Journal of School Psychology 90 (2022) 114–134

Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

Examining the validity of the Early Identification System – Student Version for screening in an elementary school sample

Wendy M. Reinke *, Keith C. Herman, Francis Huang, Chynna McCall, Shannon Holmes, Aaron Thompson, Sarah Owens

University of Missouri, Missouri Prevention Science Institute, United States of America

ARTICLE INFO

Action Editor - Craig Albers

Keywords:
Universal screening
Prevention
Social-emotional health
Elementary school

ABSTRACT

As many as 1 in 5 youth in the United States experience social, emotional, and behavioral problems. However, many students with mental health concerns are unidentified and do not receive adequate care. The purpose of this study was to evaluate the factor structure, measurement invariance, and the concurrent and predictive validity of the Early Identification System-Student Report (EIS-SR), a screener for social, emotional, and behavioral problems, using a sample of over 5000 students from Grades 3 to 5. The EIS-SR was developed by using extant literature on the risk indicators that lead to social, emotional, and behavioral challenges among children and youth. As expected, seven subscales were identified as having adequate factor loadings. Furthermore, the measure was determined to be invariant across grade level ($n = 5005$), gender ($n = 5005$), and between Black and White students ($n = 1582$). The concurrent validity of the Internalizing Behavior, Attention and Academic Issues, Emotion Dysregulation, and School Disengagement subscales was supported by correlations with comparable subscales of the Behavior Assessment System for Children-3rd Edition (BASC-3; $n = 382$). Additionally, the EIS-SR subscales administered in the fall of the school year were predictive of important outcomes in spring, including attendance ($n = 4780$), disciplinary referrals ($n = 4938$), bully victimization ($n = 4670$), math academic achievement scores ($n = 4736$), and reading ($n = 4772$) academic achievement scores. The EIS-SR holds promise as a feasible and technically adequate screening tool for use in elementary schools.

1. Introduction

As many as 1 in 5 youth in the United States experience social, emotional, and behavioral problems (Centers for Disease Control and Prevention [CDC], 2013) with the onset of many of these concerns beginning in early childhood. Youth experiencing these difficulties are at increased risk of long-term, pervasive, and deleterious problems, including poor literacy (Bulotsky-Shearer et al., 2010), peer rejection (Wood et al., 2002), delinquency (Copeland et al., 2007), substance abuse (Center for Behavioral Health Statistics and Quality, 2015), later mental health problems (Harvey et al., 2009), and multiple other outcomes that come at a significant cost to society.

Despite our knowledge of the existing mental health problems among students, many students with mental health concerns are unidentified and do not receive adequate care (Georgiadis et al., 2019; Merikangas et al., 2010; Soneson et al., 2020). Unfortunately,

* Corresponding author.
E-mail address: reinkew@missouri.edu (W.M. Reinke).

https://doi.org/10.1016/j.jsp.2021.10.001
Received 16 December 2019; Received in revised form 22 May 2021; Accepted 15 October 2021
Available online 10 December 2021

only 20%–33% of children needing mental health services receive them (Geogiadis et al., 2019; Kataoka et al., 2002). However, if all children with mental health concerns were to seek support from the mental health system, the need likely would further overwhelm available services (Costello et al., 2003; Splett et al., 2018). In fact, there have been long-standing and well-documented shortages within the mental health workforce. For example, there is not a single state in the U.S. that is adequately staffed with child psychiatrists to meet the needs of our youth (Tyler et al., 2017) despite knowledge that these social, emotional, and behavioral problems are preventable and amenable to early intervention (O'Connell et al., 2009). In schools, the ratio of school psychologists to students (1:1233; Goforth et al., 2021) is more than double the ratio recommended for optimal service delivery to students (1:500; National Association of School Psychologists [NASP], 2020). Early identification of youth mental health concerns is critical and necessary to stymie the development of serious mental illness and prevent associated adverse youth and system outcomes.

Schools often are the optimal setting to provide prevention and early intervention services to youth. Not only do youth spend a considerable portion of their day in schools, but schools are typically staffed with professionals that have the expertise and necessary training to support these efforts. In fact, students are more likely to seek mental health services when they are available in schools (NASP, 2016). Moreover, it is increasingly common for schools to adopt multi-tiered systems of support to meet students' social, emotional, behavioral, and academic needs (NASP, 2016). In these tiered systems, a continuum of prevention and intervention services are provided to promote youth's healthy functioning. The key feature of a well-functioning tiered system is the systematic use of universal screening. Universal screening assists school decision-makers in efficiently identifying at-risk students and placing them in appropriate interventions. As a result, schools need easily accessible, efficient screening measures that accurately identify students in need of supports across social, emotional, and behavioral areas of functioning.

1.1. Limitations of existing screeners

Although many behavior screeners have been developed and evaluated (e.g., Behavioral and Emotional Screening System [BESS], Kamphaus & Reynolds, 2015; Social, Academic, and Emotional Behavioral Risk Screener [SAEBRS], Kilgus & von der Embse, 2014; Social Skills Improvement System [SSIS], Gresham & Elliott, 2007; Systematic Screening for Behavioral Disorders [SSBD], Walker et al., 2014), their use in schools has been limited. For example, fewer than 15% of schools in the U.S. have reported that they routinely conduct universal behavior screenings (Bruhn et al., 2014; Romer & McIntosh, 2005). Recent surveys of representative samples of districts (N = 1330 district administrators; Dineen et al., 2021) and schools (N = 475 building administrators; Briesch et al., 2021) throughout the country revealed that only 5.5% of districts and 9% of schools reported conductinng social, emotional, or behavioral screening for all students. Their lack of use in typical school practice suggests that there may be a misfit between these measures and school contexts. In particular, although most existing screeners are appropriate and technically adequate for use in schools, they fail the usability precondition for widespread adoption and implementation identified by Glover and Albers (2007). First, many social, emotional, and behavioral problems co-occur with one another, and identifying these co-occurrences is important to maximize functionality and usability. For instance, more than 1 in 3 school-aged youth with anxiety, depression, and externalizing problems have another co-occurring mental health concern (Ghandour et al., 2018). However, many existing screeners are limited in scope and focus on a few related areas of risk (e.g., internalizing and externalizing problems; see the BESS and SSBD) rather than the constellation of risk conditions experienced by many students. Second, universal screening needs to be feasible and efficient for school personnel to administer and to readily access the data due to the ever-increasing demands placed on educators' and students' time (Dowdy et al., 2010). Unfortunately, many of the currently available screening measures are lengthy and time-intensive (Dowdy et al., 2010); even a scale that only takes 5 min to complete per student would require 1 h and 40 min for a teacher to rate a class of 20 students. Few schools have the resources or infrastructure to score, interpret, and apply the findings from existing screening tools (Briesch et al., 2018; Bruhn et al., 2014; O'Connell et al., 2009). Additionally, many existing screeners come at a high monetary cost to schools. Finally, most available screeners provide national normed comparisons to identify student risk and in turn result in the identification of many students in schools serving students with higher needs, overwhelming schools that often have limited resources (Volpe et al., 2010).

1.2. The Early Identification System (EIS)

The Early Identification System (EIS) was developed to overcome several of the existing challenges in school-based screening and to assist schools in their capacity to identify students who are experiencing social, emotional, and behavioral problems. A key feature of the EIS is that it emerged from a community investment in the form of a local tax to support youth mental health. School and research partners reviewed available screeners and determined that it was necessary to develop a new tool that could be used community-wide without incurring ongoing costs. With attention to the need to assess risks across multiple social, emotional, and behavioral challenges faced by students, a feasible and cost-effective screening system was developed.

The development of the EIS was grounded in a longstanding, well-supported theory of the origins of children's risk for social, emotional, and behavioral challenges (i.e., the cascades model; Patterson et al., 1992). Furthermore, the EIS is part of a system that includes both student and teacher report (i.e., the EIS–Teacher Report [EIS-TR] and the EIS–Student Report [EIS-SR]). Not only do these two forms provide data of risk *across* reporters, but this approach also offers an opportunity for students to report on risk items that are often difficult for teachers to observe (e.g., internalizing problems, bullying). A distinguishing feature of the EIS is that the EIS-SR can be implemented in Grades 3–12 and uses identical items across developmental levels, which allows for the data to be aggregated in unique ways. For example, the data can be aggregated at the county-, district-, and school-levels, thereby allowing for quick identification of risk patterns over time. To attend to the issue of feasibility, the EIS-SR was developed to be administered and scored online, takes less than 10 min to complete, and provides immediate data to schools via a dashboard report. The dashboard provides feedback regarding the data in a user-friendly red (high risk), yellow (some risk), and green (no risk) format, which guides intervention delivery (see Reinke, Thompson, et al., 2018).

The county youth mental health tax structure noted earlier provides the impetus for reviewing these data at the county-level to determine needed community initiatives. As a result of the utility of the EIS within this county-based initiative, the EIS is intended to be disseminated to rural schools throughout the nation as part of the Institute for Educational Sciences Center for Rural Schools Mental Health (https://www.ruralsmh.com/). Given the expected widespread adoption and use of the EIS-SR, examining the psychometric properties of the scores in school-age populations is needed.

1.3. Developmental cascades model of risk indicators

Patterson (1982) developed one of the most influential theories describing the processes and mechanisms that contribute to the development of early antisocial behaviors in youth (Patterson et al., 1989). This early work described the coercive interactions between parent and child before school entry that escalates disruptive, non-compliant, and antisocial child behaviors through a negative reinforcement cycle (Patterson, 1982). The subsequent theory—variously referred to as *developmental cascades* (Masten & Cicchetti, 2010), *dual failure/dual cascades* (Obradović et al., 2009; Patterson & Stoolmiller, 1991), and *social interaction learning* (SIL) models (Forgatch & DeGarmo, 2002)—described the emergence of numerous risk patterns that undermined youth development. According to this model, many children experience challenges adapting to the new social field when they enter school. Pre-existing problems, including coercive interactions at home, antisocial/externalizing behaviors, emotional regulation skill deficits, and attention problems set youth up for failure when they enter school. Underperformance in academic and social domains portend further deterioration of child self-perceptions, peer and teacher relations, and academic performance (Reinke et al., 2008). Repeated failure experiences and negative feedback from social interactions increase youth risk levels for internalizing symptoms, aggressive behavior, maladaptive peer interactions, and school disengagement. These cascades of failure in academic and/or social fields can escalate with each new developmental transition or challenge.

Much developmental literature supports the cascade model. For instance, early externalizing concerns predict future social and academic challenges which in turn predict internalizing symptoms in adulthood (Masten et al., 2005; Obradović et al., 2009). Moreover, without intervention, youth experiencing these developmental cascades are at significant risk for school drop-out, deviant peer relations, substance abuse, delinquency, and arrest (Masten & Cicchetti, 2010; Patterson et al., 1989). An equally impressive body of research has also shown that disrupting the cascades—by reducing risk conditions with an intervention such as effective parent behavior management—alters these negative trajectories and reduces risk for future social, emotional, behavioral, and academic concerns (Patterson et al., 2010). Thus, identifying youth experiencing elevated levels of risk on key indicators (i.e., externalizing, internalizing, attention, peer, academic, emotional regulation, and school engagement problems) and providing effective interventions and supports is critical for interrupting developmental cascades.

The pervasiveness and harmful effects associated with childhood social, emotional, and academic difficulties identified by the developmental cascades model were a driving force in the development of the screening items for the EIS-SR. Items tapping risk for externalizing behaviors, social skill deficits, difficulties with peer relationships, internalizing behaviors, inattention, and problems with academic competence and school engagement were developed due to their co-occurrence with challenging social behaviors and academic failure. Indicators of bullying experiences were also included due to the association between bullying and the risk indicators, as well as other serious adverse outcomes, including suicide attempts and death (Gini & Espelage, 2014). The following section provides a review of empirical evidence supporting the links among the various risk indicators identified by the developmental cascade model with negative social, emotional, and behavioral outcomes.

1.3.1. Externalizing problems

Externalizing problems refer to observable disruptive, non-compliant, and aggressive behaviors that exceed developmental normative behaviors in frequency, intensity, and/or duration (Liu et al., 2004). Externalizing behaviors, sometimes referred to as antisocial behaviors, took center stage in the cascade model given Patterson's (1982) early work focused on observing coercive aggressive interaction cycles between parent and child during the toddler years. Consistent with the cascade theory, early externalizing problems are among the more robust predictors of academic failure and social rejection as well as future school dropout, delinquency, drug use, and criminal involvement (Odgers et al., 2008; Temcheff et al., 2008). Many studies have found that disruptive behaviors in early childhood predict future problems across development and into adulthood (Masten et al., 2005; Obradović et al., 2009). In one study, youth that were identified with behavior problems in Grade 1 were seven times as likely to receive special education placement and nearly three times as likely to be arrested by Grade 12 as compared to youth without behavior problems in Grade 1 (Darney et al., 2013). A separate study identified a subgroup of youth who had stable-high levels of aggressive-disruptive behaviors from Grades 1 to 7, which included about 10% of the sample. These youths' odds of juvenile and adult arrest were 20 and 10 times higher, respectively, than youth with stable-low levels of these problem behaviors (Schaeffer et al., 2003).

1.3.2. Emotional dysregulation

Emotional regulation refers to the capacity to monitor and manage affect and sustain attention. Early externalizing behavior and emotion dysregulation are intimately interrelated and in combination predict long-term negative youth outcomes (Côté et al., 2006; Taylor et al., 2007; Tremblay et al., 2005). Not surprisingly, youth with emotional dysregulation problems more often engage in aggressive and disruptive behaviors given their limited capacity to regulate their emotions in response to conflict and disagreement. Thus, emotional dysregulation serves as a risk factor for children to engage in coercive interactions with parents and other adults at school entry (Reinke & Herman, 2002). Additionally, dysfunctional regulation in early elementary school serves as an independent and unique predictor of poor academic and mental health outcomes in adolescence (Deutz et al., 2020). Evidence also suggests that emotional dysregulation precedes and contributes to future risk for a wide range of psychopathology; one study with adolescents found that baseline emotional dysregulation predicted the escalation of anxiety symptoms, aggressive behavior, and eating pathology (McLaughlin et al., 2011).

1.3.3. Attention problems

Although much research on the cascade model has focused on the role of early antisocial behaviors as the antecedents to proximal and distal adverse youth outcomes, subsequent research has also implicated attention problems as a common co-occurring factor. Attention problems are defined by difficulty in sustaining attention, distractibility, forgetfulness, and disorganization as compared to developmental norms. For instance, Hinshaw (1992) reviewed evidence suggesting that the hypothesized link between conduct problems and academic problems could be explained by comorbid attention deficits. That is, studies that have examined the links between these three concerns simultaneously consistently find that only attention problems, and not conduct problems, precede and predict academic problems. Similarly, in a series of studies, Herman et al. (2007) and Herman et al. (2008) found that the relation between conduct problems and depression was no longer significant when attention problems were considered.

Often rooted in early neurodevelopmental deficits, attention problems often co-occur with a host of other issues, including impulsivity, working memory and expressive language deficits, and emotional dysregulation, all of which increase the likelihood that youth will present challenging behaviors to parents and other adults who unwittingly engage in coercive processes in an attempt to address these behaviors (Barkley & Cunningham, 1979; Danforth et al., 1991; Fischer, 1990; Mash & Johnston, 1990). Thus, it is not surprising that early attention problems are linked to the same downstream effects described by the cascade model. For instance, inattention is a significant risk factor for poor future academic and social-emotional outcomes (Herman et al., 2007; Stormont & Thomas, 2014; Zentall, 2005). Children who struggle with inattention tend to have issues with making friends and maintaining positive peer relationships (Gagnon et al., 1995; Hoza, 2007) due to poor emotion regulation (Barkley, 2010; Melnick & Hinshaw, 2000; Nigg & Casey, 2005). Students with inattention also tend to have problems completing assignments and meeting teachers' expectations, which is a proximal cause of associated school and academic problems (Barkley, 1998). Furthermore, students with inattention also have an increased risk for developing depression and experiencing peer rejection, compounding their risk for significant social and emotional problems. In a study with 426 Black students in Baltimore City, Herman et al. (2007) found that inattention in first grade was linked to depression in third grade and that the relationship was mediated by academic performance. The findings held when controlling for disruptive behaviors and academic skills in first grade, thus suggesting the path to depression is specific to attention problems rather than a more general externalizing or school readiness pathway. The authors replicated these findings with a sample of predominately White youth in Minnesota (Herman & Ostrander, 2007).

1.3.4. Peer relationship problems

Peer problems can emerge at any stage of development and interfere with social, emotional, behavioral, and academic functioning. Consistent with the cascade model, peer problems are interrelated with externalizing problems and emotional dysregulation. Peer rejection and social skill deficits are all interrelated problems associated with adverse developmental outcomes in adolescence and young adulthood (Côté et al., 2006; Taylor et al., 2007; Tremblay et al., 2005). Burt and Roisman (2010) found that the relationship between early externalizing problems and negative outcomes in adolescence was mediated by social skills deficits. Furthermore, externalizing behavior and emotion regulation problems in middle childhood are highly negatively correlated with social competence and leads to peer marginalization, low academic achievement, low levels of school bonding, school dropout, association with delinquent peers, and—in some communities—substance abuse and gang involvement (Dishion et al., 2010).

1.3.5. Bullying

This intersection of externalizing behavior, emotion regulation, peer rejection, and social skill deficits can also result in bullying behavior. Bullying is an undesired, continuous, and relationally aggressive behavior based upon one person's real or perceived power. Similar to externalizing behaviors, bullying is influenced by social relationships and social skill development (Swearer & Hymel, 2015). In fact, approximately half of the students that indicate they have engaged in bullying behaviors report that they have been bullied (Haynie et al., 2001). A recent paper found that 27.5% of students were characterized by occasional to frequent broad involvement in both bullying behavior and victimization and that nearly all students who bullied fell into one of two profiles that included at least occasional victimization experiences (Demaray et al., 2021). Furthermore, children who have trouble regulating emotions and who display aggressive behaviors within the context in which classroom norms are unfavorable to such actions are more likely to report being bully-victims (Brendgen et al., 2015).

As compared to children who are not bullied, children who experience bullying are more likely to experience depression, anxiety, health issues, and problems in school (e.g., absences, lower academic achievement; U.S. Department of Health and Human Services [DHHS], 2017). Being a victim of bullying has been directly tied to the development of internalizing problems among young children (Arseneault et al., 2008). Within Arseneault et al., monozygotic twins who had been bullied had more internalizing symptoms as compared with their co-twin who had not been bullied. This effect remained significant after controlling for preexisting internalizing problems. Additionally, children who bully others are more likely than their non-bullying counterparts to engage in risk-taking and violent behaviors as they get older, including alcohol and drug use, physical aggression, and domestic violence (DHHS, 2017). Children who witness bullying are at risk for depression, anxiety, drug and alcohol use, and repeated school absences (DHHS, 2017). Thus, the co-assessment of internalizing and bullying behaviors is important and needed.

1.3.6. Internalizing problems

In the cascade model, internalizing problems often are a downstream consequence of early aversive interactions with the social environment and repetitive failure experienced in key social and academic domains. Internalizing symptoms and disorders are some of the most prevalent emotional and behavioral problems experienced by youth. The National Comorbidity Survey–Adolescent Supplement (NCS-A), a nationally representative face-to-face survey of 10,123 adolescents aged 13–18 years in the United States, found anxiety disorders to be the most common mental disorder among youth (31.9%) and to have one of the earliest median ages of onset (6 years; Merikangas et al., 2010). Mood disorders, such as depression, were nearly as common (14.3%) as externalizing disorders in youth. Prior research has also shown that depressive symptoms that are not severe enough to meet diagnostic criteria have a comparable negative impact on youth functioning (Lewinsohn et al., 2003). In one study, Lewinsohn and colleagues (2000) found that depressive symptoms in adolescence had a strong positive, linear association with functional outcomes and that there was a not a diagnostic threshold that distinguished high and low functioning.

Depression and anxiety symptoms in youth often go unnoticed, and most do not receive any services for their symptoms (Avenevoli et al., 2015; Beardslee et al., 1993; Cheung et al., 2013; Keller et al., 1991; Lewinsohn et al., 1998). However, later depression and anxiety are predicted by a student's high levels of internalizing problems in elementary school (Mash & Wolfe, 2015). Left unaddressed, youth with elevated internalizing problems tend to have difficulties with interacting in a prosocial manner (Henricsson & Rydell, 2006) and experience challenges with their behavioral, cognitive, and academic functioning (Bub et al., 2007; Burt et al., 2008). One indication that a student may be experiencing internalizing problems, such as anxiety or depression, is school attendance. For instance, Finning et al. (2019) conducted a systematic review of the literature and found small to moderate associations between depression and poor school attendance, particularly absenteeism and unexcused absences/truancy. Thus, chronic or unexcused absences can be associated with internalizing problems.

1.3.7. Academic competence deficits

The social, emotional, and behavioral risk factors noted above all have associations with academic failure. Academic competence refers to the skills, attitudes, and behaviors of learners that contribute to success at school (DiPerna & Elliott, 2002). Academic competence deficits can be both a consequence of maladaptive social and behavioral responses and cause for subsequent negative social, emotional, and behavioral outcomes. Failure at school is associated with a range of negative behavioral and emotional outcomes; consequently, identifying early indicators of academic risk are essential in preventing these adverse outcomes. For example, students with learning problems have elevated rates of emotional disorders in both clinical and community samples (Darney et al., 2013; Herman et al., 2021; Huntington & Bender, 1993; Maughan et al., 2003; Reinke et al., 2008). In addition to the commonly cited co-occurrence between disruptive behavior problems and academic failure, researchers have shown that students with academic problems also experience higher levels of anxiety, general unhappiness, and depression (Herman et al., 2008; Herman et al., 2020; Maughan et al., 2003).

1.3.8. School disengagement

School engagement refers to cognitive, behavioral, and emotional motivation to learn and progress in school (Li & Lerner, 2012). School disengagement is a risk indicator for a host of negative outcomes, including underachievement, grade retention, school dropout, delinquency, and substance abuse (Henry et al., 2012; Rhodes et al., 2018; Wang & Fredricks, 2014). For instance, in a sample of 911 students, Henry et al. (2012) identified school disengagement in Grades 8 and 9 as a robust predictor of dropout and serious behavior problems in late adolescence and early adulthood. Similarly, Wang and Fredricks (2014) found that decreases in engagement during middle and high school ($N = 1272$ students) predicted increased rates of delinquency and substance abuse over time. Existing conceptualizations and evidence suggest that school engagement is both an early risk indicator of academic and social failure and a contributing cause of subsequent failure (Chase et al., 2014; Li & Lerner, 2012).

In summary, the risk indicators identified for use on the EIS-SR are grounded in sound theory and have strong associations with significant adverse academic, behavioral, emotional, and social outcomes. Identifying students in need of further intervention is crucial in a preventive tiered service delivery model, as identification is often considered the first step in the problem-solving process (Glover & Albers, 2007). When students are exposed to supports that are tied directly to targeted areas of need, there can be significant improvements in their social, emotional, behavioral, and academic functioning (Durlak et al., 2011; Taylor et al., 2017).

1.4. Scale development and initial evaluation

Based on these relevant and meaningful risk characteristics and their co-occurrence, the EIS development team met to discuss each domain and begin item-generation aligned with the definitions of these domains. The item pool was then reviewed by measurement experts in the field (academics with expertise in school screening measures) and by personnel in the schools (including school psychologists and teachers) who would be using the data. Experts were asked to comment on whether each item accurately represented the domain of interest and whether any other items were needed to fully capture the domain. The pool was narrowed to 42 items based on this expert input and refinement; two of these items were intended as stand-alone risk indicators (e.g., bullying and peer victimization) rather than as items intended to load on the broader domains. An initial investigation with 1590 elementary and middle school students largely supported the psychometric properties of the EIS-SR with excellent model fit and strong scale reliabilities (Huang et al., 2019). Exploratory and confirmatory factor analysis (CFA) with randomly split subsamples supported a 6-factor solution. In these analyses, items on the attention problems subscale fit under the externalizing behavior domain. The result was not entirely surprising given that externalizing problems and attention issues are highly correlated (Angold et al., 1999). Additionally, the relatively small subsamples, the mix of elementary and middle school students, and the overrepresentation of parochial school students in the prior study suggested the need for further work to examine the internal structure of the EIS-SR (Huang et al., 2019). Given the strong theoretical foundation of the EIS-SR and clear predictions about its factor structure based on this framework, further examination and replication of its internal structure with CFAs in an elementary student sample was warranted (Fabrigar & Wegener, 2012).

1.5. Validity framework

The EIS was developed from a unified view of validity (Messick, 1995); that is, the scale developers designed the measure to address all facets of construct validity, including its relevance and use, values implications, and social consequences as described by Good and Jefferson (1998). Thus, the EIS development team was interested in creating a measure that not only had adequate technical properties as defined by the Standards for Educational and Psychological Testing (e.g., test content, internal structure, relations with other variables; American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014), but also would be used by schools to draw appropriate inferences, take meaningful and impactful action (e.g., deliver effective interventions and supports), and maximize benefits for school personnel and students while minimizing harmful consequences. These latter sources of test validity would all be subsumed under the Standards' definition of consequences of test use.

1.6. Purpose and research questions

The purpose of this study was to evaluate key technical aspects of the EIS-SR's construct validity, including its (a) factor structure, (b) measurement invariance, (c) concurrent validity, and (d) predictive validity in relation to school disciplinary data, bully victimization, attendance, and academic outcomes with a large community elementary sample of students. Concurrent validity refers to how correlated the EIS-SR subscales are with other similarly measured constructs and predictive validity indicates how well the EIS-SR subscales predict certain future outcomes (Boateng et al., 2018). Regarding the EIS-SR's technical adequacy, we asked four specific research questions. First, what is the latent factor structure of the EIS-SR? It was hypothesized that the EIS-SR would have seven subscales, including Externalizing Behavior, Internalizing Behavior, Peer Relationship Problems, Attention and Academic Issues, Emotion Dysregulation, Relational Aggression, and School Disengagement. Second, are the subscales invariant across gender, grade level, and race? It was hypothesized that these subscales would be invariant, indicating that the instrument is measuring the same construct in a similar manner across these groups. Third, what is the concurrent validity of each EIS-SR subscale as compared to analogous subscales of the student report of the Behavioral Assessment System for Children (BASC-3; Reynolds et al., 2015)? As evidence of concurrent validity, where the EIS-SR and the BASC are hypothesized to be measuring the same construct, we expected that four EIS-SR subscales would be moderately correlated with the BASC-3, a commonly used behavioral and emotional assessment with a rich psychometric history. Lastly, do the EIS-SR subscales predict student behavioral and academic outcomes? We hypothesized that the EIS-SR subscales, measured in the fall of the school year, would predict student disciplinary data, bully victimization, attendance, and academic outcomes (e.g., standardized tests in communication arts and mathematics) measured in the spring of the school year when accounting for all subscales in the same model. In particular, it was expected that higher scores on Externalizing Behavior, Emotion Dysregulation, Peer Relationship Problems, School Disengagement, and Relational Aggression subscales would predict office discipline referrals (ODRs) and suspensions (in and out of school), as well as predict lower academic achievement. We expected the Attention and Academic Issues subscale to predict lower scores on academic outcomes, but not necessarily disciplinary outcomes. Furthermore, it was hypothesized that higher scores on Internalizing Behaviors and School Disengagement subscales would be associated with lower attendance. Also, we expected higher Internalizing Behaviors, Externalizing Behaviors, and Emotion Dysregulation to be associated with reports of being bullied by others.

2. Method

2.1. Participants

Participants included 5227 students (51% male; 49% female) from 27 public schools in a Midwestern state. These students were enrolled in Grades 3 (n = 1675; 32%), 4 (n = 1770; 34%), and 5 (n = 1782; 34%). Sixty-eight percent of the students identified as White, 16% as Black, 5% as Asian, 4% as Latinx, and 7% as a different race or two or more races. Fifty-eight percent of the students were eligible for free or reduced price meal (FRM; a commonly used proxy for socioeconomic status) and 11% had an identified disability. A subset of students with parental consent and who provided student assent also completed the BASC-3 (n = 382). Compared to the overall sample, the subset of students completing the BASC-3 came from 22 schools, was 48% male, and had a greater proportion of White students (83%), a lower proportion of students eligible for FRM (40%), and were primarily 3rd and 4th grade students (87%). Eighty-four percent of the eligible BASC-3 subsample provided parent and student assent to participate.

2.2. Procedure

The University Institutional Review Board and the participating school districts approved the study protocol. The EIS-SR was administered in a web-based online format to students in Grades 3–5 from participating schools. Because the EIS-SR administration was universal (i.e., K–12, countywide), it qualified as a "regular educational practice" that did not require additional active consent for research (American Psychological Association, 2017, p. 11). However, a note was sent home to parents explaining the purpose of the EIS-SR and how the information would be used by schools with an option to opt-out their child. Fewer than 1% of students were opted-out by parents each year. Students who were opted-out completed other schoolwork either in or outside the classroom when the EIS-SR was administered. Students completed the EIS-SR in the fall (October) of the 2018–19 academic year. All EIS-SR data were gathered at this one time point across all schools. Schools were given flexibility about how and when to administer the surveys and were allowed to administer the EIS-SR in a manner that best fit their routines and existing schedules. Each item was read aloud to the students by a teacher or a school mental health practitioner. A script was used that provided students with information about the purpose of the EIS-SR and how the information would be used. A debriefing was provided at the end of the survey in which students were provided contact information if they wanted to discuss the survey with an adult in the building. On average, the students completed the measure in 8 min (range 3–15 min). Next, the subset of students with parental consent and who provided student assent completed the BASC-3 ($n = 382$). These students were from the classrooms of teachers who were interested in participating in the validation study. Students completed the BASC-3 within a two-week window of when the EIS-SR was administered. Each item was read to the student by a researcher. Students answered each question by circling their best answer. Students received a small token of appreciation for completing the assessment (e.g., a snack). On average the students completed the measure within 30 min.

2.3. *Measures*

2.3.1. *Student demographics*

Schools provided demographic information for students, including gender, FRM status, disability status, and race.

2.3.2. *Early Identification System-Student Support*

The Early Identification System-Student Report (EIS-SR) includes 42 questions with Likert-type response options (0 = *Never*, 1 = *Sometimes*, 2 = *Often*, 3 = *Always*). Seven factors were hypothesized based on our literature review, item development, and expert review and feedback: Externalizing Behavior, Internalizing Behavior, Peer Relationship Problems, School Disengagement, Emotional Dysregulation, Attention and Academic Issues, and Relational Aggression. Two of the items were identified as risk factors on their own (i.e., “Other kids make fun of me at school,” and “I am bullied by others”) and were not included in the factor analyses. An initial investigation with 1590 elementary and middle school students largely supported the psychometric properties of the EIS-SR with excellent model fit and strong scale reliabilities (Huang et al., 2019). We used EIS-SR raw scores in factor analyses and logistic regression analyses and standardized scores in linear regression analyses.

2.3.3. *Measure of concurrent validity*

The BASC-3 (Reynolds et al., 2015) Student Report of Personality (SRP) is a standardized measure developed for use in assessing problem behavior, social problems, and adaptive skills. All subscale scores are nationally normed, providing a solid reference to true risk for social behavioral and emotional problems. The subscales have a mean of 50 and standard deviation of 10, resulting in a T-score. T-scores less than 60 are in the normal range, scores between 60 and 69 indicate at risk, and scores greater than 69 are in the clinically significant range. The BASC-3 has strong reliability and validity (Reynolds et al., 2015). The manual reports coefficient alpha reliabilities for children ages 8–11 years ranging from 0.79 to 0.94 on the subscales used in the present study. For the purpose of this study, BASC-3 subscales that are similar to the EIS-SR subscales were used, including Attention Problems, Attitude to School, and Depression subscales, and the Emotional Symptoms Index, Inattention/Hyperactivity, Internalizing Problems, Personal Adjustment, and School Problems composite scores.

2.3.4. *Measures of predictive validity*

We used existing data to evaluate the predictive validity of the EIS-SR, including school disciplinary data, attendance, and academic outcomes.

School Disciplinary Data. For each student, schools provided the number of office discipline referrals[1] (ODRs), in-school suspensions (ISS) received, and out-of-school suspensions (OSS) received each month. For the purpose of this study, the total number of ODRs, ISS, and OSS were calculated based on those received across the months of January through the end of the school year. Students who received a sanction one or more times were coded as a 1 and those not receiving a sanction were coded as a 0.

Bullied. Students in the spring were asked “Other kids make fun of me at school” and “I am bullied by others” ($r = 0.64$). As bullying requires repetitive interactions (Solberg & Olweus, 2003), students who answered *Often* or *Always* to either question were considered victims (bullied = 1; 497 out of 4658 or 10.7%). These data were gathered as part of the Spring implementation of the EIS-SR.

Attendance Data. For each student, schools provided the percent of days in attendance for each month. For spring attendance data, we averaged the monthly attendance data from January to May.

Academic Outcomes. The Missouri Assessment Program (MAP) exam was used in this study. The MAP is a state-mandated grade-level assessment based on the Missouri state standards for all Missouri students in Grades 3–8. The MAP is administered at each of these grade levels within the content areas of communication arts and mathematics. Primary content areas measured in communication arts

[1] Although less severe than suspensions, ODRs may signal a risk of academic as well as behavioral problems (McIntosh et al., 2009).

include speaking/writing Standard English, reading fiction/poetry/drama, reading nonfiction, writing formally/informally, and combined reading. Primary content areas measured in math include number and operations, algebraic relationships, geometric and spatial relationships measurement, and data and probability. MAP scaled scores had acceptable Cronbach's alpha coefficients. Specifically, reliability of the communication arts test was 0.90 for Grade 3, 0.91 for Grade 4, and 0.91 for Grade 5, and the mathematics test produced reliability coefficients of 0.91 for Grade 3, 0.91 for Grade 4, and 0.90 for Grade 5 (Missouri Department of Elementary and Secondary Education, 2017).

2.4. Analytic strategy

The analysis was conducted over four distinct phases. Phase 1 aimed to establish the factor structure of the EIS-SR. Phase 2 investigated the degree of measurement invariance of the measures based on gender, grade level, and race. In Phase 3, the concurrent validity of the measures were evaluated. Phase 4 examined the predictive validity of the EIS-SR subscales by using the subscale

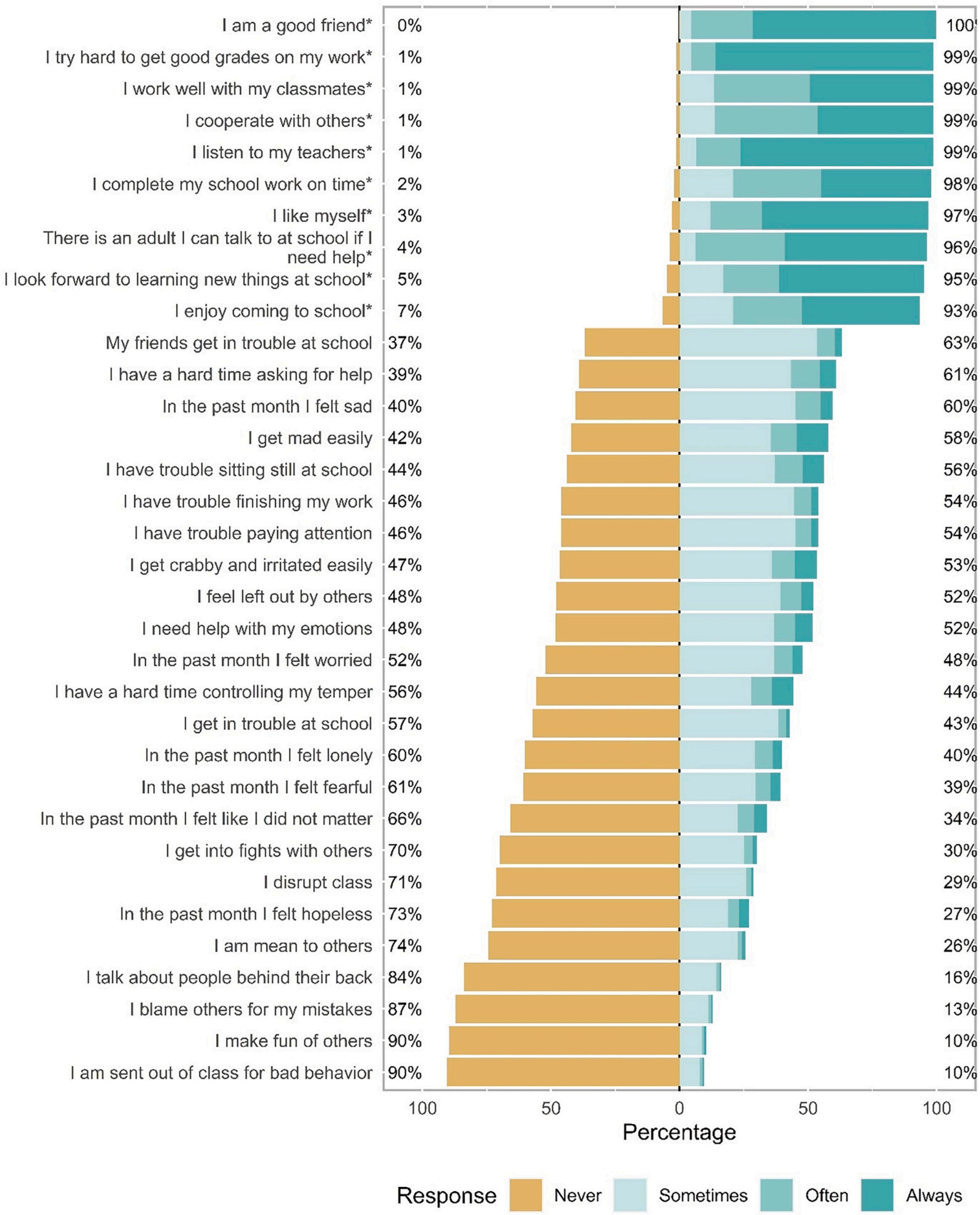

Fig. 1. Distribution of responses per question ($n = 5005$). *Questions are reverse coded when computing scales to reflect greater risk.

measures taken in the fall to predict spring student outcomes.

Data management and predictive validity analyses were conducted using R 3.5.3 (R Core Team, 2019) and factor analyses were performed using Mplus 6.1 (Muthén & Muthén, 2011). Missingness was not an issue as only complete surveys are downloaded from the online system. Prior to conducting any analyses, a validity check was performed and students who had reported that they had truthfully answered most or all the items that were included in the analysis (n = 5005; 96%). Validity checks have been shown to improve the quality of survey data (Cornell et al., 2012; Jia et al., 2018), especially in the context of mental health screening data (Furlong et al., 2017).

2.4.1. Confirmatory factor analyses

Guided by the knowledge of the items corresponding to the hypothesized factors, we conducted a confirmatory factor analysis (CFA) using two randomly split samples of valid participants ($n_1 = 2502$, $n_2 = 2503$). Although some may use exploratory factor analysis (EFA) prior to conducting a CFA, a CFA is appropriate when "the researcher has clear predictions about the number of common factors and the specific measure each common factor will influence" (Fabrigar & Wegener, 2012, p. 4) and "when the researcher has some knowledge of the underlying latent variable structure" (Byrne, 2012, p. 6). The CFA is also a more rigorous test as compared to an EFA and the use of the CFA was consistent with our goal of testing the hypothesized latent variable structure.

Although the instrument originally had 42 items, eight items were removed prior to analysis for various reasons, including two items that were deemed to be a risk factor on their own (e.g., "I am bullied by others"), one item that was questioned by respondents with regards to its social utility ("I have friends to eat lunch with at school"), and five items that were not included based on prior factor analytic work using a much smaller sample (e.g., poor loadings, presence of double loadings, redundancy; Huang et al., 2019). In the end, 34 items were examined and item distributions are shown in Fig. 1.

We were interested in the seven hypothesized risk subscales and tested for an overall risk indicator as well. To assess this, we tested a unidimensional factor model, a seven correlated-factor model, and a second-order factor model. We performed the confirmatory factor analyses using the first random split-half sample (training sample; $n = 2502$) and inspected fit indices, factor loadings, and factor correlations.

Items with low factor loadings (e.g., < 0.45) were subject to removal. For the correlated factor model, if factors had extremely high correlations (e.g., > 0.90) with one another, they were subject to further inspection as well (e.g., factors may be combined). As a specification check, we consulted modification indices, and items were evaluated based on their alignment with their theoretical postulates and the risk factors. Nested models were compared with a SBχ^2 (Satorra & Bentler, 2010) difference test to identify the best fitting model.

As a measure of model fit and given the large sample size, which leads to the over rejection of reasonably specified models using the χ^2 test (Anderson & Gerbing, 1988), several model fit indices were consulted. The fit indices included the root mean square error of approximation (RMSEA), the Tucker-Lewis index (TLI), and the comparative fit index (CFI). RMSEA values <0.08 were considered reasonable (Kline, 2011) and values for the CFI and TLI > 0.90 suggested acceptable model fit (Hu & Bentler, 1995).

Given the categorical nature of the responses, polychoric correlation matrices were used with weighted least squares with mean and variance correction (WLSMV) estimation (Finney & DiStefano, 2006). To account for the clustered nature of the data, we used the type = complex option in Mplus, which adjusted standard errors and model fit indices to account for the nesting of observations within schools (Stapleton, 2006). To aid in model generalizability and to avoid issues of overfitting, the best fitting model would then be replicated using a separate confirmatory, hold out sample ($n = 2503$). Given the limitations of alpha (McNeish, 2018), scale score reliability was estimated using categorical omega using the MBESS package (Kelley & Pornprasertmanit, 2016).

2.4.2. Measurement invariance

To assess whether the factor structures of the subscales were invariant across groups, we tested for measurement invariance using multi-group confirmatory factor analysis (MG-CFA) based on gender (male and female), grade level (3, 4, and 5), and race (Black and White students). Measurement invariance indicates that the instrument is measuring the same construct in a similar manner across different groups (Dimitrov, 2010). Although group comparisons using gender and grade level used the entire sample, when unbalanced group sizes are used for invariance testing, such as with the race variable with our data, invariance may be erroneously reported (Yoon & Lai, 2018). Because of unequal group sizes, we compared the reports of Black students ($n = 791$) with the reports of a random sample of White students ($n = 791$). The relatively smaller numbers of the other groups based on race (e.g., 181 valid Latinx respondents) precluded us from performing categorical invariance testing for Asian and Latinx students; for instance, given the small sample size for these subgroups, no respondents endorsed some response options on various items (e.g., no Asian student endorsed *never* for the question "I work well with my classmates").

Measurement invariance was conducted using three successively more restrictive models where configural, metric, and scalar/threshold invariance were investigated (Bowen & Masa, 2015; Dimitrov, 2010). Configural invariance (CI) tests whether the factors are measured by the same indicators across the specified groups (Bowen & Masa, 2015). Metric invariance (weak invariance) constrains the factor loadings to be equal among the groups to indicate that the factors are measured in the same manner. Finally, scalar invariance (or strong invariance) imposes an additional restriction where item thresholds (due to the categorical nature of the data) are held equal among groups. Although a fourth test of residual invariance is possible, the methodological literature is unclear about its necessity (Pendergast et al., 2017). We recognize as well that there are other procedures in determining measurement invariance for categorical data such as beginning with configural invariance, testing for threshold invariance, and finally testing for factor loading invariance (Svetina et al., 2019; Wu & Estabrook, 2016). As a robustness check, we also performed these steps and the results were unchanged.

To assess whether invariance or measurement equivalence was tenable using the succeeding more restrictive models (i.e., metric vs. configural models, scalar vs. metric models), several measures were consulted. Although the $SB\chi^2$ difference test is often used as one test, the test is sensitive to sample size as well (Cheung & Rensvold, 2002). A more commonly consulted measure for invariance testing is the ΔCFI, where a decrease of CFI (i.e., a worsening of model fit) in the more restrictive model by more than 0.01 suggests non-invariance (Cheung & Rensvold, 2002). In addition, Chen (2007) suggested a ΔRMSEA of <0.015 to signify measurement invariance. Improvement of model fit between models is also evidence supporting measurement invariance (Dimitrov, 2010).

2.4.3. Concurrent validity

As a measure of concurrent validity, we correlated subscale scores of the EIS-SR with selected subscales (i.e., Attention Problems, Attitude to School, Depression) and composite t-scores (i.e., Emotional Symptoms Index, Inattention/Hyperactivity, Internalizing Problems, Personal Adjustment, and School Problems) of the BASC-3 student report. Only certain BASC-3 subscales were used as not all the EIS-SR subscales (i.e., Peer Relationship Problems, Externalizing Behaviors, and Relational Aggression) had an analogous or similar score on the BASC-3. As has been adopted by others (e.g., Kilgus et al., 2013; Sun et al., 2011), we considered correlations of approximately 0.50 to be large (Cohen, 1988).

In addition, the BASC refers to respondents with T-scores two *SD* above the mean (i.e., 70 +) as clinically significant, which suggests a high level of maladjustment warranting follow-ups with the individual. Using the BASC T score threshold, we created a clinically significant and non-clinically significant group and investigated how well the corresponding EIS-SR subscales could predict clinically significant status. We conducted receiver operating classification (ROC) curve analysis (Swets et al., 2000) and computed the area under the curve (AUC) statistic, which combines measures of sensitivity and specificity. Based on Hosmer and Lemeshow's (2004, p. 162) guidelines, we considered scores between 0.70 and 0.80 to be acceptable discrimination, between 0.80 and 0.90 to be excellent, and scores at or above 0.90 to be outstanding discrimination. Although AUC may be used for diagnostic accuracy and establishing cut scores, we used AUC as a measure of model fit and classification accuracy. Classification accuracy is an essential characteristic of screening tools in order to categorize students at risk or not at risk (Stanley et al., 2019).

2.4.4. Predictive validity

To assess the predictive validity of the EIS-SR subscales, the fall EIS-SR measures were used to predict spring behavioral (i.e., receipt of ODR, ISS, OSS, bully victimization in the spring), academic (i.e., reading and math MAP scores), and attendance outcomes. Based on the dichotomous nature of the behavioral outcomes, logistic regression was used. As a measure of model fit, we also included AUC scores. For the academic outcomes, linear regression models were used and a small number of outliers (e.g., higher or lower than 3 *SD* from the mean or ~ 1% of the sample) were excluded from the analysis. The predictors of interest were the EIS-SR subscales. Student (i.e., disability status, gender, grade level, race, and FRM status) and school-level characteristics were controlled for. School fixed effects together with cluster robust standard errors were used to account for students nested within schools and to control for both observed and unobserved school characteristics (Huang, 2016). For the continuous outcomes, the dependent variables and the EIS-SR subscales were standardized ($M = 0$, $SD = 1$). As a result of using fixed effects models, only schools that had variability in the outcomes were included in the analyses (e.g., schools that did not issue an ISS, OSS, or ODR were excluded from the analyses regardless of student characteristic). Due to the exclusions, the sample size for the behavioral outcomes ranged from 3899 (e.g., for OSS, several schools did not issue an OSS in the spring) to 4938. For the academic outcomes, the sample size ranged from 4736 to 4780.

3. Results

3.1. Factor analysis

Using the training sample ($n = 2502$), a one factor model was tested. Model fit, however, was poor, indicating the multidimensional nature of the constructs investigated, TLI = 0.82, CFI = 0.81 (see Table 1). The hypothesized seven correlated factor model was then tested and all items had appreciable factor loadings and showed acceptable model fit, RMSEA = 0.03, TLI = 0.94, CFI = 0.94. However, upon inspection of model fit indices, one item (i.e., "I have friends to talk to at school") was flagged as possibly loading with several other factors. To avoid issues of cross loading and for parsimony, we excluded the item and re-estimated the model. The modified model showed better model fit, RMSEA = 0.03, CFI = 0.95, TLI = 0.95 based on all fit indices. Standardized factor loadings are shown in Table 2. The factor correlations (see Table 3) ranged from 0.17 (School Engagement with Internalizing Behaviors) to 0.57 (Externalizing Behaviors with Relational Aggression).

A second-order factor model was estimated with a general risk factor giving effects to all subscales, RMSEA = 0.03, TLI = 0.94, CFI = 0.93.[2] The general risk factor loaded most strongly on Externalizing Behaviors, Relational Aggression, and Peer Relationship Problems (see Table 4). However, a χ^2 difference test indicated that the seven correlated factor model fit the data better, $\chi^2(14) = 536$, $p < .001$, and was the preferred factor structure.

Using the confirmatory hold-out sample ($n = 2503$), the seven correlated factor model was replicated. Model fit indices again indicated that the model fit the data well, RMSEA = 0.03, TLI = 0.96, CFI = 0.95. For completeness, the second order model fit is

[2] Another alternative model specification was a bifactor model. Although bifactor models have grown in popularity, bifactor models have been shown to fit well even with the use of random data (Bonifay & Cai, 2017) which raises questions about their utility. Others have also raised similar conerns (e.g., Reise et al., 2016).

Table 1
Model fit indices of competing models using the training and confirmatory samples.

	χ2	DF	RMSEA	TLI	CFI
Training sample (n = 2502 in 27 schools)					
1 factor	4723	560	0.055	0.820	0.809
7 correlated factor (hypothesized)	1981	539	0.033	0.938	0.931
7 correlated factor (modified)	1608	506	0.030	0.953	0.947
Second order factor model	2006	520	0.034	0.936	0.931
Confirmatory sample (n = 2503 in 27 schools)					
7 correlated factor (modified)	1478	506	0.028	0.956	0.951
Second order factor model	1759	520	0.031	0.944	0.939

Note. All χ2 were statistically significant.

shown as well. Although categorical omega (see Table 2) indicated that the scales exhibited acceptable reliability, the reliabilities of the Relational Aggression scale was somewhat lower than expected ($\omega_{exp} = 0.69$; $\omega_{con} = 0.65$).

3.2. Measurement invariance

Using the entire sample (n = 5005), we first investigated measurement invariance based on gender (males = 2539, females = 2466). In the first step (see Table 5 for all fit indices), configural invariance was established, RMSEA = 0.03, CFI = 0.96, TLI = 0.95. Next, metric invariance, which fixed the factor loadings as equal between groups, was tested. Resulting model fit showed an overall improvement in fit (RMSEA = 0.03, CFI = 0.96, TLI = 0.96), supporting metric invariance. Finally, scalar/threshold invariance was tested where thresholds were constrained to be equal among groups in addition to having equal factor loadings. The model also had adequate fit and although the $SB\chi^2$ was statistically significant, $SB\chi^2(95) = 238$, $p < .001$, improvement in the other fit indices indicated that scalar/threshold invariance was supported, RMSEA = 0.02, CFI = 0.96, TLI = 0.96,

Grade level invariance (Grade 3 = 1580; Grade 4 = 1695; Grade 5 = 1730) was then tested and configural invariance was established, RMSEA = 0.03, CFI = 0.96, TLI = 0.95. Metric invariance was tested and improvement in model fit (RMSEA = 0.03, CFI = 0.97, TLI = 0.96) supported metric invariance as well. Finally, scalar/threshold invariance was tested and the model also had adequate fit, RMSEA = 0.03, CFI = 0.96, TLI = 0.96. Although in this model the $SB\chi^2$ was statistically significant, $SB\chi^2(190) = 683$, $p < .001$, and the CFI decreased by 0.002, metric invariance was supported based on the guidelines provided (Chen, 2007; Cheung & Rensvold, 2002; Dimitrov, 2010).

For invariance based on race (Black = 791, White = 791), configural invariance was supported, RMSEA = 0.03, CFI = 0.96, TLI = 0.96. Metric invariance was also shown by a non-statistically significant $SB\chi^2(27) = 26.2$, $p = .51$. Scalar invariance was also tested and supported as well, RMSEA = 0.03, CFI = 0.96, TLI = 0.96. In summary, based on MG-CFA analyses, the EIS-SR was shown to be invariant based on gender, grade level (i.e., with students from Grades 3 to 5), and race (i.e., between Black and White students).

3.3. Concurrent validity

Descriptively, the hypothesized EIS-SR subscales and their corresponding BASC-3 counterparts were moderately correlated with each other (see Table 6). The EIS-Internalizing Behaviors subscale correlated the most with the BASC Emotional Symptoms Index (r = 0.58), Internalizing Problems (r = 0.58), and Depression (r = 0.57). The EIS-Attention and Academic Issues subscale was correlated with both the Attention Problems and the Inattention/Hyperactivity subscales (rs = 0.61 and 0.60, respectively). The EIS-Emotional Dysregulation subscale correlated with the BASC Emotional Symptoms Index and Internalizing Problems (both rs = 0.51). Finally, the EIS-School Disengagement subscale correlated the most with the BASC-3 Attitude to School and School Problems subscales (rs = 0.56 and 0.54, respectively). All of the EIS-SR subscales were negatively correlated with the BASC Personal Adjustment subscale. In terms of the classification accuracy of the EIS-SR predicting clinically significant outcomes using the BASC T-scores, the EIS-SR measures had acceptable to excellent discrimination capabilities (AUCs = 0.75–0.82). Although establishing cut-scores was not part of the current study—as this would not apply to all the subscales and would require a more representative sample—the percent of students correctly classified to be at risk using the BASC ranged from 71% (for internalizing behavior) to 84% (for school disengagement). Using the National Center for Intensive Interventions (https://intensiveintervention.org/) classification accuracy guidelines and associated behavioral screening tools rubric, three EIS-SR subscales would appear to be eligible for a rating of "convincing evidence" given AUCs ≥0.75 and sensitivity and specificity measures ≥0.70.[3] The School Disengagement subscale would have a "partially convincing evidence" rating as a result of having a lower sensitivity of 0.66.

3.4. Predictive validity

To assess predictive validity, we examined interrelations among EIS-SR subscales and behavioral (e.g., disciplinary infractions),

[3] https://intensiveintervention.org/sites/default/files/NCII_BScreening_RatingRubric_2019.pdf

Table 2
Standardized factor loadings using confirmatory factor analyses from the training ($n = 2502$) and confirmatory samples ($n = 2503$).

Item	Question	PR	EX	IN	AT	ED	RA	SD
4	I am a good friend*	0.73 (0.75)						
5	I cooperate with others*	0.74 (0.74)						
6	I work well with my classmates*	0.74 (0.73)						
18	I get in trouble at school		0.78 (0.76)					
19	I am sent out of class for bad behavior		0.76 (0.78)					
20	I disrupt class		0.75 (0.76)					
21	I get into fights with others		0.72 (0.72)					
22	My friends get in trouble at school		0.52 (0.50)					
31	I listen to my teachers*		0.81 (0.77)					
32	I blame others for my mistakes		0.70 (0.69)					
7	I have a hard time asking for help			0.56 (0.54)				
9	I like myself*			0.66 (0.65)				
12	In the past month, I felt sad			0.70 (0.72)				
13	In the past month, I felt fearful			0.62 (0.71)				
14	In the past month, I felt lonely			0.76 (0.76)				
15	In the past month, I felt worried			0.66 (0.69)				
16	In the past month, I felt like I did not matter			0.85 (0.83)				
17	In the past month, I felt hopeless			0.80 (0.83)				
34	I feel left out by others			0.76 (0.76)				
23	I have trouble sitting still at school				0.70 (0.66)			
24	I have trouble finishing my work				0.69 (0.66)			
25	I have trouble paying attention				0.82 (0.80)			
29	I try hard to get good grades on my work*				0.57 (0.62)			
40	I complete my school work on time*				0.60 (0.57)			
26	I get mad easily					0.87 (0.87)		
27	I have a hard time controlling my temper					0.90 (0.89)		
33	I get crabby and irritated easily					0.81 (0.83)		
39	I need help with my emotions					0.76 (0.73)		
10	I am mean to others						0.87 (0.80)	
35	I talk about people behind their back						0.78 (0.76)	
36	I make fun of others						0.80 (0.80)	
28	I look forward to learning new things at school*							0.76 (0.77)
37	I enjoy coming to school*							0.80 (0.82)
42	There is an adult I can talk to at school if I need help*							0.57 (0.55)

(continued on next page)

Table 2 (*continued*)

Item	Question	PR	EX	IN	AT	ED	RA	SD
	Reliabilities (Omega)	0.71 (0.69)	0.86 (0.87)	0.78 (0.77)	0.74 (0.75)	0.85 (0.85)	0.69 (0.65)	0.70 (0.70)
	Intraclass correlation coefficient (ICC)	0.055	0.068	0.018	0.043	0.062	0.058	0.018

Note. Confirmatory hold-out sample factor loadings within parentheses. PR = peer relationship problems. EX = externalizing behavior. IN = internalizing behavior. AT = attention and academic issues. ED = emotional dysregulation. RA = relational aggression. SD = school disengagement.

* Reverse coded.

Table 3
Factor correlations using training sample (Lower Diagonal; n = 2502) and confirmatory samples (Upper Diagonal; *n* = 2503).

	PR	IN	EX	AT	RA	ED	SD
Peer relationship problems (PR)		0.17	0.41	0.31	0.49	0.36	0.32
Internalizing behavior (IN)	0.21		0.17	0.19	0.17	0.28	0.15
Externalizing behavior (EX)	0.46	0.19		0.39	0.52	0.44	0.29
Attention & Academic issues (AT)	0.37	0.22	0.45		0.31	0.37	0.29
Relational aggression (RA)	0.48	0.18	0.57	0.36		0.42	0.27
Emotional dysregulation (ED)	0.40	0.31	0.47	0.41	0.46		0.26
School disengagement (SD)	0.35	0.17	0.31	0.32	0.29	0.26	

Note. All correlations are statistically significant (*ps* < 0.001).

Table 4
Second-order model (general risk) standardized factor loading on subscales.

	Training sample (n = 2502)	Confirmatory sample (n = 2503)
Peer relationship problems (PR)	0.86	0.77
Internalizing behavior (IN)	0.63	0.60
Externalizing behavior (EX)	0.89	0.87
Attention & Academic issues (AT)	0.78	0.81
Relational aggression (RA)	0.87	0.86
Emotional dysregulation (ED)	0.83	0.79
School disengagement (SD)	0.60	0.61

Table 5
Fit indices for invariance testing based on gender, grade level, and race.

		$\chi 2$	DF	RMSEA	CFI	TLI
I. Gender (males = 2539; females = 2466)						
A	Configural invariance	2849	1012	0.027	0.958	0.954
B	Metric invariance	2671	1039	0.025	0.963	0.960
C	Scalar/threshold invariance	2774	1134	0.024	0.963	0.963
II. Grade level (Grade 3 = 1580; Grade 4 = 1695; Grade 5 = 1730)						
A	Configural invariance	3460	1518	0.028	0.959	0.954
B	Metric invariance	3232	1572	0.025	0.965	0.962
C	Scalar/threshold invariance	3535	1762	0.025	0.963	0.964
III. Race (White = 791; Black = 791)						
A	Configural invariance	1680	1012	0.029	0.959	0.955
B	Metric invariance	1619	1039	0.027	0.964	0.962
C	Scalar/threshold invariance	1766	1134	0.027	0.961	0.962

Note. All Satorra-Bentler $\Delta\chi^2$ were statistically significant (*ps* < 0.01) except for the difference between IIIA and IIIB; $\Delta\chi^2(27) = 26.23$, $p = .51$. Based on ΔRMSEA, ΔCFI, ΔTLI, measurement invariance was supported.

bullying events, academic (i.e., communication arts and mathematics), and attendance outcomes. Descriptive statistics for the scales are shown in Table 7. For all measures, the relationship of the scales was evaluated over and above the contribution of school characteristics (controlled through fixed effect models) and student demographic variables (i.e., gender, race, FRM status, disability status). For the behavioral and being bullied outcomes (see Table 8), which were coded as a 1 or a 0 (i.e., 1 = receipt of an ODR, ISS, OSS, or bullied), the most consistent predictor was that of Externalizing Behaviors (*ORs* = 1.08–1.18, all *ps* < 0.05), which indicated that students who acted out more were more likely to receive an ODR, suspension, and experience peer bully victimization. Higher levels of Emotional Dysregulation were also associated with higher odds of receiving an ODR (*OR* = 1.09, *p* < .001) and being bullied

Table 6
Correlation and area under the curve (AUC) statistics between the EIS-SR and BASC-3 subscales (n = 382).

	EIS Subscales						
BASC Subscales	PR	IN	EX	AT	RA	ED	SD
Attention Problems	0.38	0.32	0.47	**0.61**	0.37	0.43	0.34
Attitude to School	0.21	0.24	0.30	0.39	0.24	0.37	**0.56**
Depression	0.23	0.57	0.31	0.43	0.29	0.43	0.19
Emotional Symptoms Index	0.30	**0.58**	0.37	0.53	0.32	**0.51**	0.29
Inattention/Hyperactivity	0.36	0.36	0.48	0.60	0.37	0.46	0.32
Internalizing Problems	0.27	**0.58**	0.40	0.50	0.33	**0.51**	0.24
Personal Adjustment	−0.32	−0.51	−0.32	−0.49	−0.27	−0.43	−0.32
School Problems	0.26	0.27	0.40	0.44	0.30	0.37	0.54
AUC	na	0.82	na	0.79	na	0.80	0.75

Note. All correlations are statistically significant. Peer relationship problems (PR). Internalizing behavior (IN). Externalizing behavior (EX). Attention & Academic issues (AT). Relational aggression (RA). Emotional dysregulation (ED). School disengagement (SD). The highest correlations between the EIS and the hypothesized BASC constructs are highlighted in **BOLD**. [na] EIS subscales of PR, EX, and RA do not have an analogous match on the BASC. AUC scores use both the internalizing subscales on the BASC and EIS-SR; both attention related subscales on the BASC and EIS-SR; the Emotional Symptoms Index on the BASC and the Emotional Dysregulation subscale on the EIS-SR; and the Attitude to School scale on the BASC and the School Disengagement subscale on the EIS-SR.

($OR = 1.07, p < .01$). Lastly, students with higher levels of Internalizing Problems also were more likely to report that they were bullied (OR = 1.14, p < .001). Internalizing Problems was associated with lower odds of receiving an OSS (OR = 0.93, p < .01).

For the attendance and academic outcomes (see Table 9), the only EIS-SR subscale that predicted attendance was School Disengagement ($\beta = -0.04$, $p < .05$), with more disengaged students attending school less. Higher scores on the Attention and Academic Issues and Externalizing Behavior subscales were related to lower scores on MAP achievement outcomes (βs = −0.11 – −0.15, ps < 0.001). Unexpectantly, students with higher levels of Peer Relationship Problems (βs = 0.07–0.10, ps < 0.001) and School Disengagement scored higher on the MAP subtests (β = 0.03, p < .05 for Communication Arts). Finally, Relational Aggression, Internalizing Problems, and Emotion Dysregulation were not predictive for either communication arts or mathematics achievement outcomes.

4. Discussion

Universal social, emotional, and behavioral screening is a highly recommended practice for schools to use in identifying students needing supports and for prevention efforts (Briesch et al., 2018; Dowdy et al., 2010). The purpose of this study was to evaluate the factor structure, measurement invariance, and the concurrent and predictive validity of the EIS-SR. As expected, seven subscales were identified as having adequate factor loadings. One item (i.e., "I have friends to talk to at school") that we predicted would load on the Peer Relationship Problems subscale was removed due to possibly cross-loading with other factors. This item was very similar to the items of "In the past month, I felt lonely" and "I feel left out by others", both of which loaded onto the Internalizing Behaviors subscale. Thus, the removal of the item would appear to be justified. In addition, all the subscales except for the Relational Aggression subscale met the internal consistency reliability required for a screener (Cortina, 1993; Nunnally, 1978). Anecdotally, our school partners have indicated that some students who have higher scores on the Relational Aggression subscale and who do not have elevated risk on other subscales are often considered to be more conscientious by school personnel, suggesting the possibility that these students may be diligently reporting even rare instances of being unkind to others as compared to peers who have a higher threshold for occasional lapses in kindness. Future EIS-SR research will consider revising items for this age-group and/or the addition of other items to improve the internal consistency of this subscale.

In addition to evaluating the factor structure and internal consistency of the EIS-SR, the invariance of the subscales across gender, among students from Grades 3 to 5, and between Black and White students was evaluated. Findings indicated that the instrument measures the same construct in a similar manner across these specific groups. It is important to ensure that the measure is operating in the same way and that the underlying construct has the same theoretical structure for each group. Having invariance allows for the generalization of the interpretation of the findings across groups, meaning that the subscale scores for students in Grades 3–5 can be

Table 7
EIS-SR scale descriptive statistics (n = 5005).

Scale	M	SD	median	min	max	skew	kurtosis
Peer relationship problems	1.72	1.62	2	0	9	0.85	0.27
Internalizing behavior	5.45	4.85	4	0	27	1.30	1.78
Externalizing behavior	2.55	2.61	2	0	21	1.79	4.82
Attention & Academic issues	3.19	2.52	3	0	15	0.83	0.49
Relational aggression	0.61	1.10	0	0	9	2.38	7.40
Emotional dysregulation	3.14	3.09	2	0	12	1.09	0.51
School disengagement	2.17	2.02	2	0	9	0.85	0.11

Table 8
Logistic regression results (in odds ratios) Predicting Receipt of an Office Disciplinary Referral (ODR), In-School Suspension (ISS), and Out-of-School Suspension (OSS).

	ODR (n = 4938)	ISS (n = 4346)	OSS (n = 3899)	Bullied (n = 4670)
EIS subscales				
Peer relationship problems	1.05 (1.00–1.12)	1.07 (0.97–1.17)	1.07 (0.96–1.2)	1.05 (0.99–1.12)
Internalizing behaviors	0.99 (0.96–1.01)	1.02 (0.98–1.06)	0.93** (0.88–0.97)	1.14*** (1.12–1.16)
Externalizing behaviors	1.18*** (1.12–1.24)	1.11** (1.04–1.19)	1.12* (1.02–1.23)	1.08** (1.02–1.14)
Attention & Academic issues	1.04 (0.99–1.09)	1.04 (0.96–1.12)	1.08 (0.97–1.22)	0.99 (0.93–1.05)
Relational aggression	0.96 (0.89–1.04)	1.06 (0.92–1.22)	1.10 (0.93–1.30)	0.96 (0.88–1.06)
Emotional dysregulation	1.09 *** (1.05–1.13)	1.06 (0.99–1.14)	1.05 (0.962–1.14)	1.07** (1.02–1.12)
School disengagement	1.00 (0.96–1.05)	1.01 (0.94–1.09)	1.12 (0.98–1.27)	0.96 (0.91–1.01)
Student demographics				
Male	2.55 *** (1.93–3.38)	2.77 ** (1.5–5.12)	3.29 *** (1.92–5.62)	0.94 (0.75–1.19)
With a disability	1.58 *** (1.23–2.04)	1.18 (0.65–2.16)	3.02 *** (1.65–5.54)	1.94 *** (1.55–2.42)
Eligible for FRM	1.89 *** (1.48–2.42)	1.68 * (1.05–2.70)	4.00 *** (1.93–8.31)	1.18 (0.95–1.45)
Race: Other[1]	1.42 * (1.05–1.93)	1.62 (0.84–3.1)	0.58 (0.22–1.55)	0.96 (0.66–1.38)
Race: Asian[1]	0.25 *** (0.13–0.48)	0.35 (0.04–2.85)	†	0.91 (0.52–1.59)
Race: Black[1]	1.81 *** (1.39–2.36)	1.38 (0.80–2.38)	0.97 (0.57–1.66)	1.05 (0.82–1.36)
Race: Latinx[1]	0.79 (0.38–1.63)	1.08 (0.45–2.61)	1.39 (0.62–3.11)	0.63 (0.39–1.01)
Grade 4[2]	1.17 (0.84–1.65)	0.88 (0.58–1.33)	0.66 (0.27–1.60)	0.82 (0.64–1.03)
Grade 5[2]	1.38 * (1.05–1.83)	1.13 (0.68–1.89)	1.09 (0.56–2.13)	0.68 ** (0.51–0.91)
Area Under the Curve	0.81	0.84	0.90	0.80

Note. Models include school fixed effects. Cluster robust standard errors used. † = not applicable; no Asian students received an OSS. FRM = free or reduced price meals.
* $p < .05$,
** $p < .01$,
*** $p < .001$.
[1] White is the reference group.
[2] Grade 3 is the reference group.

Table 9
Linear regression results (cluster robust standard errors in parenthesis) predicting academic outcomes.

	Attendance (n = 4780)	MAP: Comm (n = 4772)	MAP: Math (n = 4736)
EIS subscales			
Peer relationship problems	0.03 (0.02)	0.07*** (0.02)	0.10*** (0.01)
Relational aggression	0.00 (0.02)	−0.02 (0.01)	−0.02 (0.02)
Internalizing behaviors	0.01 (0.02)	−0.05*** (0.01)	−0.03* (0.01)
Externalizing behaviors	−0.02 (0.03)	−0.06*** (0.02)	−0.06*** (0.02)
Attention & Academic issues	−0.01 (0.03)	−0.11*** (0.01)	−0.15*** (0.02)
Emotional dysregulation	−0.03 (0.02)	−0.00 (0.01)	−0.00 (0.01)
School disengagement	−0.04* (0.02)	0.03* (0.01)	0.03 (0.02)
Student demographics			
Male	0.06 (0.03)	−0.06** (0.02)	0.15*** (0.02)
With a disability	−0.27*** (0.03)	−0.42*** (0.03)	−0.38*** (0.03)
Eligible for FRM	−0.09 (0.05)	−0.63*** (0.04)	−0.64*** (0.04)
Race: Other[1]	0.03 (0.06)	−0.14* (0.06)	−0.15** (0.05)
Race: Asian[1]	0.19 (0.12)	0.01 (0.06)	0.16** (0.05)
Race: Black[1]	−0.05 (0.05)	−0.55*** (0.04)	−0.50*** (0.03)
Race: Latinx[1]	−0.12 (0.11)	−0.26*** (0.06)	−0.15* (0.08)
Grade 4[2]	0.03 (0.04)	0.49*** (0.04)	0.62*** (0.05)
Grade 5[2]	0.01 (0.04)	0.80*** (0.03)	1.09*** (0.04)
R^2	0.05	0.40	0.50

Note. All outcomes and EIS scores standardized (M = 0, SD = 1). Models include school fixed effects. COMM = communication arts. FRM = free or reduced price meals.
* $p < .05$,
** $p < .01$,
*** $p < .001$.
[1] White is the reference group.
[2] Grade 3 is the reference group.

inferred in the same manner across all grades. Having screening measures that can identify risk in the same manner across gender, grade level, and racial identity is helpful when schools use these data to determine the types of supports to put into place for students.

Furthermore, the concurrent validity of the EIS-SR was examined. It is important to know if the EIS-SR subscales are correlated with similar measures with robust psychometric properties, as this supports that the measure is measuring what is expected. The concurrent validity was examined for the EIS-SR Internalizing Behavior, Attention and Academic Issues, Emotion Dysregulation, and School Disengagement subscales using the BASC-3 SRP, which is a widely used assessment of behavioral and emotional symptoms. Each of these subscales was moderately correlated with similar subscales on the BASC-3. Specifically, the Emotional Dysregulation subscale was moderately correlated with both the Emotional Symptoms Index and Inattention/Hyperactivity subscales of the BASC-3, indicating that students with elevated risk on the EIS-SR subscale were reporting trouble regulating both their emotions and behaviors. Unfortunately, analogous subscales for Peer Relationship Problems, Externalizing Behaviors, and Relational Aggression were not available with the student report of the BASC-3. However, all EIS-SR subscales were negatively correlated with the BASC-3 Personal Adjustment scale in which higher scores on this subscale indicate better interpersonal relationships, higher self-esteem, and self-reliance.

The range of correlations between EIS-SR subscales and comparable BASC-3 subscales ($rs = 0.51–0.61$) is similar to the concurrent correlations reported for the BESS student-report total score and longer measures of externalizing and internalizing symptoms ($rs = 0.51–0.77$; Kamphaus & Reynolds, 2015). Moreover, the range of AUC statistics (0.75–0.82) for specific EIS-SR subscales identifying youth above clinical cut-offs on individual BASC subscales were comparable or higher than those reported for individual subscales (0.68–0.74) and total scores (0.82) on the SAEBRS (Kilgus & von der Embse, 2014) in predicting elevated total scores on the Strength and Difficulties Questionnaire (Goodman, 1997; von der Embse et al., 2017). These findings support the individual specificity of several of the EIS-SR subscales in predicting specific outcomes rather than simply predicting global risk as is more common for screeners. Future research on the EIS-SR should examine the concurrent validity and specificity of all subscales using established measures.

Although having adequate internal consistency, invariance across groups, and concurrent validity are important for determining the utility of a measure, demonstrating predictive validity of scores on the measure on important external criterion bolsters construct validity claims (Lord et al., 1968). In other words, are the constructs being assessed useful in predicting socially meaningful outcomes in the way theorized? To determine the predictive validity, we examined the relations among the EIS-SR subscales and future school-related behavioral and academic outcomes controlling for school- and student-level correlates of these outcomes. Most of the findings were as hypothesized. First, as expected, higher scores on Externalizing Behaviors were associated with increased risk for receiving ODRs, ISS, and OSS. Higher scores on the Emotion Dysregulation subscale also predicted an increased risk of receiving ODRs. Peer Relationship Problems, School Disengagement, and Relational Aggression subscales were not predictive of school disciplinary actions. Relationally aggressive behaviors, such as making fun of others or talking about someone behind their back, are more covert in nature, and teachers and other school personnel may be less aware of these behaviors, making it less likely these students would receive disciplinary actions. Also, students with peer relationship problems or who are disengaged from school simply may not display behaviors that are considered troublesome enough to warrant an ODR, ISS, or OSS. Thus, it is possible that some students receiving higher scores for School Disengagement are more inhibited and do not exhibit externalizing behaviors, which are the most likely behaviors to lead to exclusionary discipline.

Regarding students who were more likely to report being bullied by others in the spring, as hypothesized, the Internalizing Behavior, Emotion Dysregulation, and Externalizing Behaviors subscales were predictive of peer bully victimization. The finding that students with emotion regulation problems and externalizing behaviors are bullied may seem contrary, but students who report problems controlling their temper and being easily irritated display behaviors that often lead to peer rejection (Dishion et al., 2010) and in some instances bully victimization (Brendgen, et al., 2015). Many of these students may also engage in bully behaviors given that half of the students who bully others report that they have been bullied (Haynie et al., 2001). Furthermore, given that being a victim of bullying has been directly tied to the development of internalizing problems among young children (Arseneault et al., 2008), the association between the Internalizing Behaviors subscale and report of being bullied are not surprising. The EIS-SR asks students to indicate if they have been bullied; using this information in conjunction with the Internalizing Behaviors subscale holds great utility in intervening in bully behaviors and supporting students with internalizing problems.

As predicted, higher scores on the Attention Academic Issues subscale were predictive of lower scores on both math and reading achievement. In addition, higher Externalizing Behavior subscales scores were related to lower math and reading achievement. This is a finding consistent with past research (Odgers et al., 2008; Temcheff et al., 2008). However, some surprising relations also were found. Higher scores on the Peer Relationship Problems and School Disengagement predicted higher reading achievement scores. For instance, students with higher levels of disengagement demonstrated high achievement after controlling for all other subscales and student demographics. One possibility for this finding could be that the School Disengagement subscale is tapping into students who are not academically challenged by their instructional level and are thus not looking forward to learning new things at school. The unexpected positive association between Peer Relationship Problems and achievement was also further examined. Regression diagnostics involved examining residuals and obtaining variance inflation factors as indicators of multicollinearity, but no issues were detected. An initial hypothesis was that students with higher levels of achievement may have greater issues with getting along with peers. However, univariate analyses and entering Peer Relationship Problems in a multiple regression (without the other EIS subscales but with demographic information) predicting achievement showed a negative relationship as expected. The correlation for Peer Relationship Problems and either type of academic outcome (i.e., communication arts and mathematics) was approximately $r = -0.10$, $p < .001$. The different subscales were also entered one at a time after Peer Relationship Problems and once either Externalizing Behaviors or Attention Issues were entered, the coefficient for Peer Relationship Problems became positive suggesting a suppressor effect (Thompson & Levine, 1997). In any case, this deserves further examination.

Lastly, it was hypothesized that higher scores on Internalizing Behaviors and School Disengagement would be associated with having poorer attendance. Prior research has found that internalizing problems are associated with absenteeism (Finning et al., 2019).

Although all relations were in the correct direction, only School Disengagement predicted poorer attendance. Students who are disengaged from school and do not enjoy coming to school are more likely to be absent. In fact, lower attendance is a risk factor for school disengagement and school dropout (see Christle et al., 2007). School disengagement is an early risk indicator of academic and social failure, thus screening for and intervening in the early signs of disengagement in elementary school can help prevent subsequent failure (Chase et al., 2014; Li & Lerner, 2012).

4.1. Implications

Universal social, emotional, and behavioral screening is an important practice in school settings, particularly when using a multi-tiered framework of prevention and intervention. Validating the EIS-SR is an important step in determining the utility of the measure for schools. The EIS-SR can be used by schools as part of a system designed to identify, prevent, and intervene in student social, emotional, and behavior problems. Thus, these data can be used with problem-solving teams within each school building to determine evidence-based interventions known to support the targeted problem area (Reinke, Sims, et al., 2018; Reinke, Thompson, et al., 2018). In addition, the data can be aggregated at the school level so that school personnel can determine risk areas with large numbers of students who are reporting challenges. For instance, using a public health framework, schools are advised that if 20% or more of their students are reporting risk in a particular area, implementing a universal preventive intervention will be more productive than intervening with one student at a time (see Reinke, Thompson, et al., 2018; Thompson et al., 2017). Grade-level data can be used in a similar manner, allowing schools to determine preventive interventions. Lastly, when the data are presented to school personnel, students who exhibit some risk, but not the highest risk, are also flagged. This supports their ability to determine which students would benefit from an evidence-based selective intervention (e.g., students with risk in peer relationship problems could be linked to group social skills training). The EIS-SR was built to be a feasible screening system that schools can use within tiered models of practice.

4.2. Limitations

Although the findings of this study have important implications for measuring risk indicators for social, emotional, and behavioral problems in elementary schools, there are a few limitations that must be considered. First, the Relational Aggression subscale had a lower than expected internal consistency and was not predictive of any of the school-based outcomes. Thus, this subscale needs to be adjusted and validated or removed from the overall EIS-SR in future studies. In the near term, users should use this scale with caution and explore the possibility of false positives, especially for students who may over report being unkind to others. It may be that students in elementary school cannot accurately report on relational aggression in ways that older youth can. Second, although the concurrent validity of several subscales was conducted, a few EIS-SR subscales did not have a comparable subscale with the BASC-3. A future area of research would be to further examine the concurrent validity of all subscales. Third, students who completed both the EIS-SR and the BASC-3 were recruited based on availability and thus did not constitute a representative sample; future studies will need to recruit representative samples to increase the generalizability of these findings. Fourth, another limitation is that both the EIS-SR and the BASC are student reported measures. In the future, comparisons with other reporters (e.g., caregivers, educators) should be considered. Fifth, although we provided initial evidence for four sources of validity for the EIS-SR, more work is needed to examine response processes related to the scale. Cognitive interviews with respondents would be a possible approach to provide such evidence. Lastly, although the sample was quite large and spanned 27 schools, the EIS-SR was administered within one geographic region among a group of students who were predominantly White. Future research should be conducted in a more racially diverse setting and different contexts.

4.3. Future directions

For the predictive validity analyses, the study used readily available variables from school records that school personnel find to be meaningful (i.e., discipline sanctions, attendance, and achievement). Thus, determining whether the EIS-SR subscales can predict these outcomes was important and meaningful, but further work examining how the EIS-SR can predict other outcomes is also needed (e.g., receipt of mental health services, health behaviors, substance use). Also, future studies should explore whether the combination of elevated risk on multiple EIS-SR subscales may increase the likelihood of later adverse outcomes. Other areas for future research include determining if the EIS-SR is sensitive to change as a result of universal, selective, or indicated interventions. As noted earlier, evaluating the concurrent validity of all subscales with another common and validated social behavioral assessment would be useful. Additional information may be collected regarding the construct validity of the subscales, especially with regards to the subscales that did not have a corresponding match with the BASC. Finally, future studies should collect consequential validity as a result of the utilization of the EIS-SR in the context of widespread dissemination to rural schools.

5. Conclusion

The present study established the factor structure, measurement invariance, and concurrent and predictive validity of the EIS-SR in an elementary student sample. Although many social, emotional, and behavior school screeners exist, the EIS-SR holds the potential for widespread dissemination, treatment utility, and impact. The EIS-SR is easy to access, administer, and interpret, and can be clearly connected to evidence-based school-delivered preventive and intervention supports that can reduce the number of students with social, emotional, and behavioral dysfunction and disorders (see Reinke, Thompson, et al., 2018; Thompson et al., 2017). Further work to establish the technical adequacy of the scale across middle and high school settings, as well as in relation to the teacher-report version of the EIS, is necessary to support the wide-scale dissemination of the tools.

Acknowledgements

The research reported here was supported by the Institute of Educational Sciences, U.S. Department of Education, through Grants R305H170023, R305C190014, and R305B150028 to the University of Missouri and by the Boone County Children Service Fund. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

American Psychological Association. (2017). Ethical principles of psychologists and code of conduct (2002, Amended June 1, 2010 and January 1, 2017). Retrieved from http://www.apa.org/ethics/code/index.aspx.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411–423. https://doi.org/10.1037/0033-2909.103.3.411.

Angold, A., Costello, E. J., & Erkanli, A. (1999). Comorbidity. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 40*(1), 57–87.

Arseneault, L., Milne, B., Taylor, A., Adams, F., Delgado, K., Caspi, A., & Moffitt, T. (2008). Being bullied as an environmentally mediated contributing factor to children's internalizing problems: A study of twins discordant for victimization. *Archival of Pediatric Adolescent Medicine, 2*, 145–150.

Avenevoli, S., Swendsen, J., He, J. P., Burstein, M., & Merikangas, K. R. (2015). Major depression in the National Comorbidity Survey Adolescent supplement: Prevalence, correlates, and treatment. *Journal of the American Academy of Child and Adolescent Psychiatry, 54*, 37–44.e32.

Barkley, E. F. (2010). *The Jossey-Bass higher and adult education series. Student engagement techniques: A handbook for college faculty*. Jossey-Bass.

Barkley, R. A. (1998). *Attention-deficit hyperactivity disorder: A handbook for diagnosis and treatment* (2nd ed.). Guilford Press.

Barkley, R. A., & Cunningham, C. E. (1979). The effects of methylphenidate on the mother– Child interactions of hyperactive children. *Archives of General Psychiatry, 36*, 201–208.

Beardslee, W. R., Keller, M., Lavori, P. W., Staley, J. E., & Sacks, N. (1993). The impact of parental affective disorder on depression in offspring: A longitudinal follow-up in a nonreferred sample. *Journal of the American Academy of Child and Adolescent Psychiatry, 32*, 723–730.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*. https://doi.org/10.3389/fpubh.2018.00149.

Bonifay, W., & Cai, L. (2017). On the complexity of item response theory models. *Multivariate Behavioral Research, 52*(4), 465–484. https://doi.org/10.1080/00273171.2017.1309262.

Bowen, N. K., & Masa, R. D. (2015). Conducting measurement invariance tests with ordinal data: A guide for social work researchers. *Journal of the Society for Social Work and Research, 6*(2), 229–249.

Briesch, A. M., Chafouleas, S. M., & Chaffee, R. K. (2018). Analysis of state-level guidance regarding school-based, universal screening for social, emotional, and behavioral risk. *School Mental Health, 10*(2), 147–162.

Briesch, A. M., Chafouleas, S. M., Dineen, J. N., McCoach, D. B., & Donaldson, A. (2021). School building administrator reports of screening practices across academic, behavioral, and health domains. *Journal of Positive Behavior Interventions*. https://doi.org/10.1177/2F10983007211003335.

Bruhn, A. L., Woods-Groves, S., & Huddle, S. (2014). A preliminary investigation of emotional and behavioral screening practices in K–12 schools. *Education and Treatment of Children, 37*, 611–634. https://doi.org/10.1353/etc.2014.0039.

Bub, K. L., McCartney, K., & Willett, J. B. (2007). Behavior problem trajectories and first-grade cognitive ability and achievement skills: A latent growth curve analysis. *Journal of Educational Psychology, 99*, 653–670. https://doi.org/10.1037/0022-0663.99.3.653.

Bulotsky-Shearer, R. J., Domínguez, X., Bell, E. R., Rouse, H. L., & Fantuzzo, J. W. (2010). Relations between behavior problems in classroom social and learning situations and peer social competence in Head Start and kindergarten. *Journal of Emotional and Behavioral Disorders, 18*, 195–210. https://doi.org/10.1177/1063426609351172.

Burt, K., & Roisman, G. (2010). Competence and psychopathology: Cascade effects in the NICHD study of early child care and youth development. *Development and Psychopathology, 22*, 557–567. https://doi.org/10.1017/S0954579410000271.

Burt, K. B., Obradovic, J., Long, J. D., & Masten, A. S. (2008). The interplay of social competence and psychopathology over 20 years: Testing transactional and cascade models. *Child Development, 79*, 359–374. https://doi.org/10.1111/j.1467-8624.2007.01130.x.

Byrne, B. (2012). *Structural equation modeling with Mplus: Basic concepts, applications, and programming*. Routledge.

Center for Behavioral Health Statistics and Quality. (2015). *2014 national survey on drug use and health: Detailed tables*. Substance Abuse and Mental Health Services Administration.

Centers for Disease Control and Prevention. (2013). Mental health surveillance among children – United States, 2005–2011. *Morbidity and Mortality Weekly, 62*, 1–35.

Chase, P. A., Hilliard, L. J., Geldhof, G. J., Warren, D. J., & Lerner, R. M. (2014). Academic achievement in the high school years: The changing role of school engagement. *Journal of Youth and Adolescence, 43*(6), 884–896.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504.

Cheung, A. H., Kozlof, N., & Sacks, D. (2013). Pediatric depression: An evidence-based update on treatment interventions. *Current Psychiatry Reports, 15*, 381.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*(2), 233–255.

Christle, C. A., Jolivette, K., & Nelson, C. M. (2007). School characteristics related to high school dropout rates. *Remedial and Special Education, 28*, 325–339.

Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). Academic Press.

Copeland, W. E., Miller-Johnson, S., Keller, G., Angold, A., & Costello, J. E. (2007). Childhood psychiatric disorders and young adult crime: A prospective population-based study. *The American Journal of Psychiatry, 164*, 1668–1675. https://doi.org/10.1176/appi.ajp.2007.06122026.

Cornell, D., Klein, J., Konold, T., & Huang, F. (2012). Effects of validity screening items on adolescent survey data. *Psychological Assessment, 24*(1), 21–35. https://doi.org/10.1037/a0024824.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98–104. https://doi.org/10.1037/0021-9010.78.1.98.

Costello, E. J., Mustillo, S., Erkanli, A., Keeler, G., & Angold, A. (2003). Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry, 60*, 837–844. https://doi.org/10.1001/archpsyc.60.8.837.

Côté, S. M., Vaillancourt, T., LeBlanc, J. C., Nagin, D. S., & Tremblay, R. E. (2006). The development of physical aggression from toddlerhood to pre-adolescence: A nationwide longitudinal study of Canadian children. *Journal of Abnormal Child Psychology, 34*, 71–85. https://doi.org/10.1007/s10802-005-9001-z.

Danforth, J. S., Barkley, R. A., & Stokes, T. F. (1991). Observations of parent–child interactions with hyperactive children: Research and clinical implications. *Clinical Psychology Review, 11*, 703–727.

Darney, D., Reinke, W. M., Herman, K. C., Stormont, M., & Ialongo, N. S. (2013). Children with co-occurring academic and behavior problems in first grade: Distal outcomes in twelfth grade. *Journal of School Psychology, 51*, 117–128. https://doi.org/10.1016/j.jsp.2012.09.005.

Demaray, M. K., Malecki, C. K., Ryoo, J. H., & Summers, K. H. (2021). Deconstructing bullying roles: A longitudinal latent profile analysis of bullying participant behaviors for students in grades 4 through 12. *Journal of School Psychology, 86*, 32–48.

Deutz, M. H., Geeraerts, S. B., Belsky, J., Deković, M., van Baar, A. L., Prinzie, P., & Patalay, P. (2020). General psychopathology and dysregulation profile in a longitudinal community sample: Stability, antecedents and outcomes. *Child Psychiatry & Human Development, 51*(1), 114–126.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development, 43*(2), 121–149. https://doi.org/10.1177/0748175610373459.

Dineen, J. N., Chafouleas, S. M., Briesch, A. M., McCoach, D. B., Newton, S. D., & Cintron, D. W. (2021). Exploring social, emotional, and behavioral screening approaches in US public school districts. *American Educational Research Journal*. https://doi.org/10.3102/00028312211000043.

DiPerna, J. C., & Elliott, S. N. (2002). Promoting academic enablers to improve student achievement: An introduction to the mini-series. *School Psychology Review, 31*(3), 293–297.

Dishion, T. J., Véronneau, M. H., & Myers, M. W. (2010). Cascading peer dynamics underlying the progression from problem behavior to violence in early to late adolescence. *Development and Psychopathology, 22*, 603–619. https://doi.org/10.1017/S0954579410000313.

Dowdy, E., Ritchey, K., & Kamphaus, R. W. (2010). School-based screening: A population-based approach to inform and monitor children's mental health needs. *School Mental Health, 2*, 166–176. https://doi.org/10.1007/s12310-010-9036-3.

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*, 405–432.

von der Embse, N. P., Kilgus, S. P., Iaccarino, S., & Levi-Nielsen, S. (2017). Screening for student mental health risk: Diagnostic accuracy, measurement invariance, and predictive validity of the Social, Academic, and Emotional Behavior Risk Screener-Student Rating Scale (SAEBRS-SRS). *School Mental Health, 9*(3), 273–283.

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University Press.

Finney, S., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. Hancock, & R. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 439–492). Information Age Publishing.

Finning, K., Ukoumunne, O. C., Ford, T., Danielsson-Waters, E., Shaw, L., Romero De Jager, I., Stentiford, L., & Moore, D. A. (2019). The association between child and adolescent depression and poor attendance at school: A systematic review and meta-analysis. *Journal of Affective Disorders, 245*, 928–938. https://doi.org/10.1016/j.jad.2018.11.055.

Fischer, M. (1990). Parenting stress and the child with attention deficit hyperactivity disorder. *Journal of Clinical Child Psychology, 19*, 337–346.

Forgatch, M. S., & DeGarmo, D. (2002). Extending and testing the social interaction learning model with divorce samples. In J. B. Reid, G. R. Patterson, & J. Snyder (Eds.), *Antisocial behavior in children and adolescents: A developmental analysis and model for intervention* (pp. 235–256). American Psychological Association.

Furlong, M. J., Fullchange, A., & Dowdy, E. (2017). Effects of mischievous responding on universal mental health screening: I love rum raisin ice cream, really I do! *School Psychology Quarterly, 32*(3), 320.

Gagnon, C., Craig, W. M., Tremblay, R. E., Zhou, R. M., & Vitaro, F. (1995). Kindergarten predictors of boys' stable behavior problems at the end of elementary school. *Journal of Abnormal Child Psychology, 23*(6), 751–766. https://doi.org/10.1007/BF01447475.

Georgiades, K., Duncan, L., Wang, L., Comeau, J., Boyle, M. H., & 2014 Ontario Child Health Study Team. (2019). Six-month prevalence of mental disorders and service contacts among children and youth in Ontario: Evidence from the 2014 Ontario Child Health Study. *The Canadian Journal of Psychiatry, 64*(4), 246–255. https://doi.org/10.1177/0706743719830024.

Ghandour, R. M., Sherman, L. J., Vladutiu, C. J., Ali, M. M., Lynch, S. E., Bitsko, R. H., & Blumberg, S. J. (2018). Prevalence and treatment of depression, anxiety, and conduct problems in U.S. children. *The Journal of Pediatrics, 206*, 256–267. https://doi.org/10.1016/j.jpeds.2018.09.021.

Gini, G., & Espelage, D. L. (2014). Peer victimization, cyberbullying, and suicide risk in children and adolescents. *Journal of the American Medical Association, 312*, 545–546. https://doi.org/10.1001/jama.2014.3212.

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

Goforth, A. N., Farmer, R. L., Kim, S. Y., Naser, S. C., Lockwood, A. B., & Affrunti, N. W. (2021). Status of school psychology in 2020: Part 1, demographics of the NASP membership survey. *NASP Research Reports, 5*(2), 1–17.

Good, R. H., III, & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. R. Shinn (Ed.), *Advanced applications of curriculum-based measurement* (pp. 61–88). The Guilford Press.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology & Psychiatry, 38*, 581–586.

Gresham, F. M., & Elliott, S. N. (2007). *Social skills improvement system rating scales manual*. Pearson Education.

Harvey, E. A., Youngwirth, S. D., Thakar, D. A., & Errazuriz, P. A. (2009). Predicting attention-deficit/hyperactivity disorder and oppositional defiant disorder from preschool diagnostic assessments. *Journal of Consulting and Clinical Psychology, 77*, 349–354. https://doi.org/10.1037/a0014638.

Haynie, D., Nansel, T., Eitel, P., Crump, A., Saylor, K., Yu, K., & Simons-Morton, B. (2001). Bullies, victims, and bully/victims: Distinct groups of at-risk youth. *The Journal of Early Adolescence, 21*, 29–49. https://doi.org/10.1177/0272431601021001002.

Henricsson, L., & Rydell, A. M. (2006). Children with behaviour problems: The influence of social competence and social relations on problem stability, school achievement and peer acceptance across the first six years of school. *Infant and Child Development, 15*, 347–366. https://doi.org/10.1002/icd.448.

Henry, K. L., Knight, K. E., & Thornberry, T. P. (2012). School disengagement as a predictor of dropout, delinquency, and problem substance use during adolescence and early adulthood. *Journal of Youth and Adolescence, 41*, 156–166. https://doi.org/10.1007/s10964-011-9665-3.

Herman, K. C., Hodgson, C. G., Eddy, C. L., Cohen, D. R., Reinke, W. M., Burrell, L., Mcfarlane, E. C., & Duggan, A. K. (2020). Does child likeability mediate the link between academic competence and depressive symptoms in early elementary school? *Child Development, 91*(2), e331–e344.

Herman, K. C., Lambert, S. F., Ialongo, N. S., & Ostrander, R. (2007). Academic pathways between attention problems and depressive symptoms among urban African American children. *Journal of Abnormal Child Psychology, 35*(2), 265–274.

Herman, K. C., Lambert, S. F., Reinke, W. M., & Ialongo, N. S. (2008). Low academic competence in first grade as a risk factor for depressive cognitions and symptoms in middle school. *Journal of Counseling Psychology, 55*, 400–410. https://doi.org/10.1037/a0012654.

Herman, K. C., & Ostrander, R. (2007). The effects of attention problems on depression: Developmental, academic, and cognitive pathways. *School Psychology Quarterly, 22*(4), 483–510.

Hinshaw, S. P. (1992). Externalizing behavior problems and academic underachievement in childhood and adolescence: Causal relationships and underlying mechanisms. *Psychological Bulletin, 111*(1), 127–155.

Hosmer, D. W., & Lemeshow, S. (2004). *Applied logistic regression* (2nd ed.). John Wiley & Sons.

Hoza, B. (2007). Peer functioning in children with ADHD. *Journal of Pediatric Psychology, 32*, 655–663. https://doi.org/10.1016/j.ambp.2006.04.011.

Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In *Structural equation modeling: Concepts, issues, and applications* (pp. 76–99). Sage Publications, Inc.

Huang, F. (2016). Alternatives to multilevel modeling for the analysis of clustered data. *Journal of Experimental Education, 84*, 175–196. https://doi.org/10.1080/00220973.2014.952397.

Huang, F. L., Reinke, W. M., Thompson, A., Herman, K. C., & County Schools Mental Health Coalition. (2019). An investigation of the psychometric properties of the early identification system–student report. *Journal of Psychoeducational Assessment, 37*(4), 473–485.

Huntington, D. D., & Bender, W. N. (1993). Adolescents with learning disabilities at risk? Emotional well-being, depression, suicide. *Journal of Learning Disabilities, 28*, 59–166. https://doi.org/10.1177/002221949302600303.

Jia, Y., Konold, T. R., Cornell, D., & Huang, F. (2018). The impact of validity screening on associations between self-reports of bullying victimization and student outcomes. *Educational and Psychological Measurement, 78*, 80–102. https://doi.org/10.1177/0013164416671767.

Kamphaus, R. W., & Reynolds, C. R. (2015). *Behavior assessment system for children—Third edition (BASC-3): Behavioral and emotional screening system (BESS)*. 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. https://doi.org/10.1176/appi.ajp.159.9.1548.

Keller, M. B., Lavori, P. W., Beardslee, W. R., Wunder, J., & Ryan, N. (1991). Depression in children and adolescents: New data on "undertreatment" and a literature review on the efficacy of available treatments. *Journal of Affective Disorders, 21*, 163–171. https://doi.org/10.1016/0165-0327(91)90037-S.

Kelley, K., & Pornprasertmanit, S. (2016). Confidence intervals for population reliability coefficients: Evaluation of methods, recommendations, and software for composite measures. *Psychological Methods, 21*(1), 69–92. https://doi.org/10.1037/a0040086.
Kilgus, S. P., Chafouleas, S. M., & Riley-Tillman, T. C. (2013). Development and initial validation of the social and academic behavior risk screener for elementary grades. *School Psychology Quarterly, 28*, 210–226. https://doi.org/10.1037/spq0000024.
Kilgus, S. P., & von der Embse, N. P. (2014). *Social, academic, and emotional behavior risk screener*. Theodore J. Christ & Colleagues.
Kline, R. (2011). *Principles and practice of structural equation modeling* (3rd ed.). Guilford Press.
Lewinsohn, P. M., Rohde, P., Seeley, J. R., Klein, D. N., & Gotlib, I. H. (2003). Psychosocial functioning of young adults who have experienced and recovered from major depressive disorder during adolescence. *Journal of Abnormal Psychology, 112*, 353–363. https://doi.org/10.1037/0021-843x.112.3.353.
Lewinsohn, P. M., Rohde, P., & Seely, J. R. (1998). Major depressive disorder in older adolescents: Prevalence, risk factors, and clinical implications. *Clinical Psychology Review, 18*, 765–794. https://doi.org/10.1016/S0272-7358(98)00010-5.
Lewinsohn, P. M., Solomon, A., Seeley, J. R., & Zeiss, A. (2000). Clinical implications of" subthreshold" depressive symptoms. *Journal of Abnormal Psychology, 109*(2), 345–351.
Li, Y., & Lerner, R. (2012). Interrelations of behavioral, emotional, and cognitive school engagement in high school students. *Journal of Youth and Adolescence, 42*, 20–32. https://doi.org/10.1007/s10964-012-9857-5.
Liu, J., Raine, A., Venables, P. H., & Mednick, S. A. (2004). Malnutrition at age 3 years and externalizing behavior problems at ages 8, 11, and 17 years. *American Journal of Psychiatry, 161*(11), 2005–2013.
Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Addison-Wesley.
Mash, E. J., & Johnston, C. (1990). Determinants of parenting stress: Illustrations from families of hyperactive children and families of physically abused children. *Journal of Clinical Child Psychology, 19*, 313–328.
Mash, E. J., & Wolfe, D. A. (2015). *Abnormal child psychology*. Wadsworth.
Masten, A. S., & Cicchetti, D. (2010). Developmental cascades. *Development and Psychopathology, 22*(3), 491–495.
Masten, A. S., Roisman, G. I., Long, J. D., Burt, K. B., Obradović, J., Riley, J. R., Boelcke-Stennes, K., & Tellegen, A. (2005). Developmental cascades: Linking academic achievement and externalizing and internalizing symptoms over 20 years. *Developmental Psychology, 41*, 733–746.
Maughan, B., Rowe, R., Loeber, R., & Stouthamer-Loeber, M. (2003). Reading problems and depressed mood. *Journal of Abnormal Child Psychology, 31*, 219–229.
McIntosh, K., Campbell, A. L., Carter, D. R., & Zumbo, B. D. (2009). Concurrent validity of office discipline referrals and cut points used in schoolwide positive behavior support. *Behavioral Disorders, 34*(2), 100–113.
McLaughlin, K. A., Hatzenbuehler, M. L., Mennin, D. S., & Nolen-Hoeksema, S. (2011). Emotion dysregulation and adolescent psychopathology: A prospective study. *Behaviour Research and Therapy, 49*(9), 544–554.
McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods, 23*(3), 412–433. https://doi.org/10.1037/met0000144.
Melnick, S. M., & Hinshaw, S. P. (2000). Emotion regulation and parenting in AD/HD and comparison boys: Linkages with social behaviors and peer preference. *Journal of Abnormal Child Psychology, 28*, 73–86. https://doi.org/10.1023/A:1005174102794.
Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., Benjet, C., Georgiades, K., & Swendsen, J. (2010). Lifetime prevalence of mental disorders in U.S. adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child and Adolescent Psychiatry, 49*, 980–989. https://doi.org/10.1016/j.jaac.2010.05.017.
Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749. https://doi.org/10.1037/0003-066X.50.9.741.
Missouri Department of Elementary and Secondary Education. (2017). Missouri Assessment Program grade-level assessments technical report 2017. Retrieved from https://dese.mo.gov/sites/default/files/asmt-gl-tech-report-2017.pdf.
Muthén, L. K., & Muthén, B. O. (2011). *Mplus User's guide sixth edition ed*. Muthén & Muthén.
National Association of School Psychologists. (2016). School-based mental health services: Improving student learning and well-being. Retrieved from https://www.nasponline.org/resources-and-publications/resources-and-podcasts/mental-health/school-psychology-and-mental-health/school-based-mental-health-services.
National Association of School Psychologists. (2020). Model for comprehensive and integrated school psychological services. In *The professional standards of the National Association of School Psychologists*. https://www.nasponline.org/x55315.xml.
Nigg, J., & Casey, B. (2005). An integrative theory of attention-deficit/hyperactivity disorder based on the cognitive and affective neurosciences. *Development and Psychopathology, 17*, 785–806. https://doi.org/10.1017/S0954579405050376.
Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.
Obradović, J., Burt, K. B., & Masten, A. S. (2009). Testing a dual cascade model linking competence and symptoms over 20 years from childhood to adulthood. *Journal of Clinical Child & Adolescent Psychology, 39*(1), 90–102.
O'Connell, M. E., Boat, T., & Warner, K. E. (Eds.). (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. National Academies Press.
Odgers, C., Moffitt, T., Broadbent, J., Dickson, N., Hancox, R., Harrington, H., Poulton, R., Sears, M. R., Thomson, W. M., & Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to adult outcomes. *Development and Psychopathology, 20*, 673–716. https://doi.org/10.1017/S0954579408000333.
Patterson, G. R. (1982). *A social learning approach: Vol. 3. Coercive family process*. Castalia.
Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. *American Psychologist, 44*, 329–335.
Patterson, G. R., Forgatch, M. S., & DeGarmo, D. S. (2010). Cascading effects following intervention. *Development and Psychopathology, 22*(4), 949–970.
Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial boys (A social international approach)*. Castalia.
Patterson, G. R., & Stoolmiller, M. (1991). Replications of a dual failure model for boys' depressed mood. *Journal of Consulting and Clinical Psychology, 59*(4), 491–498.
Pendergast, L. L., von der Embse, N., Kilgus, S. P., & Eklund, K. R. (2017). Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. *Journal of School Psychology, 60*, 65–82.
R Core Team. (2019). R: A language and environment for statistical computing. Retrieved from https://www.R-project.org/.
Reinke, W. M., & Herman, K. C. (2002). Creating school environments that deter antisocial behaviors in youth. *Psychology in the Schools, 39*, 549–559.
Reinke, W. M., Herman, K. C., Petras, H., & Ialongo, N. S. (2008). Empirically derived subtypes of child academic and behavior problems: Co-occurrence and distal outcomes. *Journal of Abnormal Child Psychology, 36*, 759–770. https://doi.org/10.1007/s10802-007-9208-2.
Reinke, W. M., Sims, W., Cohen, D., & Herman, K. C. (2018). Problem solving within an RTI framework: Roles and functions of effective teams. In P. C. Pullen, & M. J. Kennedy (Eds.), *Handbook of response to intervention and multi-tiered systems of support* (pp. 80–103). Routledge.
Reinke, W. M., Thompson, A., Herman, K. C., Holmes, S., Owens, S., Cohen, D., Tanner-Jones, L. A., Henry, L., Green, A., & Copeland, C. (2018). The county schools mental health coalition: A model for community-level impact. *School Mental Health, 10*(2), 173–180.
Reise, S. P., Kim, D. S., Mansolf, M., & Widaman, K. F. (2016). Is the bifactor model a better model or is it just better at modeling implausible responses? Application of iteratively reweighted least squares to the Rosenberg Self-esteem scale. *Multivariate Behavioral Research, 51*, 818–838. https://doi.org/10.1080/00273171.2016.1243461.
Reynolds, C. R., Kamphaus, R. W., & Vannest, K. J. (2015). *BASC3: Behavior assessment system for children*. PsychCorp.
Rhodes, J., Thomas, J. M., & Liles, A. R. (2018). Predictors of grade retention among children in an elementary school truancy intervention. *Journal of At-Risk Issues, 21*(1), 1–10.
Romer, D., McIntosh, M., Foa, E. B., Gur, R. E., Hendin, H., O'Brien, C. P., Seligman, M. E. P., & Walsh, B. T. (2005). The roles and perspectives of school mental health professionals in promoting adolescent mental health. In D. Evans (Ed.), *Treating and preventing adolescent mental health disorders: What we know and what we don't know* (pp. 597–615). Oxford University Press. https://doi.org/10.1093/9780195173642.003.0032.
Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y.

Schaeffer, C. M., Petras, H., Ialongo, N., Poduska, J., & Kellam, S. (2003). Modeling growth in boys' aggressive behavior across elementary school: Links to later criminal involvement, conduct disorder, and antisocial personality disorder. *Developmental Psychology, 39*(6), 1020–1035.

Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire. *Aggressive Behavior, 29*(3), 239–268. https://doi.org/10.1002/ab.10047.

Soneson, E., Howarth, E., Ford, T., Humphrey, A., Jones, P. B., Coon, J. T., Rogers, M., & Anderson, J. K. (2020). Feasibility of school-based identification of children and adolescents experiencing, or at-risk of developing, mental health difficulties: A systematic review. *Prevention Science, 21*(5), 581–603. https://doi.org/10.1007/s11121-020-01095-6.

Splett, J. W., Trainor, K. M., Raborn, A., Halliday-Boykins, C. A., Garzona, M. E., Dongo, M. D., & Weist, M. D. (2018). Comparison of universal mental health screening to students already receiving intervention in a multitiered system of support. *Behavioral Disorders, 43*(3), 344–356. https://doi.org/10.1177/0198742918761339.

Stanley, C., Petscher, Y., & Pentimonti, J. (2019). *Classification accuracy*. U.S. Department of Education, Office of Elementary and Secondary Education, Office of Special Education Programs, National Center on Improving Literacy.

Stapleton, L. (2006). Using multilevel structural equation modeling techniques with complex sample data. In G. Hancock, & R. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 345–383). Information Age Publishing.

Stormont, M., & Thomas, C. N. (2014). *Simple strategies for teaching children at risk, K–5*. Corwin Press.

Sun, J., Dunne, M. P., Hou, X., & Xu, A. (2011). Educational stress scale for adolescents: Development, validity, and reliability with Chinese students. *Journal of Psychoeducational Assessment, 29*, 534–546. https://doi.org/10.1177/0734282910394976.

Svetina, D., Rutkowski, L., & Rutkowski, D. (2019). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using M plus and the lavaan/semTools Packages. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(1), 1–20.

Swearer, S. M., & Hymel, S. (2015). Understanding the psychology of bullying: Moving toward a social-ecological diathesis–stress model. *American Psychologist, 70*, 344–353. https://doi.org/10.1037/a0038929.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Better decisions through science. *Scientific American, 283*, 82–87.

Taylor, L. D., Davis-Kean, P., & Malanchuk, O. (2007). Self-esteem, academic self-concept, and aggression at school. *Aggressive Behavior, 33*, 130–136. https://doi.org/10.1002/ab.20174.

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development, 88*, 1156–1171. https://doi.org/10.1111/cdev.12964.

Temcheff, C., Serbin, L., Martin-Storey, A., Stack, D., Sheilagh, H., Ledingham, J., & Schwartzman, A. (2008). Continuity and pathways from aggression in childhood to family violence in adulthood: A 30-year longitudinal study. *Journal of Family Violence, 23*, 231–242. https://doi.org/10.1007/s10896-007-9147-2.

Thompson, A. M., Reinke, W. M., Holmes, S., Danforth, L., Herman, K. C., & the County School Mental Health Coalition. (2017). The county school mental health coalition: A model for a systematic approach to supporting youth. *Children & Schools, 39*, 209–218.

Thompson, F. T., & Levine, D. U. (1997). Examples of easily explainable suppressor variables in multiple regression research. *Multiple Linear Regression Viewpoints, 24*(1), 11–13.

Tremblay, R. E., Nagin, D. S., Séguin, J. R., Zoccolillo, M., Zelazo, P. D., Boivin, M., Pérusse, D., & Japel, C. (2005). Physical aggression during early childhood: Trajectories and predictors. *Canadian Child and Adolescent Psychiatry Review, 14*, 3–9.

Tyler, E. T., Hulkower, R. L., & Kaminski, J. W. (2017). *Behavioral health integration in pediatric primary care: Considerations and opportunities for policymakers, planners, and providers*. Millbank Memorial Fund.

U.S. Department of Health and Human Services. (2017). Effects of bullying. Retrieved from https://www.stopbullying.gov/at-risk/effects/index.html.

Volpe, R. J., Briesch, A. M., & Chafouleas, S. M. (2010). Linking screening for emotional and behavioral problems to problem-solving efforts: An adaptive model of behavioral assessment. *Assessment for Effective Intervention, 35*, 240–244. https://doi.org/10.1177/1534508410377194.

Walker, H., Severson, H., & Feil, E. (2014). *Systematic screening for behavior disorders* (2nd ed.) (Ancora).

Wang, M. T., & Fredricks, J. A. (2014). The reciprocal links between school engagement, youth problem behaviors, and school dropout during adolescence. *Child Development, 85*, 722–737.

Wood, J. J., Cowan, P. A., & Baker, B. L. (2002). Behavior problems and peer rejection in preschool boys and girls. *The Journal of Genetic Psychology, 163*, 72–88. https://doi.org/10.1080/00221320209597969.

Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika, 81*(4), 1014–1045.

Yoon, M., & Lai, M. H. C. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling: A Multidisciplinary Journal, 25*(2), 201–213. https://doi.org/10.1080/10705511.2017.1387859.

Zentall, S. S. (2005). Theory- and evidence-based strategies for children with attentional problems. *Psychology in the Schools, 42*, 821–836. https://doi.org/10.1003/pits.20114.