Screening for Social, Emotional, and Behavioral Problems at Kindergarten Entry: Utility and Incremental Validity of Parent Report

Julie Sarno Owens, Jennifer Storer, Alex S. Holdaway, Verenea J. Serrano, Yuko Watabe, Lina K. Himawan, Rebecca E. Krelko, and Katherine J. Vause
Ohio University

Erin Girio-Herrera
University of Miami

Nina Andrews
Logan Hocking School District

Abstract. The current study examined the utility and incremental validity of parent ratings on the Strengths and Difficulties Questionnaire and Disruptive Behavior Disorders rating scale completed at kindergarten registration in identifying risk status as defined by important criterion variables (teacher ratings, daily behavioral performance, and quarterly grades). The participants were 252 kindergarten students from one school district. Receiver operating characteristic analyses and area-under-the-curve values indicated that most subscales had low to moderate utility in identifying children showing at-risk academic performance and social, emotional, and behavioral problems. However, forward linear regression analyses indicated that parent ratings provided incremental validity relative to the academic screening tool used by the school district, accounting for an additional 3.69% to 22.37% of the variance in kindergarten outcomes. Implications for the use of parent ratings in universal screening for social, emotional, and behavioral problems at kindergarten entry are discussed.

Epidemiologic data indicate that 15% to 20% of children ages 6 to 17 years have men-tal health problems and few of these children receive adequate care (Kataoka, Zhang, &

Wells, 2002). Many children who are at risk of mental health problems manifest social (e.g., difficulty getting along with others), emotional (e.g., anxious, despondent), or behavioral (e.g., inattention, hyperactivity–impulsivity, aggression) problems (or a combination thereof) in early childhood (Pihlakoski et al., 2006; Wakschlag et al., 2007). Left unaddressed, social, emotional, and behavioral (SEB) problems place children at risk of school dropout, substance use, criminal behavior, and physical and mental health problems in adulthood and are costly for the education, healthcare, and justice systems (see National Research Council and Institute of Medicine, 2009, for review).

One way to prevent negative outcomes associated with SEB problems is to provide early intervention in schools or through community agencies. Indeed, several randomized trials document the benefits of systematic, high-quality education and intervention in early childhood for children's cognitive, language, social–emotional, and behavioral development (Conduct Problems Prevention Group & Dodge, 2007; Ramey & Ramey, 2004). Thus, there is a need for tools to identify children who may benefit from such early intervention. Although there is strong support for the utility of teacher ratings in screening for SEB problems (Feeney-Kettler, Kratochwill, Kaiser, Hemmeter, & Kettler, 2010; Harrison, Vannest, & Reynolds, 2013), teacher ratings are only available after the child has acclimated to the school environment and started to demonstrate failure in social, behavioral, or academic domains.

Identification of children with SEB problems should be done as early as possible to help children avoid early failure experiences and to maximize efficiency of intervention efforts. Kindergarten registration is an opportune time to systematically screen children for SEB problems via parent ratings. Studies examining the psychometric properties of parent-rated screening measures in young samples are emerging (Feeney-Kettler et al., 2010; Harrison et al., 2013). However, in most studies the criterion variable that is predicted is another rating scale rather than an

outcome that is important to parents, teachers, or administrators (e.g., grades, severe behavioral infractions) and on which educational decisions (e.g., grade retention, suspension) are made. Given that parent ratings could identify risk status prior to the child experiencing school failure, research is needed to determine if parent ratings at kindergarten registration predict indicators of school success that have high ecological validity.

Our university–community partnership examined (a) the utility of two publicly available parent rating scales that assess child SEB problems, completed at kindergarten registration, in identifying risk status on teacher ratings, as well as criterion variables of priority to school district personnel (i.e., grades, daily classroom behavior), and (b) the incremental validity of parent ratings in predicting those criterion variables beyond that predicted by the school district's academic screening tool used at kindergarten registration. To our knowledge, no studies have examined the incremental validity of parent ratings over other school-based tools used to detect risk.

SEB PROBLEMS AFFECT SCHOOL SUCCESS

There is substantial evidence that SEB problems can be reliably and accurately detected at an early age (Keenan & Wakschlag, 2004; Mathiesen & Sanson, 2000) and that such problems are associated with negative outcomes in multiple domains of functioning, including school performance (see Suldo, Gormley, DuPaul, & Anderson-Butcher, 2014, for review). For example, as early as preschool, inattention, hyperactivity, low-energy, social-reticence, and withdrawn behaviors observed by teachers at the beginning of the school year predict social and academic difficulties at the end of the year (Fantuzzo, Bulotsky, McDermott, Mosca, & Lutz, 2003). Teacher-rated social and self-regulation skills assessed in kindergarten are related to growth in academic skills through sixth grade. For example, children with lower levels of social and self-regulation skills start behind their peers academically in kindergarten and fall

further behind by second grade (McClelland, Acock, & Morrison, 2006). Furthermore, there is evidence that SEB problems persist into later childhood (Pihlakoski et al., 2006) and young adulthood (Kessler et al., 2005), affecting important school outcomes such as grade retention and graduation rates (Loe & Feldman, 2007). Together, these studies document significant negative academic outcomes associated with SEB problems in young children, in both the short and long term. Given the importance of school success to future employment and economic status (U.S. Department of Labor, 2013), identifying and addressing problems that interfere with academic success are important to students, families, schools, and society. However, children with SEB problems often go undetected in schools (Walker, Nishioka, Zeller, Severson, & Feil, 2000).

CURRENT PROCEDURES FOR IDENTIFYING SEB PROBLEMS IN STUDENTS

Several strategies can be used to identify children with SEB problems, including observational methods, teacher referral, and universally administered teacher ratings. Each method has advantages and disadvantages. Systematic direct observation of student behavior by an external observer is least susceptible to bias and captures the dynamic between student behavior and context (Landau & Swerdlik, 2005). However, observations require extensive personnel resources and multiple observations to obtain an adequate sample of behavior (Briesch, Chafouleas, & Riley-Tillman, 2010; Volpe, McConaughy, & Hintze, 2009).

Teachers are useful reporters of aberrant student behavior (e.g., Racz, King, Wu, Witkiewitz, & McMahon, 2013); thus, school districts may rely on teacher referrals to trigger assessment or service provision to students. However, individual teachers have different thresholds for behaviors warranting a referral for services, and a sizable minority may not refer children because they believe that it is not part of their role or they feel uncomfort-

able or unqualified to do so (Reinke, Stormont, Herman, Puri, & Goel, 2011). Thus, relying on teacher referral (rather than universally administered teacher ratings) can result in undetected and unaddressed risk in children. Indeed, studies show that 50% to 60% of children in first through sixth grade identified as at risk of SEB problems via universal teacher ratings were not referred for services by their teachers (Eklund & Dowdy, 2013; Eklund et al., 2009). Teacher referral has less predictive utility in identifying academic risk than universally completed teacher ratings (VanDer-Heyden, Witt, & Naquin, 2003). Together, these data support seeking alternative or complementary strategies beyond teacher referral to identify at-risk students.

ADVANTAGES AND DISADVANTAGES OF UNIVERSAL SCHOOL-BASED SCREENING

There are many benefits of universal school-based screening for SEB problems. First, universal screening (by either parent or teacher) ensures that all children are evaluated, reducing the likelihood of undetected risk. Second, screening results can identify target areas for early intervention and provide a baseline against which intervention efforts can be compared (at the child, classroom, or school level). Third, early detection, when connected with early intervention, interrupts the pathway between early risk factors and negative long-term outcomes (Conduct Problems Prevention Group & Dodge, 2007; Walker et al., 2009; Webster-Stratton, Rinaldi, & Reid, 2010). Thus, parent ratings obtained prior to school entry may be particularly important because teacher ratings are only available after the child has shown some level of social, behavioral, or academic failure. Furthermore, there is evidence that successful early intervention can produce up to a $2 million savings per youth treated (estimates take into account the costs of high school dropout and a career of adult criminal behavior and substance use, all of which are linked to juvenile antisocial behavior; Cohen, 1998). Although nearly 60% of children in the United

States attend preschool or receive center-based care (U.S. Census Bureau, 2013), systematic screenings with teacher report may not be feasible given the diversity in infrastructure across such settings.

Kindergarten registration represents a natural opportunity to systematically obtain ratings of child behavior and functioning from all parents. However, there are also challenges to using universal screening, particularly with young children. It can be difficult to identify risk status during a period in which variation in behavior is developmentally normative. False-positive cases could create unnecessary distress among parents and unnecessary financial expenditures for school districts, depending on follow-up actions taken. Similarly, screening tools have imperfect psychometric properties; thus, some students may still go undetected and unserved (false negatives). Lastly, screening may only be helpful to the extent that resources can meet the needs indicated by the screening results. Given the large number of unidentified and untreated youth on the one hand and the potential for possible negative outcomes on the other hand, research is needed to identify valid and reliable screening tools and to formulate data-driven guidelines for universal screening procedures.

PARENT-BASED SCREENING TOOLS

There is strong support for the utility of teacher ratings in screening for SEB problems in preschool and early elementary school students (Feeney-Kettler et al., 2010; Harrison et al., 2013). Teacher ratings in kindergarten are correlated with student grades (rs ranged from –.40 to –.20), test scores (rs ranged from –.30 to –.29), and discipline infractions (rs ranged from .37 to .57) 2 years later (DiStefano & Kamphaus, 2007). However, data supporting the predictive utility of parent-rated screening tools are limited. Several articles have summarized the psychometric properties of the parent versions of the Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007); the Ages and Stages Questionnaires: Social–Emotional (ASQ-SE; Squires, Bricker, & Twombly, 2003); the Pre-

school Behavior Screening System (PBSS; Feeney-Kettler, Kratochwill, & Kettler, 2011); and the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001). The reviews document that these measures have acceptable internal and test–retest reliability and concurrent validity (Feeney-Kettler et al., 2010; Harrison et al., 2013). Preliminary diagnostic utility statistics are also available for all measures. However, in most studies the criterion variable used to calculate the diagnostic utility statistics was another rating scale completed by the same rater (e.g., parent BESS with parent Behavior Assessment System for Children–Second Edition [BASC-2] from which the BESS was developed; Phase 1 of the parent PBSS with Phase 2 of the parent PBSS). This common-rater and common-method approach is an important context in which to interpret the data because both could contribute to inflated associations between measures (which are used as evidence for concurrent validity). In addition, these outcomes do not indicate whether parent measures predict ecologically valid indicators of risk (e.g., grades, discipline referrals).

Although psychologists prioritize rating scales and diagnoses, educators and parents often do not. Instead, parents and educators often judge school success by the child's grades and how well the child follows school rules and gets along with teachers and peers on a daily basis. Thus, in addition to using teacher rating scales as a criterion, we selected indicators of school success with high ecological validity (Suldo et al., 2014). Namely, student grades (i.e., grade point average [GPA]) and daily social and behavioral functioning according to the daily level achieved in the school-wide positive behavior support (SW-PBS) program (e.g., green, yellow, or red level) were considered critical outcomes. If research on universal screening is to make an impact on school policy, we need to know how well the screening tool predicts the indicators that are most important to consumers (i.e., parents, teachers, administrators) and that are used to make educational decisions. For example, it is unlikely that a child would be retained or suspended because of a score on a

rating scale. However, a child is very likely to be retained because of low grades or to be suspended because of an act of disrespect, defiance, or aggression, which would be reflected by the child's daily behavioral level achieved in an SW-PBS program.

CURRENT STUDY

To identify risk status prior to the onset of school failure experiences, we need to understand the utility of parent ratings, completed before the start of school, in identifying students at risk of academic and SEB problems (because teacher ratings are not available at this time). We sought to identify a free, publicly available, psychometrically sound parent-rated measure that could help school personnel screen for SEB problems at kindergarten registration and throughout the kindergarten year. Interestingly, most of the parent-rated screening measures in the literature did not fit the aforementioned characteristics (e.g., the BESS is not free; the PBSS is not publicly available; the ASQ-SE has an upper age limit that prohibits ongoing assessment). In this context, the team selected the SDQ because it is free, publicly available, and psychometrically sound and assesses a broad array of SEB problems, including peer problems, prosocial behavior, emotional problems, behavioral problems, and hyperactivity–inattention. As a comparison, we selected another parent rating scale that also met the aforementioned criteria but offered depth in the assessment of disruptive behavioral problems because these problems are predictive of future negative academic and social outcomes for children (e.g., Dodge, Greenberg, Malone, & Conduct Problems Prevention Research Group, 2008; Moilanen, Shaw, & Maxwell, 2010) and are associated with significant teacher stress, negative teacher–student interactions, and even workforce turnover (Greene, Beszterczey, Katzzenstein, Park, & Goring, 2002; Ingersoll, 2001). Namely, the Disruptive Behavior Disorders (DBD) rating scale (Pelham, Gnagy, Greenslade, & Milich, 1992) assesses the presence and frequency of the clinical symptoms of inattention, hyperactivity–impulsivity, and oppositional defiant behavior. Although there is overlap between the SDQ and DBD in the constructs assessed, it was important to determine the relative utility of depth versus breadth of assessment domains. The DBD has nine items for inattention and nine items for hyperactivity–impulsivity; the SDQ has five items on the hyperactivity–inattention subscale. If the shorter subscale of the SDQ is as predictive of outcomes as either longer subscale, then the measure with fewer items is more feasible to administer.

Our primary aims were to examine (a) the utility of two publicly available parent rating scales that assess child SEB problems in identifying risk status as defined by the teacher BESS ratings, quarterly grades, and daily behavioral performance in the SW-PBS program and (b) the incremental validity of parent ratings in predicting those criterion variables beyond that predicted by the existing school district academic screening tool. This study advances the literature by examining the utility and incremental validity of parent ratings, completed at kindergarten registration, in relation to multiple ecologically valid kindergarten outcomes.

METHOD

The parents of 273 kindergarteners (94% of all kindergarten students enrolled in the district), along with their teachers, consented to participate. Of the kindergarten students, 252 (92.3%) had sufficient rating scale data to be included in the analyses (21 had entire rating scales missing because families moved either prior to the start of school or before the end of Quarter 1). Those included in the analyses did not differ from those excluded regarding age, sex, race, school building assignment, mother's or father's highest level of education, current receipt of mental health services, or problem severity on the parent screening measures. See Table 1 for sample characteristics. Most respondents were mothers (85.7%), and the remainder were fathers (8.7%), grandmothers (3.6%), or other caregivers (2%). The data in Table 1 indicate

Table 1. Demographic Characteristics

Variable	n (%) or M (SD)
Sex (male)	127 (50.4%)
Race (White)	239 (94.8%)
Mother's highest education level	
No high school diploma	21 (8.3%)
High school diploma	75 (29.8%)
Partial college or associate degree	116 (46.1%)
Bachelor's degree	22 (8.7%)
Graduate degree	11 (4.4%)
Father's highest education level	
No high school diploma	23 (9.1%)
High school diploma	111 (44.0%)
Partial college or associate degree	72 (28.5%)
Bachelor's degree	14 (5.6%)
Graduate degree	3 (1.2%)
Attended preschool	193 (76.6%)
Age, years	4.87 (0.41)
SDQ subscales	
Emotional problems	1.42 (1.76)
Behavioral problems	1.37 (1.66)
Peer problems	1.18 (1.40)
Hyperactivity–inattention	3.53 (2.51)
Prosocial behavior	8.70 (1.94)
DBD	
Inattention	0.50 (0.54)
Hyperactivity–impulsivity	0.69 (0.57)
Oppositional defiant behavior	0.42 (0.48)
BESS total t score	46.13 (10.27)
KRA-L	19.57 (6.16)
% green days[a]	91.80 (13.59)
GPA	
Q1	3.53 (0.70)
Q4	3.67 (0.61)

Note. $N = 252$. For some categories, percentages do not sum to 100% because of missing data. BESS = Behavioral and Emotional Screening System; DBD = Disruptive Behavior Disorders Rating Scale; GPA = grade point average; KRA-L = Kindergarten Readiness Assessment–Literacy; Q = quarter; SDQ = Strengths and Difficulties Questionnaire.
[a]Percentage of green days across kindergarten year.

that the sample is representative of the communities in the Appalachian region of Ohio from which it was drawn (U.S. Census Bureau, 2006). All 12 teachers were female and White.

Measures

Parents completed a demographic questionnaire to document child sex, age, race or ethnicity; family moves in the past year; preschool experience and other socialization opportunities; and parental education and employment. Parents also completed two additional rating measures, as described in the following sections.

DBD Rating Scale

The DBD (Pelham et al., 1992) is a 45-item rating scale that assesses the presence of symptoms of inattention, hyperactivity–impulsivity, oppositional defiant behavior, and conduct disorder behavior, as listed in the fourth and fifth editions of the *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, 2000, 2013) for children in preschool through high school. Parents rated each item on a 4-point scale ranging from 0 (*not at all present*) to 3 (*very much present*). For this study, conduct disorder items were omitted owing to low base rates of these items in this age group. Internal consistency ($\alpha > .85$) and 4-week test–retest stability ($>.75$) were adequate for research, and construct validity of the inattention and hyperactivity–impulsivity scales has been demonstrated through significant correlations with similar subscales (DuPaul, Power, McGoey, Ikeda, & Anastopoulos, 1998). In the current sample, internal consistency estimates were .92 (inattention), .90 (hyperactivity–impulsivity), and .88 (oppositional defiant behavior). A subscale average score of 1 or greater has been used to denote at-risk status on this measure (Owens, Johannes, & Karpenko, 2009; Swanson et al., 2001). By use of this cut score, 16.7%, 25.4%, and 11.1% of students were showing at-risk levels of inattention, hyperactivity–impulsivity, and oppositional defiant behavior, respectively.

Strengths and Difficulties Questionnaire

The SDQ (Goodman, 2001) is a 25-item questionnaire in which items are rated on a 3-point scale ranging from 0 (*not true*) to 2 (*certainly true*). We used the parent version for children ages 4 to 10 years. Four studies

(see Stone, Otten, Engels, Vermulst, & Janssen, 2010, for review) confirm the five-sub-scale factor structure: emotional problems, behavioral problems, peer problems, hyperactivity–inattention, and prosocial behavior. Across 26 studies, the weighted average internal consistency of the parent subscales ranged from .53 (peer problems) to .76 (hyperactivity–inattention) and the stability over 4 to 6 months ranged from .65 to .71. High scores on the parent SDQ are also associated with increased risk of a psychiatric disorder (Goodman, 2001; Stone et al., 2010). Higher scores on the prosocial behavior subscale indicate more acceptable behaviors; higher scores on all other subscales indicate more problematic behaviors. In our sample, internal consistency estimates were .70 (emotional problems), .70 (behavioral problems), .83 (hyperactivity–inattention), .48 (peer problems), and .72 (prosocial behavior). Using the recommended cut scores (see www.sdqinfo.org), 16.7% (peer problems), 17.9% (hyperactivity–inattention), 18.3% (emotional problems), 20.2% (behavioral problems), and 20.6% (prosocial behavior) of students were at risk depending on the subscale.

BASC-2 BESS–Teacher Version

The BESS (K–12 form) is a 26-item measure that screens for internalizing, externalizing, and adaptive behavioral problems in children (Kamphaus & Reynolds, 2007). Data from a large normative sample show that the teacher version total scores have acceptable interrater ($\geq$.70) and internal reliability ($\geq$.94) for research, as well as evidence of convergent and predictive validity for standardized achievement test scores and student GPAs over 4 years (rs ranged from .34 to .63). There is also evidence of strong test–retest reliability ($>$.85, with approximately 40-day interval; Feeney-Kettler et al., 2010). In this study, BESS total scores were used as a criterion variable. By use of a cut score for age-based t score of 61 or higher, 10% of students in this sample were at risk.

Grade Point Average

A quarterly GPA was calculated using a 4-point scale (based on participants' grades in reading, spelling, math, science, and social studies) and used as a criterion variable. Student grades were classified as follows: above satisfactory (4.0), satisfactory (3.0), below satisfactory (2.0), needs improvement (1.0), and unsatisfactory (0.0). Given that a GPA lower than 3.0 was considered below satisfactory, a cut score of less than 3.0 was used as the threshold to define at-risk status for academic problems. Across quarters, 8% to 18% of students fell below this cut score.

Behavioral Functioning Indicator

All schools in the participating district used an SW-PBS framework. Each teacher used a behavior wheel that documented the students' daily rule-following behavior. Each student began the school day with his or her clip on the green segment of the wheel and moved his or her clip with each additional rule violation, such that the yellow segment represented a first warning, the orange segment represented a second warning (with a possible mild consequence), and the red segment represented a referral to the office or parent notification (or both). This daily status was documented on the student's monthly calendar. This measure has strong ecological validity and is a metric commonly used by school staff and parents to describe a child's day. The percentage of green days achieved for each student was calculated by dividing the total number of green days achieved between August and June by the total number of days the student attended school in that time. In collaboration with teachers, at-risk status was defined as achieving fewer than 80% green days. Teachers reported that a typical student could be assigned a level lower than green ("off green") one day per week but that two days off green raised concerns about behavioral control and was often associated with other problems. In the sample, 12% of students fell below this cut score (sample range was 15%–100% green days; 75% of students had $\geq$90% green days).

Kindergarten Readiness Assessment–Literacy

The Kindergarten Readiness Assessment–Literacy (KRA-L) is a 25-item screen-

ing tool developed by the Ohio Department of Education to assess early literacy skills in three areas: oral language, phonological awareness, and print awareness. Total scores, ranging from 0 to 29, are organized into three bands, each with a recommendation to guide decisions about the need for additional assessment and instructional programming: 0 to 13 (assess broadly for intense instruction), 14 to 23 (assess for additional instruction in targeted areas), and 24 to 29 (assess for enriched or advanced instruction). Across the school district, the average score was 19.49 and 22.42% of students fell in the lowest band, 46.98% fell in the middle band, and 30.60% fell in the highest band. In the current sample, the average score was 19.57 and 21.2% of students fell in the lowest band, 47.6% fell in the middle band, and 31.2% fell in the highest band. Although the KRA-L has been required by the state since 2007, few psychometric data are available (http://education.ohio.gov/Topics/Early-Learning/Guidance-About-Kindergarten/KRAL).

Procedures

Parents were recruited to participate in the project when they registered their child for kindergarten (between April and August 2011). Once parents provided consent, they completed the demographic questionnaire, DBD, and SDQ. Parents received a small educational gift as compensation for their participation. The majority of parents completed study questionnaires in April and May (78%), whereas the remaining parents completed study measures in June (1%), July (1%), August (16%), and September (4%). The child's KRA-L scores, as well as the summary of the child's screening, were inserted into the child's kindergarten registration file for the principal and teacher to use as desired. Teachers provided consent to participate and completed the teacher BESS for all consented children in October. This time frame was chosen to allow adequate time for teachers to become familiar with their students. Teachers were

compensated with $80. Indicators of academic achievement and school behavior were collected on a regular basis throughout the academic year (August through June).

RESULTS

Descriptive statistics for all predictor and criterion variables are presented in Table 1. The data were used to examine the utility of the parent rating scales identifying risk status, as well as the incremental validity of parent ratings in predicting criterion variables beyond that predicted by the existing school district academic screening tool.

Utility in Identifying Risk Status

Children were first categorized as at risk or typically developing for each criterion variable (i.e., BESS teacher rating scale, percent green days, GPA for Quarters 1 and 4). Receiver operating characteristic (ROC) analyses were used to assess the utility of each subscale of the DBD and SDQ in identifying student risk status on each criterion. ROC analyses produce an area-under-the-curve (AUC) value that represents a ratio of sensitivity and specificity for identifying at-risk status. AUC values above .5 indicate that the parent subscale score predicts at-risk status on the indicator at a rate greater than chance (see Table 2). There is variability in the qualitative terms used to describe the magnitude of AUCs, without definitive agreement on these terms (e.g., Pintea & Moldovan, 2009; Swets, 1996). Thus, we describe the results using the following ranges: .50 to .70, low to moderate; .70 to .90, moderate to strong; and .90 and above, strong.

Identifying Risk Status Based on Teacher Ratings

In identifying risk status on the BESS, the AUCs for the SDQ and DBD subscales ranged from .50 to .68 (see Table 2). With the exception of the SDQ emotional problems subscale (AUC = .50), confidence intervals (CIs) indicated that all subscales were similar in identifying risk status on the BESS.

Table 2. AUC Statistics for Parent Ratings Predicting Teacher Ratings, Behavioral Functioning, and GPA

Subscale	BESS[a]			% Green Days[bc]			Q1 GPA[d]			Q4 GPA[e]		
	AUC	SE	95% CI	AUC	SE	95% CI	AUC	SE	95% CI	AUC	SE	95% CI
SDQ												
Emotional problems	.50	.06	[.37, .62]	.45	.06	[.34, .56]	.57	.05	[.47, .67]	.56	.06	[.44, .69]
Behavioral problems	.68	.06	[.55, .80]	.69	.05	[.59, .80]	.62	.05	[.52, .72]	.60	.06	[.48, .72]
Peer problems	.66	.06	[.55, .77]	.66	.05	[.56, .77]	.65	.05	[.56, .74]	.64	.06	[.52, .75]
Hyperactivity–inattention	.68	.06	[.57, .79]	.67	.06	[.56, .78]	.66	.05	[.57, .75]	.63	.05	[.52, .73]
Prosocial behavior	.61	.06	[.50, .72]	.71	.05	[.61, .81]	.66	.05	[.57, .75]	.57	.06	[.45, .69]
DBD												
Inattention	.66	.06	[.53, .78]	.60	.06	[.47, .72]	.69	.05	[.59, .78]	.67	.05	[.56, .78]
Hyperactivity–impulsivity	.68	.07	[.55, .80]	.67	.06	[.55, .79]	.65	.05	[.55, .74]	.63	.06	[.51, .75]
Oppositional defiant behavior	.63	.06	[.51, .75]	.62	.06	[.50, .73]	.66	.05	[.56, .75]	.59	.06	[.46, .71]

Note. AUC = area under the curve; BESS = Behavioral and Emotional Screening System; CI = confidence interval; DBD = Disruptive Behavior Disorders Rating Scale; GPA = grade point average; Q = quarter; SDQ = Strengths and Difficulties Questionnaire. [a]n = 26 at risk (N = 252 total). [b]n = 31 at risk (N = 252 total). [c]Percentage of green days across kindergarten year. [d]n = 44 at risk (n = 248 total). [e]n = 29 at risk (n = 235 total).

Identifying Risk Status Based on Daily Behavior

In identifying risk status based on SW-PBS green days, the AUCs for the SDQ and DBD subscales ranged from .45 to .71 (see Table 2). Consistent with results for the BESS, with the exception of the SDQ emotional problems subscale (AUC = .45), CIs indicated that all subscales were similar in identifying risk status on this criterion variable.

Identifying Risk Status Based on Grades

The AUCs for first-quarter and fourth-quarter grades ranged from .56 to .69. (The AUCs for the second- and third-quarter grades all fell within the CIs of those for the first and fourth quarters and are not presented because of space limitations.) Across both quarters, the DBD inattention subscale had the highest AUC statistic for identifying risk status based on GPA and the SDQ emotional problems subscale had the lowest.

Incremental Validity

To better understand which subscales best predicted the criterion variables, relative to each other and beyond the predictive utility of the school district's academic screening tool (KRA-L), a forward linear regression analysis was conducted for each outcome indicator. To account for the nesting of students within teachers, a regression model with random intercept was applied. Random effects (resulting from students nested within teachers) were not applied to the predictors because they were collected before the start of the academic year; thus it was reasonable to assume there was no effect of teacher on them. A manual forward analysis was conducted as follows. In the first step, the KRA-L total was entered. In the second step, the best predictor (among the three DBD subscales and five SDQ subscales) was chosen and entered based on the one that provided the best improvement in three information criteria: Akaike information criterion, corrected Akaike information criterion, and Bayesian information criterion. The second step was repeated until either no improvement in all three information criteria

could be made or no significant predictors could be added. See Table 3 for correlations between predictor and criterion variables and Table 4 for regression results.

Predicting Teacher Rating Scores

In the first step, the KRA-L accounted for 20.82% of the variation in the teacher BESS scores. In the second step, the SDQ behavioral problems subscale offered the most improvement in prediction based on the information criteria, accounting for an additional 10.40% of the variation in BESS scores. In the third, fourth, and fifth steps, the DBD hyperactivity–impulsivity (2.02%), SDQ emotional problems (2.01%), and DBD oppositional defiant behavior (1.36%) subscales each accounted for additional variance in BESS scores.

Predicting Daily Behavior

In the first step, the KRA-L accounted for 6.18% of the variation in green days. On the basis of the information criteria, the SDQ behavioral problems subscale offered the most improvement in the second step, accounting for an additional 13.99% of variation in green days. In the third, fourth, and fifth steps, the DBD hyperactivity–impulsivity (2.55%), SDQ emotional problems (4.15%), and DBD oppositional defiant behavior (1.68%) subscales each produced significant improvement in prediction.

Predicting Grades

In the first step, the KRA-L accounted for 40.50% and 23.36% of the variation in the first- and fourth-quarter GPA, respectively. For the first-quarter GPA, in the second and third steps, the SDQ peer problems (3.31%) and prosocial behavior (1.63%) subscales offered significant improvement in prediction based on the information criteria. For the fourth-quarter GPA, only the SDQ peer problems subscale (3.69%) accounted for additional variance.

Determining Possible Cut Scores

Once the subscales with significant incremental validity were identified, diagnostic

Table 3. Correlations Among Predictor and Criterion Variables

Domain	1	2	3	4	5	6	7	8	9	10	11	12	13
1. DBD inattention	—												
2. DBD hyperactivity–inattention	.86*	—											
3. DBD oppositional defiant behavior	.71*	.71*	—										
4. SDQ emotional problems	.46*	.42*	.48*	—									
5. SDQ behavioral problems	.62*	.61*	.74*	.40*	—								
6. SDQ peer problems	.44*	.41*	.40*	.35*	.44*	—							
7. SDQ hyperactivity–inattention	.71*	.75*	.52*	.32*	.59*	.43*	—						
8. SDQ prosocial behavior	−.38*	−.37*	−.45*	−.14*	−.52*	−.28*	−.29*	—					
9. BESS total	.32*	.37*	.27*	.05	.35*	.23*	.35*	.34*	—				
10. % green days[a]	−.26*	−.38*	−.25*	.02	−.36*	−.24*	−.34*	−.32*	−.61*	—			
11. Q1 GPA	−.29*	−.24*	−.25*	−.09	−.22*	−.27*	−.26*	−.28*	−.58*	.32*	—		
12. Q4 GPA	−.22*	−.24*	−.23*	−.09	−.17*	−.24*	−.23*	−.24*	−.47*	.31*	.61*	—	
13. KRA-L	−.25*	−.23*	−.26*	−.11	−.20*	−.17*	−.24*	−.19*	−.42*	.26*	.61*	.42*	—

Note. BESS = Behavioral and Emotional Screening System; DBD = Disruptive Behavior Disorder; GPA = grade point average; KRA-L = Kindergarten Readiness Assessment–Literacy; Q = quarter; SDQ = Strengths and Difficulties Questionnaire.
[a]Percentage of green days across kindergarten year.
*$p < .01$.

Table 4. Random Coefficient Regression Results for Parent Ratings Predicting Teacher Ratings, Behavior, and GPA

	ΔR^2	β	SE
BESS at-risk status			
Step 1: KRA-L total	20.82%	−0.618*	0.084
Step 2: SDQ behavioral problems	10.40%	1.864*	0.434
Step 3: DBD hyperactivity–impulsivity	2.02%	5.148*	1.251
Step 4: SDQ emotional problems	2.01%	−0.741	0.320
Step 5: DBD oppositional defiant behavior	1.36%	−3.905	1.755
Additional R^2 because of Step 2 to 5	15.79%		
Total R^2	36.61%		
80% green for behavior			
Step 1: KRA-L total	6.18%	0.004*	0.001
Step 2: SDQ behavioral problems	13.99%	−0.025*	0.005
Step 3: DBD hyperactivity–impulsivity	2.55%	−0.110*	0.025
Step 4: SDQ emotional problems	4.15%	0.016*	0.005
Step 5: DBD inattention	1.68%	0.065	0.026
Additional R^2 because of Step 2 to 5	22.37%		
Total R^2	28.55%		
Q1 GPA			
Step 1: KRA-L total	40.50%	0.066*	0.005
Step 2: SDQ peer problems	3.31%	−0.071*	0.024
Step 3: SDQ prosocial behavior	1.63%	0.045	0.018
Additional R^2 because of Step 2 to 3	4.94%		
Total R^2	45.44%		
Q4 GPA			
Step 1: KRA-L total	23.36%	0.042*	0.006
Step 2: SDQ peer problems	3.69%	−0.082*	0.025
Additional R^2 because of Step 2	3.69%		
Total R^2	27.05%		

Note. Depending on variables in model, *n* ranges from 231 (model with Q4 GPA) to 248 (80% green). DBD = Disruptive Behavior Disorders Rating Scale; GPA = grade point average; KRA-L = Kindergarten Readiness Assessment–Literacy; Q = quarter; SDQ = Strengths and Difficulties Questionnaire.
*$p < .01$.

utility statistics (sensitivity, specificity, positive predictive power, negative predictive power) were examined to determine cut scores that balanced false positive to false negative ratios. Although sensitivity, specificity, and positive predictive power rates exceeding .75 are considered ideal (Glover & Albers, 2007), other authors have noted that rates between .40 and .60 may be acceptable in the first stage of a screening protocol (Kettler & Feeney-Kettler, 2011). The cut scores presented in Table 5 are those that represent the best balance of sensitivity and specificity, assuming a base rate of 20%, consistent with epidemiologic data (Kataoka et al., 2002).

DISCUSSION

This study provides preliminary evidence for the utility and incremental validity of score inferences derived from parent SDQ and DBD rating scales, completed at kindergarten registration, in relation to important kindergarten outcomes (i.e., teacher ratings, grades, and daily behavior reports). In brief, the ROC analyses and the diagnostic effi-

Table 5. Diagnostic Efficiency Statistics for Parent-Rated Subscales Predicting Criterion Variables

	SEN	SPE	EFF	PPV	NPV	TP	FP	TN	FN
SDQ behavioral problems subscale identifying at-risk status on BESS									
Cut score of 2	.58	.68	.67	.31	.87	6.0%	28.6%	61.1%	4.4%
Cut score of 3	.46	.83	.79	.40	.86	4.8%	15.5%	74.2%	5.6%
Cut score of 4	.31	.93	.87	.52	.84	3.2%	6.3%	83.3%	7.1%
DBD hyperactivity–impulsivity subscale identifying at-risk status on BESS									
Cut score of 0.5	.73	.49	.51	.26	.88	7.5%	46.0%	43.7%	2.8%
Cut score of 1	.54	.79	.77	.39	.87	5.6%	18.7%	71.0%	4.8%
DBD oppositional defiant behavior subscale identifying at-risk status on BESS									
Cut score of 0.5	.50	.67	.65	.27	.84	5.2%	29.8%	60.0%	5.2%
Cut score of 1	.31	.91	.85	.47	.84	3.2%	7.9%	81.7%	7.1%
SDQ emotional problems subscale identifying at-risk status on BESS									
Cut score of 1	.58	.39	.41	.19	.79	6.0%	54.8%	34.9%	4.4%
Cut score of 2	.35	.65	.62	.20	.80	2.6%	31.3%	58.3%	6.7%
Cut score of 3	.23	.82	.76	.25	.81	2.4%	15.9%	73.8%	7.9%
SDQ behavioral problems subscale identifying at-risk status on green days									
Cut score of 2	.58	.69	.68	.32	.87	7.1%	27.4%	60.3%	5.2%
Cut score of 3	.45	.83	.79	.40	.86	5.6%	14.7%	73.0%	6.7%
Cut score of 4	.23	.92	.84	.42	.83	2.8%	6.7%	81.0%	9.5%
DBD hyperactivity–impulsivity subscale identifying at-risk status on green days									
Cut score of 0.5	.74	.51	.54	.28	.89	9.1%	42.9%	44.8%	3.2%
Cut score of 1	.48	.79	.75	.37	.86	6.0%	18.3%	69.4%	6.3%
DBD inattention subscale identifying at-risk status on green days									
Cut score of 0.5	.52	.63	.62	.26	.84	6.3%	32.5%	55.2%	6.0%
Cut score of 1	.32	.86	.79	.36	.84	4.0%	12.7%	75.0%	8.3%
SDQ emotional problems subscale identifying at-risk status on green days									
Cut score of 1	.52	.38	.40	.17	.76	6.3%	54.4%	33.3%	6.0%
Cut score of 2	.29	.64	.60	.17	.78	3.6%	31.3%	56.3%	8.7%
Cut score of 3	.16	.81	.73	.18	.80	2.0%	16.3%	71.4%	10.3%
SDQ peer problems subscale identifying at-risk status on Q1 GPA									
Cut score of 1	.75	.48	.58	.27	.89	12.3%	42.7%	39.5%	4.4%
Cut score of 2	.50	.71	.67	.30	.85	8.9%	23.8%	58.5%	8.9%
Cut score of 3	.30	.86	.76	.34	.83	5.2%	11.7%	70.6%	12.5%
SDQ prosocial behavior subscale[a] identifying at-risk status on Q1 GPA									
Cut score of 7	.35	.83	.77	.22	.90	4.3%	14.9%	72.8%	8.1%
Cut score of 6	.21	.88	.79	.19	.89	2.6%	10.6%	77.0%	9.8%
Cut score of 5	.10	.95	.85	.23	.88	1.3%	4.3%	83.4%	11.1%
SDQ peer problems subscale identifying at-risk status on Q4 GPA									
Cut score of 1	.72	.46	.49	.25	.87	8.9%	47.7%	40.0%	3.4%
Cut score of 2	.52	.30	.68	.30	.85	6.4%	26.0%	61.7%	6.0%
Cut score of 3	.31	.14	.79	.36	.83	3.8%	12.3%	75.3%	8.5%

Note. BESS = Behavioral and Emotional Screening System; DBD = Disruptive Behavior Disorder; EFF = diagnostic efficiency; FN = false negatives; FP = false positives; GPA = grade point average; NPV = negative predictive value; PPV = positive predictive value; Q = quarter; SDQ = Strengths and Difficulties Questionnaire; SEN = sensitivity; SPE = specificity; TN = true negatives; TP = true positives.
[a]For the prosocial behavior subscale, lower scores are indicative of greater problems in this area.

ciency statistics indicated that most parent-rated subscales offered low to moderate utility in identifying children demonstrating at-risk academic performance in the first and last quarter of the year (i.e., GPA) and SEB problems in the first quarter (teacher BESS ratings) and throughout the year (percent green days). However, parent ratings of child SEB problems provided incremental validity relative to the academic screening tool used by the school district, accounting for an additional 3% to 22% of the variance in kindergarten outcomes. These data highlight (a) the potential utility of obtaining parent report prior to kindergarten entry to identify children who are likely to struggle to achieve grade-level expectations for academic, social, and behavioral functioning and (b) the importance of incorporating screening results into a more comprehensive, multi-informant, and multistep process that is linked to a service-delivery framework. In the following sections, we discuss our findings in relation to previous studies with samples of young children and offer recommendations for research and practice.

Identifying and Predicting SEB Functioning

We used the teacher BESS as an indicator of SEB functioning and SW-PBS green days as an indicator of behavioral functioning. The strongest predictor of these criterion variables was SDQ behavioral problems, accounting for 10% to 14% of variance in these outcomes beyond that accounted for by the KRA-L. Furthermore, the depth of items on the DBD hyperactivity–impulsivity and oppositional defiant behavior subscales predicted additional variance in BESS scores, and the depth of the DBD inattention subscales predicted additional variance in green days. The SDQ emotional problems subscale also explained a small incremental proportion on both outcomes. Collectively, these results are consistent with evidence that SEB problems, detected in young children at the beginning of the year, are associated with peer rejection, student–teacher conflict, and impaired class-

room functioning at the end of the year (Fantuzzo et al., 2003; Ladd, Birch, & Buhs, 1999).

The ROC analyses provide important information about dichotomous decisions that school personnel must make using the screening results. The AUCs identifying at-risk status on the teacher BESS (e.g., behavioral problems = .68; 95% CI [.55, .80]) and green days (e.g., prosocial behavior = .71; 95% CI [.61, .81]) were low to moderate yet comparable to findings from other studies that have examined utility statistics in a cross-informant or cross-method approach. For example, during the process of developing the teacher BESS, DiStefano and Kamphaus (2007) reported that teacher ratings on the 23-item screening tool identified children with a clinical diagnosis (compared with children without a diagnosis) at a rate greater than chance; the AUC was .70 (95% CI [.64, .77]). Other studies have found higher AUCs, ranging between .70 and .80, for both parent and teacher SDQs identifying children with clinical diagnoses (e.g., Stone et al., 2010). This study advances the literature by documenting this relationship with more distal and ecologically valid indicators of risk (i.e., diagnoses are often made concurrently and have considerable conceptual overlap with screening measures).

The diagnostic utility statistics document the trade-offs that have to be made regarding false-positive and false-negative rates at different cut scores. On the one hand, the negative predictive values for most subscales are above .80, raising confidence that if the child's screening results are negative, he or she is highly unlikely to show at-risk behavior during kindergarten. On the other hand, the sensitivity and positive predictive values indicate that it is premature to recommend that parent ratings on these measures be used in isolation or as a first gate in multiple-gating procedures because the false-positive rates (i.e., inaccurately telling parents that their kindergarten child is at risk) may create undue stress for parents and teachers alike and the false-negative rates may contribute to undetected academic risk. Future research that follows children longitudinally across multiple

years could better inform the manner in which parent ratings can be incorporated into school district screening, monitoring, and assessment procedures.

This preliminary evidence that parent ratings completed at kindergarten entry can predict teacher perceptions and daily behavioral functioning across the entire year represents a significant contribution to the literature. Given that disruptive classroom behavior is (a) stressful for teachers (Greene et al., 2002), (b) one of the most common reasons that teachers refer children for services (Lloyd, Kauffman, Landrum, & Roe, 1991), and (c) a significant contributor to teachers leaving the profession (Ingersoll, 2001), being able to identify such risk prior to school entry provides the opportunity to better target children for monitoring, further assessment, or early intervention.

Identifying and Predicting Academic Functioning

We used quarterly GPA as the indicator of academic functioning. Because the KRA-L is designed to assess early literacy skills, it is not surprising that KRA-L scores accounted for a large proportion of the variance in the Quarter 1 GPA (41%) and a much lower portion of the variance in the Quarter 4 GPA (23%). Interestingly, the SDQ peer problems subscale was the strongest significant predictor of student GPA at the beginning of the year and end of the year, accounting for a small but perhaps meaningful portion of incremental variance. Unlike the KRA-L, the proportion of variance explained by the SDQ peer problems subscale remained fairly stable across the year. This pattern underscores the importance of attending to social functioning when attempting to understand and influence academic functioning. This finding is consistent with other studies that have documented the positive relationship between peer problems and academic problems (DiPerna, Volpe, & Elliott, 2001), as well as studies documenting the independent and additive contributions of peer rejection and diagnostic status on academic impairment (Mikami & Hinshaw,

2006). Although a better understanding of this cross-domain relationship is important, the diagnostic utility statistics indicate that parent ratings of peer relations have limited sensitivity and positive predictive power. Thus, additional research is needed to examine other cut scores to define at-risk academic performance and other indicators of academic functioning in this age group. In addition, it could be fruitful to explore the sensitivity and specificity of other measures of parent-rated social functioning at this age. Our study findings suggest that such research has merit.

Conducting Universal Screening at Kindergarten Registration

Although nearly 60% of children in the United States attend preschool or receive center-based care (U.S. Census Bureau, 2013), systematic screenings prior to kindergarten may not be feasible given the diversity of infrastructure across such preschool settings. The structure of kindergarten registration represents an opportune context in which to obtain parent-based screening information that enhances school professionals' preparedness to meet the unique needs of every child. However, detecting at-risk status in a young child is a sensitive issue. Arguments cautioning against early mental health screening include the wide variability in normative behavior at this developmental stage, the risk of false positives, the burden placed on schools to address the needs of children detected, and the potential financial and personnel costs. Given the low base rate of the behaviors to be detected, the sensitivity and positive predictive power rates for screening tools identifying cross-informant or cross-method outcomes may be limited. Although these concerns warrant careful attention, the studies reviewed in the introduction offer important evidence that the advantages of early intervention may outweigh the disadvantages associated with the aforementioned concerns. Furthermore, regarding concern about false positives, untoward outcomes that may arise as a function of a false-positive identification can be mitigated by how detected cases are handled. For exam-

ple, researchers and school districts could take different actions based on different levels of screening scores. Exceeding the lowest cut score could simply trigger closer monitoring of the student and parent communication, whereas exceeding a higher cut score could trigger a cross-informant assessment or the provision of early intervention in the general education environment.

It is important to recognize that universal screening does not produce additional need; rather, screening simply identifies students who might have unmet need. Although some investigators may argue that there is no harm in waiting a few months to obtain teacher report, the results of our study suggest that universal screening identifies many of those children who are likely to emerge as needing assistance (e.g., through current procedures such as lower grades and referrals to the office) but identifies them earlier, offering the opportunity for parent–teacher communication, the development of monitoring systems, and early intervention. Furthermore, given the moderate correlations between parent and teacher ratings, parent screening identifies a unique subset of children who may not be detected by teacher ratings. Lastly, there is some evidence that children who are detected via a teacher-rated universal screener, but are deemed to be false positives because they did not qualify for special education, are still impaired and in need of intervention because of lower scores on measures of intelligence, language, and academic achievement as compared with children who did not fail the screening test (true negatives; Glascoe, 2001).

Limitations

Although the current data are of interest to researchers and practitioners, they should be considered within their limitations. First, the teacher sample is homogeneous regarding race and the child sample is homogeneous regarding race and socioeconomic status. Thus, replication with larger samples and more racially and economically diverse samples is warranted. Second, because teachers had access to the results of the parent screen-

ing, such access could have influenced teacher ratings or services provided during the year. However, because the screening was new to the district and systematic procedures for how to use the screening results had not been imposed, this influence was minimal. Indeed, informal interviews with school staff suggested that the reports were used minimally and inconsistently across teachers and school buildings. This highlights the importance of developing and evaluating systematic procedures and accountabilities for using screening data. Third, because quarterly grades in kindergarten are typically not based on quizzes, tests, and homework assignments, GPA in kindergarten may simply represent another rating of teacher's perception (and possibly bias) of student progress toward the academic standards. Fourth, because psychometric data are not available for the KRA-L, the incremental validity for parent ratings may differ when using a different academic screening tool. Lastly, longitudinal data are needed to understand the longer-term outcomes of identified children.

Preliminary Recommendations for Practice

Despite the aforementioned limitations, the results suggest that parent report at kindergarten entry offers significant, unique, and meaningful data about child risk status beyond that offered by the school district's academic screening and thus is worthy of inclusion in future research and practice. Given the consistency of our findings with the empirical literature, as well as the consistency of the AUCs across multiple domains of functioning, across multiple methods of assessment (teacher rating scales, grades, and daily behavior), and across time, preliminary recommendations can be made. First, the data support the use of either the DBD or the SDQ in future research. Given that the SDQ behavioral problems subscale was the strongest predictor of teacher BESS scores and SW-PBS green days and that the SDQ offers information about social functioning (i.e., peer problems and prosocial behavior) that is not offered by the DBD, it may

be advantageous to use the SDQ. A similar recommendation has been made for teacher screening tools because broader screening tools (assessing more domains) are more useful than narrower tools that assess one domain in predicting later teacher perceptions of the student, peer relations, and grades (Flanagan, Bierman, Kam, & Conduct Problems Prevention Research Group, 2003).

Second, these data suggest that school personnel should consider how to include parent ratings in their screening procedures. Identification on such ratings could trigger a monitoring procedure in which identified children are reviewed and discussed at intervention assistance team meetings, as well as a process for proactive communication with parents before problems further develop. Then, additional screening, assessment, and intervention services could be titrated over time to address problems that persist. Such a process could lead to increased efficiency in identification and facilitate the provision of early intervention to reduce the cost and intensity of interventions needed later.

Lastly, parent rating scales completed at kindergarten registration provide a unique opportunity to obtain a "preview" of incoming students. This information could be used to distribute children with varying strengths and weaknesses evenly across classrooms or to promote better student–teacher fit. Using screening data, school staff may be able to prevent the development of some negative teacher–student relationships or to prevent one teacher from being assigned the majority of the most demanding students by chance. Future studies could also examine the extent to which children identified as at risk in the spring prior to kindergarten would benefit from some type of school readiness programming during the summer prior to kindergarten entry.

Summary

Obtaining parent ratings during kindergarten registration gives school staff a head start of several months to assess individual student needs and make decisions related to monitoring and intervention. Although the AUC rates indicate that the parent ratings on the DBD and SDQ could not be used as a screening tool in isolation, they do offer incremental information on social, behavioral, and academic functioning during the kindergarten year, accounting for up to 22% of additional variance beyond that explained by the academic screener used by the school district. These data provide an important foundation for future research examining the utility of parent ratings in multistep, multi-informant screening and intervention processes.

REFERENCES

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Briesch, A. M., Chafouleas, S. M., & Riley-Tillman, T. C. (2010). Generalizability and dependability of behavioral assessment methods: A comparison of systematic direct observation and direct behavior rating. *School Psychology Review, 3,* 408–421.

Cohen, M. A. (1998). The monetary value of saving a high-risk youth. *Journal of Quantitative Criminology, 14,* 5–33. doi:10.1177/10634266070150020401

Conduct Problems Prevention Group & Dodge, K. A. (2007). Fast track randomized controlled trial to prevent externalizing psychiatric disorders: Findings from grades 3 to 9. *Journal of the American Academy of Child and Adolescent Psychiatry, 46,* 1250–1262. doi:10.1097/chi.0b013e31813e5d39

DiPerna, J. C., Volpe, R. J., & Elliott, S. N. (2001). A model of academic enablers and elementary reading/language arts achievement. *School Psychology Review, 3,* 298–312.

DiStefano, C. A., & Kamphaus, R. (2007). Development and validation of a behavioral screener for preschool-age children. *Journal of Emotional and Behavioral Disorders, 15,* 93–102.

Dodge, K. A., Greenberg, M. T., Malone, P. S., & Conduct Problems Prevention Research Group. (2008). Testing an idealized dynamic cascade model of the development of serious violence in adolescence. *Child Development, 79,* 1907–1927.

DuPaul, G. J., Power, T. J., McGoey, K. E., Ikeda, M. J., & Anastopoulos, A. D. (1998). Reliability and validity of parent and teacher ratings of attention-deficit/hyperactivity disorder symptoms. *Journal of Psychoeducational Assessment, 16,* 55–68. doi:10.1177/073428299801600104

Eklund, K., & Dowdy, E. (2013). Screening for behavioral and emotional risk versus traditional school identification methods. *School Mental Health, 6,* 40–49. doi:10.1007/s12310-013-9109-1

Eklund, K., Renshaw, T., Dowdy, E., Jimerson, S., Hart, S., Jones, C., & Earhart, J. (2009). Early identification of behavioral and emotional problems in youth: Uni-

versal screening versus teacher-referral identification. *California School Psychologist, 14,* 89–95.

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.

Feeney-Kettler, K. A., Kratochwill, T. R., Kaiser, A. P., Hemmeter, M. L., & Kettler, R. J. (2010). Screening young children's risk for mental health problems: A review of four measures. *Assessment for Effective Intervention, 35,* 218–230.

Feeney-Kettler, K. A., Kratochwill, T. R., & Kettler, R. J. (2011). Identification of preschool children at risk for emotional and behavioral disorders: Development and validation of a universal screening system. *Journal of School Psychology, 49,* 197–216. doi:10.1016/j.jsp.2011.02.001

Flanagan, K. S., Bierman, K. L., Kam, C. M., & Conduct Problems Prevention Research Group. (2003). Identifying at-risk children at school entry: The usefulness of multi-behavioral problem profiles. *Journal of Clinical Child and Adolescent Psychology, 32,* 396–407. doi:10.1207/S15374424JCCP3203_08

Glascoe, F. P. (2001). Can teachers' global ratings identify children with academic problems? *Journal of Developmental and Behavioral Pediatrics, 22,* 163–168. doi:10.1097/00004703-200106000-00002

Glover, T., & Albers, C. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology, 45,* 117–135. doi:10.1016/j.jsp.2006.05.005

Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 1337–1345. doi:10.1097/00004583-200111000-00015

Greene, R., Beszterczey, S. K., Katzzenstein, T., Park, K., & Goring, J. (2002). Are students with ADHD more stressful to teach? Patterns of teacher stress in an elementary school sample. *Journal of Emotional and Behavioral Disorders, 10,* 79–89. doi:10.1177/10634266020100020201

Harrison, J. R., Vannest, K. J., & Reynolds, C. R. (2013). Social acceptability of five screening instruments for social, emotional, and behavioral challenges. *Behavioral Disorders, 38,* 171–189.

Ingersoll, R. M. (2001). Teacher turnover and teacher shortages: An organizational analysis. *American Educational Research Journal, 38,* 499–534. doi:10.3102/00028312038003499

Kamphaus, R., & Reynolds, C. (2007). *Behavior Assessment System for Children–Second Edition: Behavioral and emotional screening system (BESS).* Bloomington, MN: Pearson.

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. *The American Journal of Psychiatry, 159,* 1548–1555. doi:10.1176/appi.ajp.159.9.1548

Keenan, K., & Wakschlag, L. S. (2004). Are oppositional defiant and conduct disorder symptoms normative behaviors in preschoolers? A comparison of referred and nonreferred children. *American Journal of Psychiatry, 161,* 356–358. doi:10.1176/appi.ajp.161.2.356

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry, 62,* 593–602. doi:10.1001/archpsyc.62.6.593

Kettler, R. J., & Feeney-Kettler, K. A. (2011). Screening systems and decision making at the preschool level: Application of a comprehensive validity framework. *Psychology in the Schools, 48,* 430–441. doi:10.1002/pits.20565

Ladd, G. W., Birch, S. H., & Buhs, E. S. (1999). Children's social and scholastic lives in kindergarten: Related spheres of influence? *Child Development, 70,* 1373–1400. doi:10.1111/1467-8624.00101

Landau, S., & Swerdlik, M. E. (2005). What you see is what you get: A commentary on school-based direct observation systems. *School Psychology Review, 34,* 529–536.

Lloyd, J. W., Kauffman, J. M., Landrum, T. J., & Roe, D. L. (1991). Why do teachers refer pupils for special education? An analysis of referral records. *Exceptionality, 2,* 115–126. doi:10.1080/09362839109524774

Loe, I. M., & Feldman, H. M. (2007). Academic and educational outcomes of children with ADHD. *Journal of Pediatric Psychology, 32,* 643–654. doi:10.1093/jpepsy/jsl054

Mathiesen, K. S., & Sanson, A. (2000). Dimensions of early childhood behavior problems: Stability and predictors of change from 18 to 30 months. *Journal of Abnormal Child Psychology, 28,* 15–31. doi:10.1023/A:1005165916906

McClelland, M. M., Acock, A. C., & Morrison, F. J. (2006). The impact of kindergarten learning-related skills on academic trajectories at the end of elementary school. *Early Childhood Research Quarterly, 21,* 471–490. doi:10.1016/j.ecresq.2006.09.003

Mikami, A. Y., & Hinshaw, S. P. (2006). Resilient adolescent adjustment among girls: Buffers of childhood peer rejection and attention-deficit/hyperactivity disorder. *Journal of Abnormal Child Psychology, 34,* 823–837. doi:10.1007/s10802-006-9062-7

Moilanen, D. L., Shaw, D. S., & Maxwell, K. L. (2010). Developmental cascades: Externalizing, internalizing and academic competence from middle childhood to early adolescence. *Development and Psychopathology, 22,* 635–653. doi:10.1017/S095457941000033

National Research Council and Institute of Medicine. (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities.* Washington, DC: National Academies Press.

Owens, J. S., Johannes, L. M., & Karpenko, V. (2009). The relation between change in symptoms and functioning in children with ADHD receiving school-based mental health services. *School Mental Health, 1,* 183–195. doi:10.1007/s12310-009-9020-y

Pelham, W. E., Gnagy, E. M., Greenslade, K. E., & Milich, R. (1992). Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 3,* 210–218.

Pihlakoski, L., Sourander, A., Aromaa, M., Rautava, P., Heleius, H., & Sillanpaa, M. (2006). The continuity of psychopathology from early childhood to preadolescence: A perspective cohort study of 3–12-year-old children. *European Journal of Child and Adolescent Psychiatry, 15,* 409–417. doi:10.1007/s00787-006-0548-1

Pintea, S., & Moldovan, R. (2009). The receiver-operating characteristic (ROC) analysis: Fundamentals and applications in clinical psychology. *Journal of Cognitive and Behavioral Psychotherapies, 9,* 49–66.

Racz, S. J., King, K. M., Wu, J., Witkiewitz, K., & McMahon, R. J. (2013). The predictive utility of a brief kindergarten screening measure of child behavior problems. *Journal of Consulting and Clinical Psychology, 81,* 588–599. doi:10.1037/a0032366

Ramey, C. T., & Ramey, S. L. (2004). Early learning and school readiness: Can early intervention make a difference? *Merrill-Palmer Quarterly, 50,* 471–491.

Reinke, W., Stormont, M., Herman, K., Puri, R., & Goel, N. (2011). Supporting children's mental health in schools: Teacher perceptions of needs, roles, and barriers. *School Psychology Quarterly, 26,* 1–13. doi:10.1037/a0022714

Squires, J., Bricker, D., & Twombly, E. (2003). *Ages and stages questionnaires: Social-emotional.* Baltimore, MD: Brookes.

Stone, L. L., Otten, R., Engels, R., Vermulst, A., & Janssen, J. (2010). Psychometric properties of the parent and teacher versions of the Strengths and Difficulties Questionnaire for 4- to 12-year-olds: A review. *Clinical Child and Family Psychology Review, 13,* 254–274.

Suldo, S. M., Gormley, M. J., DuPaul, G. J., & Anderson-Butcher, D. (2014). The impact of school mental health on student and school-level academic outcomes: Current status of the research and future directions. *School Mental Health, 6,* 84–98. doi:10.1007/s12310-013-9116-2

Swanson, J. M., Kraemer, H. C., Hinshaw, S. P., Arnold, L. E., Conners, C. K., Abikoff, H. B., . . . Wu, M. (2001). Clinical relevance of the primary findings of the MTA: Success rates based on severity of ADHD and ODD symptoms at the end of treatment. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 168–179.

Swets, J. A. (1996). *Signal detection theory and ROC analysis in psychology and diagnosis: Collected papers.* Mahwah, NJ: Erlbaum.

U.S. Census Bureau. (2006). *State and county quick facts.* Retrieved September 12, 2013, from http://quickfacts.census.gov/qfd/states/39000.html

U.S. Census Bureau. (2013). *America's children: Key national indicators of well-being, 2013.* Retrieved September 12, 2013, from http://www.childstats.gov/americaschildren/famsoc3.asp

U.S. Department of Labor. (2013). *Education pays in higher earnings and lower unemployment rates.* Retrieved May 19, 2013, from http://www.bls.gov/emp/ep_chart_001.htm

Wakschlag, L. S., Briggs-Gowan, M. J., Carter, A. S., Hill, C., Danis, B., Keenan, K., & Leventhal, B. L. (2007). A developmental framework for distinguishing disruptive behavior from normative misbehavior in preschool children. *Journal of Child Psychology and Psychiatry, 48,* 976–987. doi:10.1111/j.1469-7610.2007.01786.x

Walker, H., Nishioka, V., Zeller, R., Severson, H., & Feil, E. (2000). Causal factors and potential solutions for the persistent underidentification of students having emotional or behavioral disorders in the context of schooling. *Assessment for Effective Intervention, 26,* 29–39. doi:10.1177/073724770002600105

Walker, H. M., Seeley, J., Small, J. R., Severson, J., Graham, H. H., Feil, B. A., . . . Forness, S. R. (2009). A randomized controlled trial of the first step to success early intervention: Demonstration of program efficacy outcomes in a diverse, urban school district. *Journal of Emotional and Behavioral Disorders, 17,* 197–212. doi:10.1177/1063426609341645

Webster-Stratton, C., Rinaldi, J., & Reid, J. M. (2010). Long-term outcomes of Incredible Years Parenting Program: Predictors of adolescent adjustment. *Child and Adolescent Mental Health, 16,* 38–46. doi:10.1111/j.1475–3588.2010.00576.x

VanDerHeyden, A. M., Witt, J. C., & Naquin, G. (2003). Development and validation of a process for screening referrals to special education. *School Psychology Review, 32,* 204–227.

Volpe, R. J., McConaughy, S. H., & Hintze, J. M. (2009). Generalizability of classroom behavior problem and on-task scores from the Direct Observation Form. *School Psychology Review, 38,* 382–401.

Julie Sarno Owens, PhD, is an associate professor of psychology and Co-Director of the Center for Intervention Research in Schools at Ohio University. Her program of research is focused on (a) evaluating the effectiveness, feasibility, and sustainment of evidence-based interventions under typical school conditions and (b) developing or modifying procedures to address barriers to intervention implementation and positive student outcomes.

Jennifer Storer, PhD, is a supervised clinical psychologist with KVC West Virginia, a nonprofit foster care and child welfare agency. She is also a clinical assistant professor of behavioral medicine and psychiatry at West Virginia University School of Medicine. Her work focuses primarily on engagement of families in treatment and perceived barriers to treatment, as well as the assessment and treatment of children with developmental and trauma-related disorders.

Alex S. Holdaway is a doctoral candidate at the Center for Intervention Research in Schools at Ohio University. His program of research is focused on (a) developing and evaluating evidence-based training and professional development for school-based professionals and (b) modifying procedures to address barriers to intervention implementation and positive student outcomes.

Verenea J. Serrano is a doctoral candidate at the Center for Intervention Research in Schools at Ohio University. Her program of research is focused on understanding factors that contribute to social and emotional impairment in youth with attention deficit hyperactivity disorder.

Yuko Watabe, PhD, is a psychology fellow at Children's Hospital Los Angeles, where she provides psychotherapy and assessment services for children with developmental disabilities. Her research interests include the effectiveness and sustainability of evidence-based treatment for ethnic-minority populations.

Lina K. Himawan is a biostatistician in the Department of Psychology at Ohio University. She assists faculty members and graduate students in the department with writing grant proposals, conducting power analyses, selecting research designs, conducting data analyses, and presenting results of analyses for publication.

Rebecca E. Krelko is a second-year master's student in the Department of Social Work at The Ohio State University. She holds a bachelor of arts from Ohio University.

Katherine J. Vause graduated from Ohio University in 2012. She is currently working as a faculty research assistant as a data manager in the human development and quantitative methodology department.

Erin Girio-Herrera, PhD, is a clinical research scientist at the Center for Intervention Research in Schools at Ohio University. Her research focuses on understanding and developing strategies to promote mental health service utilization for youth through (a) improving identification of mental health problems, (b) reducing parental barriers to mental health service utilization, and (c) engaging parents in mental health service utilization.

Nina C. Andrews is a licensed school psychologist who practices in rural southeastern Ohio. Her areas of interest include early intervention, social–emotional development, school mental health, and dropout prevention. She works closely with local resources including Ohio University and community mental health centers to provide school mental health services to students across the district and is a certified Youth Mental Health First Aid provider.