration	Impaired adjustment	Worse than peers	Occurs across multiple settings	Requires professional help
Long duration	Parent				
	Pre-K teacher				
	Kindergarten teacher				
Impaired adjustment	.24*				
Worse than peers	.21*	.39***			
	—	.63***			
	—	.41***			
Occurs in multiple settings	.41***	.34***	.26**		
	—	.21*	.30***		
	—	.26***	.30***		
Requires professional help	.18	.43***	.24*	.33***	
	—	.69***	.57***	.25**	
	—	.55***	.41***	.42***	
Actually referred for services	.18	.42***	.25**	.33***	.63***
	—	.24**	.10	.81***	.35**
	—	.10	.03	.78***	.29***

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 8
Percent of Children Rated As Having Serious Problems Who Met Each Severity Indicator

	Rater		
	Parents	Pre-K teacher	Kindergarten teacher
Problem has occurred for over 6 months	88.6%	—	—
Child's behavior is worse than peers	37.0%	48.9%	67.2%
Child problem impairs adjustment to Pre-K	31.6%	22.4%	33.9%
Problem occurs in multiple settings	30.2%	64.7%	27.1%
Problem needs professional help to correct	25.3%	24.2%	36.6%
Child referred for evaluation and services	21.0%	30.9%	30.0%

their fall or spring Pre-K scores on PPVT, Woodcock Johnson applied problems (math), knowledge of alphabets, or ability to count.

Table 9 presents means on the ABLE II total problems scale scores and the TCRS problem and competence scales (Pre-K and kindergarten) by case status (case vs. noncase). The term case is applied to a child who is screened positively as having at least one concern that for which more than one severity item was endorsed. In practice settings, this might be a child whose risk of problems is serious enough to merit follow-up. Case status was related as expected to parent and teacher ratings of problems in Pre-K and in kindergarten. A child was designated as a "case" when parent or teacher reported a concern and endorsed two or more severity items related to that concern. In each instance, the differences were significant with the cases being higher on problems and lower on competence than the noncases. As further evidence of the consistency in case designations, there was a moderate level of agreement on which children were cases and which were not between Parent and Teachers (76% for Pre-K and 72% for kindergarten) and between Pre-K and kindergarten teachers (80%).

Sensitivity and Selectivity of Severity Indicators

Next, data are provided on the effectiveness with which children were screened by parents and teachers as having severe problems. We examined how well the screen selected children who received special education services at the end of Pre-K and kindergarten using ABLE screens completed by parents in the middle of the Pre-K year. By the time children entered first grade, about 11.9% had been referred for evaluation and/or received special education services; 7.3% in Pre-K and 8.6% in kindergarten. Most children referred for IEP in Pre-K programs have historically been for language difficulties. By convention, the sensitivity of a screening tool addresses the question of how many of the children who in time become eligible for special services are identified in earlier screening. The sensitivity rate in this case would be the percentage of children who received special services by the end of kindergarten (sometimes labeled "true positive") who were screened by ABLE as a case, that is, having a serious concern.

Selectivity is the ability of the instrument to rule out or screen out children who truly do not have serious problems (sometimes referred to as "true negatives"). This addresses the possibility that an instrument might miss children who truly have difficulty. In this case, the selectivity rate is the percentage of children who are not referred for services who are screened out (i.e., not designated as having serious problems) by the ABLE.

The sensitivity and selectivity of the severity indicators are presented in Table 10. Estimates are presented separately by rater (parents, Pre-K teacher, and kindergarten teacher) and by time (end of Pre-K and end of kindergarten). The sensitivity rate is marginally acceptable for parents, poor for Pre-K teachers and

Table 9
Means, Standard Deviations and MANOVA Univariate F Values for ABLE II Total Scale Score and Hightower TCRS Scores by Case Status (Parent Rated)

	Groups		
	Non-Cases (children without serious problems) (n = 261)	Cases (children screened with serious problems) (n = 82)	Univariate F-Value
Pre-K year			
ABLE II-total problem score	46.75 (7.0)	58.57 (9.8)	124.6***
TCRS total problem score	1.42 (.5)	1.62 (.6)	8.9**
TCRS competence	3.70 (.7)	3.33 (.8)	12.0***
Kindergarten			
TCRS total problem score	1.50 (.04)	1.70 (.07)	6.32***
TCRS competence	1.63 (.06)	2.03 (.11)	10.46***

Note. Multiple $F(3, 339) = 42.3$. $p < .001$.
** $p < .01$. *** $p < .001$.

Table 10
Sensitivity and Selectivity of Severity Indicators in Predicting to IEP Status at End of Pre-K and at the End of Kindergarten

Sensitivity	
Selectivity	
IEP status End of Pre-K	
Parent	
.59	
.83	
Pre-K teacher	
.42	
.91	
IEP status End of K	
Parent	
.60	
.84	
Pre-K teacher	
.36	
.91	
Kindergarten teacher	
.65	
.80	

good for kindergarten teachers. The selectivity rate of the ABLE was very good for all raters at both time points. Overall 4.6% of children in the kindergarten sample had problems serious enough that they were referred for IEP evaluations at the end of the year. In this group were 25% of the children identified as cases by parents in Pre-K because of problems with aggression and 38% of the children identified as having a serious problem related to speech and language. At the end of kindergarten, 13.1% of the children about whom parents had a serious concern were referred for IEPs compared with 3.7% of children for whom a serious concern was not expressed. This means that cases identified in Pre-K by parents using ABLE were 3.5 times more likely than noncases to be seen as having serious problems.

Study 2: Pre-K Mental Health Screening

Study 1 was used to develop and assess the psychometric properties of both ABLE I screening and severity items and norms for the ABLE II scales. The purpose of Study 2 was to replicate the psychometric assessment of ABLE I and II by providing additional evidence of the validity and reliability and to assess the effectiveness of ABLE I and II to provide useful screening information when used in practice as part of a system of mental health assessment and service delivery in urban and rural preschool settings.

Method

Participants

The children evaluated in Study 2 included both three and 4-year olds enrolled in the Head Start program operated by a school district in a large Midwestern city ($N = 4,843$) or in early childhood programs receiving partial state financial support in three rural counties in a Southeastern state ($N = 734$). Over 80% of the children in the Midwestern Head Start sample were from low-income African American or Latino families. Respondents from the Southern rural counties were approximately 70% White and low income.

Procedures

In contrast to the multistate data that were gathered using personal interviews as part of a larger research study, the mental health screenings in Study 2 were integral to the service delivery in early childhood programs. As part of his or her duties, teachers and teaching assistants completed ABLE I on each child using paper and pencil forms. Before completing the ABLE, Teachers received 1 day of training on the constructs and methods for completing the ABLE. Parents completed the ABLE with the assistance of a teacher or family coordinator at the time of registration or soon thereafter. All respondents completed the 10 ABLE I screening items and if there were a concern they completed the seven severity items for the most pressing of their concerns. If they endorsed at least two severity items, they were instructed to complete the 40 items for the ABLE II scales. Use of the normed scales of ABLE II in this way provides another check on or confirmation of the ABLE I ratings. All assessments were completed within the first eight weeks of the program. Parents of children enrolled in these programs completed the screening using paper and pencil format. A Spanish language version was available to parents who preferred Spanish.

Teachers in the urban program entered their own responses to ABLE I and II received the results online. Urban parents and both parents and teachers in the rural programs responded on printed forms that were gathered and scored centrally. Teachers screened 66% of the urban children and 87% of the rural children. In addition, ABLE screenings were also completed by the parents of 83% of the urban and 56% of the rural children. Results were shared with teachers and parents when either reported a serious concern. In all settings, social workers and behavioral health consultants with training on the ABLE followed up with the teachers and parents who reported serious concerns on the ABLE. They first corroborated the accuracy of the screening results by interviewing teachers and parents and by observing the child at school. Less than 1% of these follow-ups resulted in a decision that the ABLE erred in its determination that either the parent or the teacher had a serious concern about the child. Once problems were corroborated, follow up services were provided by program staff and consultants or the children and families referred to language therapists, pediatricians, or psychologists (Barbarin, 2006).

Results

A confirmatory factor analysis was completed on the combined urban and rural samples for the ABLE II scales using a procedure similar to that used in the original construct validation, that is, principal components analysis, Varixmax rotation with Kaiser Normalization in the final solution and using .40 as the minimum criterion for factor loading. In six iterations, the final factor solutions replicated the five-factor solution generated in Study 1. Moreover, the Cronbach's estimates of internal consistency were very similar to those obtained in the analysis of the Study 1 sample: .91 for *Inattention/Overactivity,* .85 for *Aggression*, .91 for *Opposition*, .86 for *Language*, and .73 for *Emotional Dysregulation.*

Teachers in the urban Head Start identified 8.8% of the children screened as having a serious problem and parents identified 15.3%. Parents and teachers agreed 77% of the time about children who

did not have problems. In all, 880 children (18.8%) were identified as cases and referred for mental health services because of serious concerns expressed by either parents or teachers.

Table 11 presents the prevalence of problems reported by Pre-K teachers and parents in the urban sample. Parents reported concerns more often than Pre-K teachers (15.3% vs. 8.8%). Among cases identified by parents and teachers combined, prevalence was approximately 10% for each of the areas of concern. Teachers reported disobedience and language most often; parents most often reported aggression, emotional dysregulation, and inattention. Table 12 presents reports of problems identified in the mental health screening by the urban sample of teachers and parents disaggregated by gender. Overall teachers and parents reported more problem behavior in boys than in girls. When social workers completed follow-up interviews within 4 weeks of the screening, the rate of children incorrectly identified as having a serious problem (false positives) was extremely low.

In the rural sample, 29% of children had serious issues according to parents and 34% of children were screened for a serious concern by teachers on ABLE I. Parents and teachers agreed about whether there was a serious concern for 70% of the children. The rates of serious problems reported on ABLE I did not differ by age but serious concerns were expressed more often for boys (43%) than girls (27%), $\chi^2(1) = 21.4$, $p < .001$. For rural parents, there were no differences by gender in the number of children identified with a serious concern but children 3 years of age and younger were more often identified as cases on ABLE I (35%) than children 4 years of age and older (24%), $\chi^2(1) = 5.7$, $p \leq .05$. Underscoring the value of the ABLE II follow-up screening, for both parents and teachers neither age nor gender were significant in tests of mean differences for any of the six ABLE II scales.

Discussion

The frequencies of specific problems detected by ABLE in the two studies reported here are consistent with prevalence rates reported in other studies (Costello, 1988; Costello, Angold, Burns, Erkanli, Stangl, & Tweed, 1996; Fantuzzo, Bulotsky, McDermott, Mosca, & Lutz, 2003; Offord, Boyle & Racine, 1989). Between 10% and 20% of children in Study 1 were identified as having a serious adjustment problem. Rates for rural and younger children were higher. Overall, problems with behavior and inattention/overactivity are the most prevalent concerns expressed by parents and teachers. The prevalence of inattention/overactivity is noteworthy and may be a consequence of an increasing academic focus being adopted in Pre-K. The accelerated demands for didactic instruction largely in response to the No Child Left Behind Act are transforming the nature of Pre-K and kindergarten classroom interactions and practices. Beginning with Pre-K, children increasingly are being asked to accommodate to a stricter, more academically demanding regime than was true in the past. The screening of attention problems in a child should lead not only to efforts to change child behavior but should also give rise to questions about how a preschool program and its curriculum might accommodate to the developmental needs of young children. Fewer than half the children had just one problem. When a problem existed, it tended to occur in tandem with other problems. For example, the data showed a high comorbidity of bad temper and aggression with other problems.

No relationship was found between teacher characteristics and report of child problems. Teachers' race/ethnicity, education, and Pre-K teaching experience were unrelated to reports of child problems. There was a modest inverse relationship between the program quality as measured by the ECERS language interaction factor and child problem rating. In programs with higher quality language stimulation and teacher student interactions, teachers reported fewer problems.

The SES and race/ethnic group differences noted with measures such as the CBCL (Achenbach & Edelbrock, 1991) were not found in the ABLE screenings. The lack of ethnic and income differences in the selection of children as cases may result from ratings that make judgments in comparison to peers, use of severity indicators, and the development of the ABLE on sample that included a large proportion of children drawn from a population of children at risk because of poverty status. Consequently, the norms may be more appropriate for this group of children (Raadal, Milgrom, Cauce, & Manci, 1994)

Weighing the accuracy of information on children's problems is a critical step in determining how much credence to give the estimates generated in this study. Taken together, the psychometric data reported here provide support for the reliability and for discriminant and convergent construct validity of the screening scales. Estimates of internal consistency fall in the acceptable range for each of the ABLE II scales and support for the construct validity of ABLE I and II ratings is convincing. Triangulating the data of parents, teachers, and trained observers is a common approach to estimating accuracy. For most problem ratings, higher concordance exists between parent and teachers than between parents and classroom observers. Observers are in the class for a

Table 11
Concerns Identified for Children in the Mental Health Screening Study by Problem Type and by Rater

	Children rated by Pre-K teachers (T) (n = 4,843)		Children rated by parents (P) (n = 4,012)		Children rated by either T or P (n = 4,843)	
Concerns	*N*	%	*N*	%	*N*	%
Inattention/overactivity	191	3.9	357	8.9	514	10.6
Disobedience	211	4.4	340	8.5	508	10.5
Emotional dysregulation	155	3.2	365	9.1	511	10.5
Aggression	157	3.2	361	9.0	488	10.0
Language	216	4.5	275	6.9	450	9.0

Table 12
Prevalence of Problems Reported on ABLE I Within the Group of Children Identified as Cases by Reporter and Gender of Child

	Teacher (T) rated		Parent (P) rated		Either T or P rated	
Concerns	Boys	Girls	Boys	Girls	Boys	Girls
Inattention/overactivity	12.1%	5.4%	19.0%	11.9%	25.7%	15.6%
Disobedience	17.3%	7.7%	16.6%	11.5%	27%	15.7%
Emotional dysregulation	10.3%	7.6%	18.4%	17.0%	24.7%	21.7%
Aggression	10.4%	4.5%	16.4%	12.4%	22.8%	14.8%
Language	14.0%	6.4%	15.6%	9.6%	22.9%	13.4%

limited time. Parents and teachers have the advantage of observing children over a much longer period and over more setting. This difference is very likely to affect ratings of emotional dysregulation that can be discerned over time and low probability behaviors such as aggression that may require very long periods to witness. Modest correlations can be accounted for in part by the fact that parents, teachers and, in this case, classroom observers have access to somewhat different behaviors that correspond to the variations in home, school, and community settings and the periods over which observations were possible. For these reasons, although, parent-teacher-observer agreement can be informative, they do not constitute an unassailable index of reliability or accuracy of information on children's problems.

ABLE I includes an algorithm for gauging the severity of concerns about children that are not transient and have the potential for becoming deleterious to the child's life at home and at school. The severity indicators predict later referral in that children who meet the threshold for problem severity are 3 to 8 times more likely to be referred for services in Pre-K or kindergarten. Severity indicators can also help resolve parent and teacher ambivalence about labeling a child as disordered on the one hand and not wanting to ignore serious problems that might be nipped in the bud on the other hand. Even when a parent or teacher is clear that some form of intervention is warranted, they may be uncertain about (a) how intensively to respond, and (b) what level of resources should be enlisted in addressing the problem. At the program level, severity indicators can provide guidance to staff about where to begin intervention. For example, high frequency low severity problems might be addressed most efficiently through staff training and program adjustments. Very serious concerns may be addressed through child-centered classroom consultation or individualized interventions.

Limitations

ABLE is neither a comprehensive screen nor an instrument that produces a DSM clinical diagnosis. Its scope is limited to concerns most commonly cited by early childhood staff and most often identified in the research literature on young children's mental health. Other problems, for example, autism spectrum disorder, are not covered by this screening method. ABLE distinguishes children who do not have serious problems from children with overt symptoms that may be indicative of or precursors to serious disorders. These are problems of midrange difficulty that often respond to intentional, focused intervention by teachers. ABLE is not able to distinguish between symptoms that are serious enough for intervention by parents and teachers from problems that involve more subtle symptoms of neurological defects or psychopathology. Neurological problems or serious psychopathologies require assessment that is more intensive, personal interview and observations by persons with refined diagnostic skills. Thus, as a screening tool it is different from diagnostic assessment that usually entails multiple methods, multiple reporters generating information that is integrated into complex formulations and leading to intervention that is more intensive. These are beyond the purview of any single and simple assessment instrument.

The job of validating and testing the psychometrics of any measure is never complete. This is certainly true of ABLE. The effect sizes of the correlations on claims of validity rest are small. Although the data support the reliability of the measure as robust, additional work is needed to strengthen claims of construct validity. Also needed are additional studies of selectivity and sensitivity using criteria more appropriate than IEP referral.

These limitations notwithstanding, the evidence supports the assertion that ABLE is a valid and reliable instrument for use in preschool screening of children who are at risk of problems of attention, behavior, language, and emotions that might interfere with their adjustment and success at school. This instrument is noteworthy for its relevance to common problems in Pre-K, its brevity and its efficiency. It can be completed in a few minutes. It includes an empirical algorithm for assessing severity of concerns. It also uses a two-step procedure in which the longer normed scales are completed only if the concerns pass the initial threshold of severity. Moreover, the normed scales provides a check and balance for the initial screening by comparing parent and teacher responses to those of a representative group of children in public sponsored Pre-K.

Implications for Practice and Policy

The results of screening with the ABLE underscore the importance of mental health evaluation in early childhood. A substantial number of children (15–20%) enter Pre-K with concerns that could interfere with long-term adjustment if they are not addressed. As states move toward universal access to Pre-K for 4-year olds within public schools, it has become a platform on which to build an integrated approach to mental health evaluation and services such as that called for by Knitzer (2000). As an example, Barbarin (2006) has proposed a cost-effective model of mental health service delivery that is based on universal screening at the start of Pre-K and using screening data to design professional development for teachers, classroom consultation for problem behaviors,

and referral to mental health and language specialists for the most serious cases. Nevertheless, the challenges of implementing effective universal screening and service provision in Pre-K are many and should not be taken lightly. Even if adequate financial resources were available, competing demands on time and already high reporting requirements may make Pre-K teachers and parents less than enthusiastic reporters. Problems that stem from family distress, inappropriate "drill and kill" curricula, ineffective teaching practices, and classroom disorganization may be incorrectly attributed to the children themselves. Dangers of over- and underreporting also exist. Demoralized teachers may underreport problems for fear of stigmatizing children or because reporting seems futile because services are not available.

A principal motivation of mental health screening is to identify children who need and who might benefit from intervention. Services were provided by the programs in which the Mental Heath Screening study was conducted (Study 2). Lack of access to appropriate mental health services is still a major problem for children. Despite numerous studies demonstrating the mental health needs of children, only 21% of children nationally who might benefit from mental health screening and evaluations actually receive them (Burns, Phillips, Wagner, Barth, Kilko, Campbell, & Yandsverk, 2004) and only 20% of those who needed services were provided them (Kataoka, Zhang, & Wells, 2002; Knitzer & Cooper, 2006) Moreover, only 1 in 7 of the children from the Multi-State Study who were identified on the ABLE in Pre-K as having a serious problem actually received professional services during the year or were referred for an IEP by the end of kindergarten. It is possible that these children will be false positives who are misidentified by ABLE because of measurement error. This explanation is not tenable in light of the low rates of false positives found in follow-up phase of the Mental Health Screening Study (Study 2). It is also possible that the children in the Multistate study simply got better by the end of the school year through the intercession of skilled classroom teachers or on their own. For a variety of reasons some children did get better and no longer needed services but this was not true of all of them. Another plausible explanation is that ABLE screened for problems that still are not categorized as eligible for public sponsored services. Even if all these are true, many children remain who needed services but did not receive them.

Many programs face tough choices about how to allocate their meager behavioral health resources. Some programs in our study implemented policies that restricted services to language problems or developmental disabilities such as Autism and Mental Retardation. The dilemma is a policy, not a conceptual, issue. Even when resources exist, policy decisions must be made about how intensively parents and teacher should intervene before referral is made for professional services.

The resource problem is serious and has generated much discussion but little action. Several policy changes could make a difference. An important first step toward address the mental health needs of children is universal access to health care for children that place mental health on a par with physical health. A second step that could be taken immediately is to broaden the definition of disability in IDEA (Disabilities Education Improvement Act of 2004) to include incapacitating behavioral and emotional problems. So few children in public Pre-K were referred for IEP evaluation and received special education services because of policy decisions to restrict covered services to speech and language problems. Consequently, a majority of the referrals for special education services at the Pre-K level are for language deficits in the school districts participating in this study.

Conclusion

A substantial number of children enters Pre-K with a variety of behavioral, attentional, and emotional difficulties that threaten to undermine their well-being and their success in school. Expanded access to early childhood programs especially for children at-risk for behavioral difficulties offers a remarkable opportunity for systematic early intervention. Universal screening in these programs is an efficient and necessary first step. Assessing young children's mental health status is not an easy task. Recommending the use of a fairly simple screening tool is not intended to trivialize or underestimate the complexity of children's mental health or to suggest that the task of assessing it is over once an instrument such as the ABLE has been completed. ABLE is not a clinical diagnostic instrument nor can it substitutes for the clinical judgment of skilled mental health professionals. It is a method for gathering information about the concerns parents and teachers have about preschool children organized in a way that facilitates decision making about whether the concern is serious enough to seek professional consultation and follow up. In addition to facilitating decision-making about individual children, it can be used as a universal screening instrument to aid the allocation of mental health resources and the planning of specific multilevel interventions. Assessment with any instrument is just the beginning.

References

Achenbach, T., & Edelbrock, C. (1991). *Manual for the CBCL and 1991 profile.* Burlington, VT: University of Vermont, Department of Psychiatry.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Barbarin, O. (2006). ABLE: A system for mental health screening and care for preschool children. In E. Moore & B. Bowman (Eds.). *School readiness and social and emotional development: Exploring diverse cultural perspectives* (chap. 6). Washington, DC: National Black Child Development Institute, Inc.

Barbarin, O. (2006). ABLE manual. Chapel Hill, NC: HPT1, LLC.

Barbarin, O., Bryant, D., McCandies, T., Burchinal, M., Early, D., Clifford, R., et al. (2006). Children enrolled in public pre-k: The relation of family life, neighborhood quality, and socio-economic resources to early competence. *American Journal of Orthopsychiatry, 76,* 265–276.

Barbarin, O., & Crawford, G. (2006). Acknowledging and reducing stigmatization of African American boys. *Young Children, 61,* 79–86.

Barbarin, O., & Soler, R. (1993). Behavioral, emotional and academic adjustment in a national probability sample of African American children: Effects of age, gender and family structure. *Journal of Black Psychology, 19,* 423–446.

Behar, L., & Stringfield, S. (1974). A behavior rating scale for the preschool child. *Developmental Psychology, 10,* 601–610.

Berls, A. T., & McEwen, I. R. (1999). Battelle developmental inventory. *Physical Therapy, 79,* 776–783.

Blankemeyer, M., Culp, R. E., Hubbs-Tait, L., & Culp, A. M. (2002). A tool for identifying preschoolers' deficits in social competence: The preschool taxonomy of problem situations. *Education and Treatment of Children, 25,* 208–223.

Bricker, D., Squires, J., & Mounts, L. (1995). *Ages and Stages Question-*

naire: A parent-completed, child monitoring system. Baltimore: Brookes.

Brigance, A. H. (1991). *BRIGANCE diagnostic inventory of early development, Rev. ed. (BDIED-R*). North Billerica, MA: Curriculum Associates.

Burchinal, M., Pianta, R., Howes, C., Bryant, D., Early, D., Clifford, R., et al. (2004). Predicting child outcomes at the end of kindergarten from the quality of Pre-Kindergarten teaching, instruction, activities, and caregiver sensitivity. Manuscript submitted for publication.

Burns, B., Phillips, S., Wagner, H., Barth, R., Kilko, D., Campbell, Y., et al. (2004). Mental health need and access to mental health services by youths involved with child welfare: A national survey. *Journal of the American Academy of Child and Adolescent Psychiatry, 43,* 960–970.

Campbell, S. B. (1995). Behavior problems in preschool children: A review of recent research. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 36,* 113–149.

Canino, I., & Spurlock, J. (1994). *Culturally diverse children and adolescents: Assessment, diagnosis, and treatment.* New York: The Guilford Press.

Canivez, G. L., & Rains, J. D. (2002). Construct validity of the adjustment scales for children and adolescents and the preschool and kindergarten behavior scales: Convergent and divergent evidence. *Psychology in the Schools, 39,* 621–633.

Carrow-Woolfolk, E. (1995). *Oral and Written Language Scales (OWLS*). Circle Pines, MN: American Guidance Service.

Cauce, A. M. (1995). Of norms and cutoffs: Response. *Journal of the American Academy of Child and Adolescent Psychiatry, 34,* 537–538.

Clifford, R. M., Barbarin, O., Chang, F., Early, D. M., Bryant, D., Howes, C., et al. (2005). What is pre-kindergarten? Characteristics of public pre-kindergarten programs. *Applied Developmental Science, 9,* 126–143.

Conners, C. K. (1990). *Conners rating scales manual.* Toronto, Ontorio, Canada: Multi-Health Systems.

Costello, E. J. (1988). Child psychiatric disorders and their correlates: A primary care pediatric sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 28,* 851–855.

Costello, E. J., Angold, A., Burns, B. J., Erkanli, A., Stangl, D. K., & Tweed, D. L. (1996). The great smoky mountains study of youth. Functional impairment and serious emotional disturbance. *Archives of General Psychiatry, 53,* 1137–1143.

Dubrow, N. F., & Garbarino, J. (1989). Living in the war zone: Mothers and young children in a public housing development. *Child Welfare, 68,* 3–20.

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.) Circle Pines, MN: American Guidance Service.

Fallon, T. J., & Schwab-Stone, M. (1994). Determinants of reliability in psychiatric surveys of children aged 6–12. *Journal of Child Psychology & Psychiatry & Allied Disciplines, 35,* 1391–1408.

Fantuzzo, J., Bulotsky, R., McDermott, P., Mosca, S., & Lutz, M. N. (2003). A multivariate analysis of emotional and behavioral adjustment and preschool educational outcomes. *School Psychology Review, 32,* 185–203.

Fantuzzo, J., Coolahan, K., Mendez, J., McDermott, P., & Sutton-Smith, B. (1998). Contextually-relevant validation of peer play constructs with African American Head Start children: Penn Interactive Peer Play Scale. *Early Childhood Research Quarterly, 13,* 411–431.

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.

Fantuzzo, J., Stoltzfus, J., Lutz, M. N., Hamlet, H., Balraj, T., Turner, C., et al. (1999). An evaluation of the special needs referral process for low-income preschool children with emotional and behavioral problems. *Early Childhood Research Quarterly, 14,* 465–482.

Garber, J. (1984). Classification of childhood psychopathology: A developmental perspective. *Child Development, 55,* 30–48.

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

Harms, T., Clifford, R. M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale, Rev. ed.* New York: Teachers College Press.

Harrison, P. L. (1990). *AGS early screening profiles.* Circle Pines, MN: American Guidance Service.

Head Start Quality Research Consortium. (2003). *Head Start FACES: A whole-child perspective on program performance.* Washington DC: U.S. Department of Health and Human Services.

High/Scope Educational Research Foundation. (1992). *High/Scope Child Observation Record (COR) for ages 2 1/2*–6 [Assessment instrument]. Ypsilanti, MI: High/Scope Press.

Hightower, A. D., Work, W. C., Cowen, E. L., Lotyczewski, B. S., Spinell, A. P., Guare, J. C., et al. (1986). The Teacher-Child Rating Scale: A brief objective measure of elementary children's school problem behaviors and competencies. *School Psychology Review, 15,* 393–409.

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.

Keenan, K., & Wakschlag, L. S. (2000). More than the terrible twos: The nature and severity of behavior problems in clinic-referred pre-school children. *Journal of Abnormal Child Psychology, 28,* 33–46.

Kellam, S. G., Branch, J. D., Agrawal, K. C., & Ensminger, M. E. (1975). *Mental health and going to school: The Woodlawn Program of Assessment, early intervention and evaluation.* Chicago: University of Chicago Press.

Knitzer, J. (2000). Early childhood mental health services: A policy and systems development perspective. In J. Shonkoff & S. Meisels (Eds.) Handbook of early childhood intervention (2nd ed., pp. 416–438). New York: Cambridge University Press.

Knitzer, J., & Cooper, J. (2006). Beyond integration: Challenges for children's mental health. *Health Affairs, 25,* 670–679.

Kohn, M. (1988). *Kohn Problem Checklist/Kohn Social Competence Scale.* New York: Psychological Corporation.

Ladd, G. W., & Profilet., S. M. (1996). The child behavior scale: A teacher-report measure of young children's aggressive, withdrawn, and prosocial behaviors. *Developmental Psychology, 32,* 1008–1024.

Lutz, M. N., Fantuzzo, J., & McDermott, P. (2002). Multidimensional assessment of emotional and behavioral adjustment problems of low-income preschool children: Development and initial validation. *Early Childhood Research Quarterly, 17,* 338–355.

Mash, E. (2003). Child psychopathology: A developmental systems perspective. In E. J. Mash & R. A. Barkley (Eds.), *Child psychopathology* (2nd ed., pp. 3–71). New York: The Guilford Press.

McDermott, P. A., & Weiss, R. V. (1995). A normative typology of healthy, subclinical, and clinical behavior styles among American children and adolescents. *Psychological Assessment, 7,* 162–170.

Naglieri, J., LeBuffe, P., & Pfeiffer, S. (1994). *Devereux Scales of Mental Disorders (DSMD*). San Antonio, TX: The Psychological Corporation.

National Center for Early Development and Learning (NCEDL). (2000). *Multi-State study of public Pre-K.* Chapel Hill, NC: Frank Porter Graham Child Development Institute, University of North Carolina.

National Center for Education Statistics (NCES). (1993). *Public school kindergarten teachers' views on children's readiness for school.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Offord, D. R., Boyle, M. H., & Racine, Y. (1989). The Ontario Child Health Study: Correlates of disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 28,* 856–860.

Raadal, M., Milgrom, P., Cauce, A. M., & Manci, L. (1994). Behavior problems in 5-year-old to 11-year-old children from low-income fami-

lies. *Journal of the American Academy of Child and Adolescent Psychiatry, 33,* 1017–1025.

Raver, C. (2002). Emotions matter: Making the case for the role of young children's emotional development for early school readiness. *Social Policy Report: Giving Child & Youth Development Knowledge Away, 16,* Whole No. 3.

Rimm-Kauffman, S., Pianta, R., & Cox, M. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly, 15,* 147–166.

Rogler, L. H. (1993). Culture in psychiatric diagnosis: An issue of scientific accuracy. *Psychiatry-Interpersonal and Biological Processes, 56,* 324–327.

Rose, S. L., Rose, S. A., & Feldman, J. F. (1989). Stability of behavior problems in very young children. *Development and Psychopathology, 1,* 5–19.

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

U.S. Department of Health and Human Services. (1999). *Mental health: A report of the Surgeon General.* Rockville, MD: U.S. Department of Health and Human Services.

U.S. Department of Health and Human Services. (2001). *Report of the Surgeon General's conference on children's mental health: A national action agenda.* Washington, DC: U.S. Department of Health and Human Services.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement.* Itasca, IL: Riverside Publishing.

Zill, N. (1985). *The behavior problem scales.* Washington, DC: Child Trends, Inc.

Received August 25, 2005
Revision received May 15, 2007
Accepted June 6, 2007 ■