

REFINING SCHOOL MENTAL HEALTH SCREENING USING MIXTURE  
MODELING: EVALUATION FOR USE IN APPLIED SETTINGS

---

A Dissertation  
presented to  
the Faculty of the Graduate School  
at the University of Missouri-Columbia

---

In Partial Fulfillment  
of the Requirements for the Degree  
Doctor of Philosophy

---

by  
CAROLINE G. HODGSON  
Dr. Wendy Reinke, Dissertation Supervisor

July 2020

The undersigned, appointed by the dean of the Graduate School, have examined the dissertation entitled

REFINING SCHOOL MENTAL HEALTH SCREENING USING MIXTURE  
MODELING: EVALUATION FOR USE IN APPLIED SETTINGS

presented by Caroline Hodgson,

a candidate for the degree of doctor of philosophy,

and hereby certify that, in their opinion, it is worthy of acceptance.

---

Professor Wendy Reinke

---

Assistant Professor Wesley Bonifay

---

Associate Professor Aaron Thompson

---

Associate Professor Stephen Kilgus

## DEDICATION

To my husband, Tony, who has shown me grace upon grace, who cultivated my courage, and who persuaded me to finish graduate school, one pizza roll at a time. To my beloved friends, who have cared for, laughed with, and prayed for me to shape who I've become. And finally, for my parents, whose selfless blazed trails for me to pursue my goals.

## ACKNOWLEDGEMENTS

A major thank you is due to my advisor, Wendy Reinke. Your vision inspired me to dream big with research, believe in my ideas, and shoot for expansive goals, and your support enabled me to actually execute on those plans, even while away during internship and in the midst of a pandemic. Many thanks are due to Wes Bonifay, without whom I likely never would have developed any expertise or interest in item response theory. Your guidance was essential to developing my identity as a scientist, and your commitment to asking bigger questions about what measurement even *is* continues to inspire and challenge me. Thank you to Aaron Thompson, for your genuine curiosity about anything that could positively impact children's well-being, and for always encouraging me to learn more. Finally, thank you to Stephen Kilgus. I will always be grateful for your example as a scientist and mentor, as well as your commitment to making real change in children's well-being through rigorous and meaningful research.

Many school stakeholders, including students, teachers, and administrators, have made this work possible through their collaboration in screening children for mental health concerns. Thank you to the Boone County Schools Mental Health Coalition for their continuing efforts to move the needle on youth mental health in central Missouri and to create lasting change in our schools and the lives of local youth.

Special thanks are due to my incredible graduate school colleagues, including Colleen Eddy, Amanda Allen, Lauren Henry, and many others. Their support and friendship was integral to my completion of this degree, and their intelligence, thoughtfulness, passion, and commitment to excellence have inspired me to do better research for all the right reasons.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... ii

LIST OF ILLUSTRATIONS .....v

ABSTRACT..... vi

CHAPTER I: INTRODUCTION.....1

    Background ..... 1

    Statement of the Problem..... 2

    Purpose of the Study ..... 3

    Research Questions ..... 5

CHAPTER II: LITERATURE REVIEW .....7

    Youth Mental Health..... 7

    Public Health and Prevention Strategies ..... 10

        Schools as Key Sites for Intervention..... 12

        Multi-Tiered Systems of Support..... 14

        Boone County Schools Mental Health Coalition..... 16

    School-based Behavioral and Emotional Screening ..... 17

    Advanced Methodology in Mental Health Measurement..... 20

        Latent Class and Latent Profile Analysis..... 20

        Item Response Theory ..... 21

        Research Applications to Youth Mental Health ..... 24

        Multidimensional IRT..... 25

        Mixture IRT ..... 25

    Purpose..... 28

CHAPTER III: METHODS .....31

    Research Design..... 31

    Participants..... 31

        Student Respondents..... 31

    Measures ..... 32

        Early Identification System – Student Report (EIS-SR)..... 32

    Procedures..... 34

    Data Analytic Plan ..... 35

        Exploratory and Confirmatory Factor Analysis..... 35

Evaluating Item Response Theory Models .....	36
Person-Centered Analyses .....	38
CHAPTER IV: RESULTS.....	40
RQ1: Dimensionality and Diagnostics.....	40
Confirmatory Factor Analysis.....	41
RQ2: IRT Item Diagnostics .....	42
RQ3: Person-Centered Analysis (Latent Profile Analysis) .....	42
Latent Class Regression .....	44
CHAPTER V: DISCUSSION.....	48
Factor Structure.....	48
Latent Profile Analysis .....	52
Profiles of Distress and Dysregulation .....	55
Academic and Behavioral Outcomes.....	58
Latent Profile Analysis in School Screening Research.....	60
Multidimensional IRT in Screening Research .....	60
Limitations .....	62
Future Directions .....	64
Applications of Person-Centered Analyses to School-Based Screening .....	64
Data Use and Applied Interpretation of Person-Centered Research.....	66
Integration of Additional Informants .....	67
Conclusion .....	70
References.....	71
VITA.....	110

## LIST OF ILLUSTRATIONS

Table 1. Comparison of candidate mixture models for estimating EIS item properties and latent profiles of risk concurrently.....	100
Table 2. Fit statistics for exploratory and confirmatory factor analyses of the EIS-SR. .	101
Table 3. Standardized factor loadings for the exploratory (confirmatory) factor analysis of the EIS-SR. ....	102
Table 4. Subscale correlations, reliabilities, and descriptive statistics based on factors of the EIS-SR. ....	103
Table 5. Fit indices for each latent profile solution evaluated. ....	104
Table 6. Probability of most likely class membership (row) by latent class (column) for the five-profile LPA solution. ....	105
Table 7. Proportion of likely class membership in the five-profile solution. ....	106
Table 8. Outcome means, SE, and equality tests across profiles of student risk (n = 5553). ....	107
Figure 1. Scree plot of the EIS-SR.....	108
Figure 2. Mean z-scores of the five-profile LPA solution across EIS-SR indicators. ....	109

## ABSTRACT

Universal socioemotional and behavior screening procedures continue to gain momentum and adoption in schools; however, a major research-to-practice gap remains in translating these data into meaningful interventions for identified students. Despite advances in methodology in school psychology research, parallel efforts have not been made to translate screening data into highly individualized, data-based interventions. The current study sought to apply person-centered analytic strategies to an existing youth mental health screening tool (EIS-Student; Reinke et al., 2020) to determine the extent to which empirically-derived latent subgroups could reveal clinically relevant conclusions in universal screening data. The sample was drawn from existing universal screening data from fall 2018 EIS-Student participants across nine high schools in a single Midwestern county ( $n = 5,860$ ). Results revealed a stable 7-factor structure for the EIS-Student as well as a 5 latent profiles. Profiles corresponded with meaningfully different scores on subscales of mental health risk from the EIS-Student, and significant relationships were observed between profile membership and academic and behavioral outcomes. Characteristics of each profile are discussed for possible clinical conceptualization and applications within a socioemotional screening program. Implications for universal screening efforts are discussed, including the use of mixture modeling to identify subgroups of student need, the value of person-centered analyses in school intervention decision-making, and considerations for key stakeholders in applied contexts.



## **CHAPTER I: INTRODUCTION**

### **Background**

According to epidemiological research, roughly 20% of children and adolescents experience mental health difficulties before reaching adulthood (Merinkangas et al., 2010), roughly half of which would warrant a serious mental health diagnosis (Walker et al., 1996; Walker, Nishioka, Zeller, Severson, & Feil, 2000). Youth whose socioemotional difficulties persist unaddressed are more likely to experience academic failure (Reinke, Herman, Petras, & Ialongo, 2008), difficulties with social relationships, and substance abuse (Levitt & Merrell, 2009). Public health models provide an alternate focus for conceptualizing the problem of youth mental health on a population scale through considerations of prevention and early identification efforts. Schools have been identified as a primary site for conducting population-level youth mental health surveillance and prevention because of their unique access to a majority of youth for a large portion of their daily lives (Herman, Merrell, Reinke, & Tucker, 2004). To address student behavioral, socioemotional, and academic needs, schools are increasingly adopting multi-tiered frameworks that follow a public health model of universal screening and prevention, targeted prevention and early intervention for those identified as at-risk, and more intensive services for those with continued needs (Bradshaw et al., 2008; Kilgus, Reinke, & Jimerson, 2015).

Essential to a public health approach is the use of accurate and efficient assessment tools to quickly identify at-risk youth from a universal sample. Because screening is the primary pathway to identification for individualized supports as well as systems-level needs, highly accurate and efficient screeners are necessary to optimize all

other prevention and early intervention efforts. As measurement methodology continues to advance in sophistication in other areas of educational (Wu, Kuo, & Wang, 2017) and psychological (Finch & Pierson, 2011) research, school-based behavioral and emotional screening research should take advantage of these tools to optimize the results of Multi-tiered Systems of Support (MTSS) efforts.

### **Statement of the Problem**

Rigorous methodology in school-based emotional and behavioral screening can bolster the accuracy of our identification and provide richer detail on the nature of student risk to guide case conceptualization and intervention individualization. Many technically adequate screeners have been developed and evaluated for their use in schools, such as the Student Risk Screening Scale (SRSS; Drummond, 1994), the Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007), and several others. Studies describing the psychometric properties of these screeners have primarily relied upon classical test theory (CTT) metrics, such as Cronbach's coefficient alpha (Cronbach, 1951), test-retest reliability, and correlational measures of convergent, divergent, and predictive validity (Glover & Albers, 2007). Although some screeners have been evaluated with more novel item response theory (IRT) methodology, (e.g. Kilgus, Bonifay, von der Embse, Allen, & Eklund, 2018), which allows researchers to evaluate and refine each item's ability to reduce statistical error and estimate student ability, few of these studies have evaluated item characteristics beyond basic item functioning. Additionally, these studies rarely take place during early stages of test development, so they mostly function to describe ongoing practice rather than optimize developing measures.

Finally, previous research has done little to apply advanced quantitative analysis of socioemotional assessment to school-based intervention decision-making. Published screening measures rely primarily on static standard score risk categories (based on standard scores and deviance from an established norm sample) or sum score cutoffs (based on relationships to gold standard measures and optimizing screener sensitivity and specificity), from which school decision-makers determine whether a student should be considered in need of services. Further, at the systems level, there is a mismatch between the unit of analysis with which school stakeholders are tasked with examining when making intervention decisions. School decision-makers are often provided variable-centric screening data (e.g. number of students considered “at risk” across several risk domains) that they must apply to decisions about individual students, each of which may contain socioemotional risk across multiple interconnected domains, requiring individualized intervention planning. Person-centered analytic strategies, including latent class analysis, latent profile analysis, and mixture item response theory (concurrent modeling of latent profiles and dimensional IRT scores for each dimension of risk), allow for student classification into one of several empirically-derived risk categories, which may represent a more appropriate match for quickly assessing student needs at the school level. Modeling latent profiles of participants is not new in educational or psychological research, but its potential for rich application in youth mental health prevention has remained largely untapped in previous studies.

### **Purpose of the Study**

The purpose of the current study is to apply advanced measurement methodology to a universal, school-based, socioemotional screener to 1) evaluate the factor structure

and basic measurement properties of items, 2) to attempt to model item functioning with IRT and person-centered analysis (either latent profile analysis or novel mixture IRT methodology), thus extending advanced methods to youth mental health measurement, and 3) to demonstrate the tenability of harnessing information generated by person-centered analysis of screening tools to support school-based decision-making. Data were drawn from the Boone County Schools Mental Health Coalition (the Coalition; Reinke et al., 2018), a collaboration among six independent school districts intended to promote prevention and early intervention in student behavioral and emotional concerns through universal screening, consultation, and direct service provision. The Coalition is built on a public health model, emphasizing proactive identification of students in need of support through regular universal screening and supporting capacity-building in early mental health intervention.

Specifically, this study evaluated the Early Identification System- Student Report (EIS-SR; Huang, Reinke, Thompson, Herman, & County Schools Mental Health Coalition, 2019), the novel universal screening measure utilized by the Coalition. We aimed to conduct factor-analytic research to determine the suitability of the measure for polytomous IRT analyses. These would allow examination of item performance and mixture IRT analysis (de Ayala & Santiago, 2017) to apply an advanced measurement technique to better understand our sample. Additionally, from the results of the final model of sample characteristics, sample reports were generated to visualize and explain the data at the school and individual level to support applied, systems-level intervention recommendations. This will support future evaluation of whether practitioners interpret the results of advanced models differently from traditional models and determining the

tenability of translating advanced measurement models for applied practice to glean new insights from student data.

### **Research Questions**

1. What is the dimensionality and factor structure of the EIS-SR in a high school sample?

Although we predicted that the EIS-SR would be multidimensional (6 factors, Huang et al., 2019), initial diagnostics will be conducted to determine dimensionality of the scale and suitability for unidimensional or multidimensional IRT analysis. Based on the breadth of the EIS-SR scale, we predicted a factor structure consisting of between four and seven factors. However, we did not predict that all items would perform strongly, whether in confirmatory factor analysis (CFA) or IRT analyses, and these items would be subject to review and possible removal from the checklist.

2. Should an IRT model be appropriate for these data, when the EIS-SR is evaluated using a polytomous (and possibly multidimensional) IRT model, do any items warrant removal due to poor performance?

Should IRT analyses be determined appropriate, we predicted that most items would possess sufficient discriminant ability ( $a > 1.0$ ) and appropriate locations (also known as “difficulty” or “severity” coefficients; de Ayala & Santiago, 2017) along the latent trait(s) ( $-2 < b < 2$ , or measuring most precisely for individuals within 2 standard deviations of the mean of the latent trait).

3. What classes of individuals will emerge in the school-based sample when EIS-SR data are modeled using either a multidimensional mixture IRT model or latent profile analysis?

We predicted that several classes of individuals at varying levels of risk along different traits would emerge. We predicted that class membership would be predicted by a variety of demographic variables (such as race, gender, income level, standardized test scores) and that each class would represent a profile of at-risk students who may be apt for systems-level or small-group intervention.

## **CHAPTER II: LITERATURE REVIEW**

This chapter reviews the following themes relevant to the present work: the importance and prevalence of youth mental health problems, models for prevention and early intervention to address the problem at a large scale, existing methods of identifying youth at risk for mental health problems through their schools, methodological advances that can evaluate and improve screening approaches, and the application of screening results to decision-making for student interventions.

### **Youth Mental Health**

Research estimates that 13-20% of children experience mental health difficulties, and the prevalence of mental health disorders in children is increasing (Perou, et al., 2013). Among these disorders, the top four experienced by youth include attention-deficit/hyperactivity disorder (6.8%), conduct problems (3.5%), anxiety (3%), and depression (2.1%; Perou, et al., 2013). Such concerns have been found to be related to and predictive of other concerns as well: prevalence rates for co-occurring academic and behavior problems have been found at a rate of 12.5% (Reinke, Herman, Petras, & Ialongo, 2008). Youth mental health difficulties predict a variety of negative outcomes throughout adolescence, including academic failure, poor social skills, suspension from school, affiliation with deviant peers, conduct problems, and substance abuse (Reinke et al., 2008; Levitt & Merrell, 2009; Nail et al., 2015). Further, the occurrence of mental health difficulties in adolescence greatly increases the likelihood of reemerging symptoms in adulthood (Tram & Cole, 2006).

Unfortunately, it is estimated that nearly 80% of youth who require mental healthcare will not access those services (Kataoka, Zhang, & Wells, 2002). Many mental

health conditions are under- or misdiagnosed in clinical settings, especially for youth of color (Angold et al., 2002), and an even greater number of youth face barriers to pursuing community-based clinical services in the first place, including cost of treatment, transportation, perceptions of mental health services available, and other factors (Owens et al., 2002, Wu et al., 2001).

Most common youth mental health difficulties can be classified into one of two areas: externalizing and internalizing problems (Achenbach, 1966; Blanco et al., 2015; Kamphaus, 2012). This distinction was first proposed and evaluated as a way of understanding youth psychopathology (Achenbach, 1966), only to be replicated later in adult populations (Achenbach, Krukowski, Dumenci, & Ivanova, 2005; Forbush & Watson, 2013). Although the *Diagnostic and Statistical Manual for Mental Disorders Fifth Edition* (DSM-5; American Psychiatric Association, 2013) does not divide diagnostic categories along these lines, emerging systems of psychopathology classification favor this distinction (e.g., Kotov et al., 2017). Quantitative approaches to psychopathological classification systems, based on observed covariations among symptoms, allow researchers to describe sets of symptoms based on their meaningful interrelationships rather than predetermined assumptions about the nature of psychopathology (Kotov, 2016).

Externalizing problems include symptoms such as attention difficulties, outbursts of anger, and antisocial behavior (Walton, Ormel, & Krueger, 2011; Krueger & Markon, 2014). Traditional DSM-5 disorders that qualify as moderate to severe examples of externalizing difficulties include attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), conduct disorder (CD), intermittent explosive



disorder (IED), and disruptive mood dysregulation disorder (DMDD; American Psychiatric Association, 2013). Internalizing problems are characterized by symptoms of sadness, withdrawal, fearfulness, panic, and somatic symptoms such as stomach aches or disturbances in sleep or appetite (Achenbach & Edelbrock, 1978; Kotov et al., 2017). These correspond with many of the symptoms of major depressive disorder (MDD) and anxiety disorders including generalized anxiety disorder (GAD), social anxiety disorder (SAD), panic disorder (PD), specific phobias, and obsessive-compulsive disorder (OCD; American Psychiatric Association, 2013).

Mental health difficulties, internalizing or externalizing, can be conceptualized as existing on a continuum ranging from the presentation of only one or two symptoms to qualifying for one or more DSM-5 diagnoses (Haslam, Holland & Kuppens, 2012). Although categorical definitions of mental disorders are useful for insurance purposes and inter-clinician communication, dimensional measures of all kinds of youth psychopathology have improved reliability and validity over categorical measures (Markon, Chmielewski, & Miller, 2011; Walton et al., 2011) and account for more nuanced understandings of mental illness. For example, evidence suggests that even subsyndromal depressive symptom levels can impact youth functioning (Carrellas, Biederman, & Uchida, 2017; Backenstrass et al., 2006), develop into full disorders in adulthood if untreated (Shankman et al., 2009), and remit with intervention (Tadić et al., 2010). Systematically identifying youth in the population who are experiencing early mental health symptoms, whether or not they meet full diagnostic criteria for a disorder, can highlight students who may not otherwise receive mental health supports, open opportunities for early intervention programs, and diagnose population-level needs that

could be addressed through universal intervention programming. Furthermore, identifying the presence of common risk factors in youth who currently have no mental health symptoms can aid in selection for prevention programming.

### **Public Health and Prevention Strategies**

Multiple national calls have been made for the development of programming to support prevention, early identification, and early intervention for a variety of youth mental health difficulties (National Research Council and Institute of Medicine, 2009; Murthy, 2015). Public health models provide an alternate focus for conceptualizing the problem of youth mental health on a population scale: consideration for prevention and early identification efforts. Prevention scientists have developed a set of strategies and values for research and intervention programs designed to reduce the incidence, severity, and overall population-level burden of problems like mental illness (Kellam, Koretz, & Mosciki, 1999; Coie et al., 1993).

The developmental prevention science model proposed by Kellam and colleagues (1999) outlines an interdisciplinary approach to reducing the public health burden of mental health difficulties at large. Kellam and colleagues' model focuses on youth and intervening during critical stages of development, but the model is conceptualized as preventing population mental health difficulties at all ages and not just those experienced during youth. Primarily, this model involves identifying characteristics of youth and their environments that may operate as risk or protective factors for mental health concerns, then developing interventions that either reduce risk factors or bolster protective factors, with the aim to prevent the distal development of mental health concerns (Kellam et al., 1999). Similarly, Coie and colleagues (1993) outlined a national agenda for prevention

science research in youth mental health, emphasizing the importance of developmental theory, identification of risk factors common to multiple disorders, and the unique roles of both universal and indicated prevention programming. Standardized prevention programs developed under this model can be universal with the aim to prevent problems in a whole population, or indicated, in which select participants with salient risk factors are invited to participate before developing problems of interest.

The set of iterative research strategies described by the developmental prevention research model (Kellam et al., 1999) is based on the model of the National Cancer Institute (Greenwald & Cohen, 1984) and other existing public health models for the prevention of disease and the elimination of its societal burden. These models typically include multitiered logic, involving a continuum of supports from primary, universal prevention to individualized interventions (Mrazek & Haggerty, 1994; Walker et al., 1996). This model has been applied to youth mental health research at all stages of the research cycle internationally (Catalano et al., 2012).

Because of the costs of universal prevention efforts, as well as the potential for their failure (e.g., Ennett, Tobler, Ringwalt, & Flewelling, 1994), it is important to critically evaluate the prevention science model and its associated interventions prior to recommending wide dissemination of prevention programming. Evidence within prevention science takes many forms, ranging from randomized controlled trials of specific prevention programs to studies that identify key malleable risk factors and demonstrate their sensitivity to intervention (Davis, 2002). Kellam and Langevin (2003) described how varying standards of evidence are appropriate for different strategies within the prevention science research cycle; for example, efficacy trials are intended to

determine a program's effects and potential in "perfect" experimental conditions, whereas effectiveness trials evaluate the program within authentic implementation conditions. Along similar themes, the Society for Prevention Research (SPR) outlined a series of criteria for determining given prevention programs to be efficacious, effective, or ready for broad dissemination (Flay et al., 2005).

Randomized trials of developmentally-oriented mental health prevention programs continue to provide evidence for reliance on public health and prevention science models. For instance, computerized or automated prevention programs have shown promise for youth mental illness prevention, as one systematic review found that 60% of online anxiety programs and 83% of online depression programs were found to be effective on at least one measure (Christensen et al., 2010). Further, many community- and family-based programs have demonstrated effectiveness at reducing population-level mental health concerns (Hawkins, Catalano, & Arthur, 2002; Spoth, Trudeau, Gyll, Shin, & Redmond, 2009; Wolchik et al., 2002). One statewide report identified dozens of efficacious prevention programs for youth mental health, delinquency, and substance abuse, and evaluated their cost-effectiveness; these recommendations were incorporated into state policy to advance the impact of preventive services (Lee et al., 2012). Most evaluations of structured mental health prevention programs have taken place in schools, as they are an apt setting for these efforts.

### **Schools as Key Sites for Intervention**

Although prevention programming can and has taken place in community centers and medical offices, schools represent a key site for the identification and treatment of youth mental health difficulties (Masia-Warner, Nangle, & Hansen, 2006). Most school-

age children spend a majority of their weekdays in school buildings, whereas they may only visit a pediatrician, for example, once annually. Attention to and intervention within ecological systems are a key theme within the field of prevention science (Coie et al., 1993), and schools are a primary system within which youth develop. Within a public health framework, schools are uniquely equipped for universal screening, universal and indicated prevention efforts, and early intervention for mild to moderate symptoms (Herman, et al., 2004). Indeed, schools are the primary provider of mental health services for school-aged children (Rones & Hoagwood, 2000), and 70-80% of children with mental health needs will access services through their school (Burns et al., 1995). Fewer racial disparities exist in utilization of school-based mental health services when compared to community-based mental health services (Angold et al., 2002), and utilization of school-based services is associated with higher rates of community-based service utilization, suggesting that schools may provide access to a continuum of appropriate supports (Tegethoff, Stalujanis, Belardi, & Meinschmidt, 2014). Furthermore, school psychologists and other school-based mental health professionals are uniquely equipped to implement preventive strategies, as their training emphasizes ecological systems theory, priming them to adjust students' environments to support their flourishing (Abrams, Flood, & Phelps, 2006; Eklund, Meyer, Way, & McLean, 2017).

Applying prevention focused public health models to mental health in school settings is not new. Merrell and Buchanan (2006) explicitly applied the public health framework to school intervention delivery to increase the capacity of schools to meet student needs. Stormont, Reinke, and Herman (2010) described a series of intervention evaluations, established curricula, and implementation challenges all involving the

application of prevention science to mental health in schools, and there have been calls for school mental health professionals to play key roles in arranging and advocating for prevention efforts (e.g., Herman et al., 2004). Randomized trials have found that school-based universal and selective prevention programming can be efficacious in preventing depression (Calear & Christensen, 2010), anxiety (Neil & Christensen, 2009), aggressive and disruptive behavior (Wilson & Lipsey, 2007; Chapman, Buckley, Sheehan, & Shochet, 2013), and other concerns. Beyond prevention and early intervention efforts, school-based intensive services can bridge an important gap in access to mental healthcare. Research has found that providing mental health services to identified students in their schools can reduce schoolwide rates of suicidal ideation and behavior (Paschall & Bersamin, 2018). One of the most widely disseminated applications of prevention to school settings, however, is through the use of multi-tiered systems of support frameworks.

### **Multi-Tiered Systems of Support**

Calls for proactive solutions to child socioemotional and academic difficulties resulted in the development of multi-tiered systems of support, or MTSS (Jimerson, Burns, & VanDerHeyden, 2007; Tilly, 2008), a framework designed to apply prevention approaches to school settings. MTSS is a type of school-based model developed generally on the public health model (Merrell & Buchanan, 2006), describing a continuum of supports for preventing and intervening with problems at the universal, selective, and intensive levels. Typically, models involve three tiers of support, with corresponding interventions and strategies: Tier 1 represents universal prevention strategies and curricular changes for all students, Tier 2 includes students with identified

risk or some early symptoms matched to general interventions, often delivered in small groups, and Tier 3 involves more intensive, individualized interventions for students who did not respond adequately to Tier 2 supports (Severson, Walker, Hope-Doolittle, Kratochwill, & Gresham, 2007).

Much evidence has accumulated in support of the use of MTSS models to address social behavior in schools, frequently through a model called Positive Behavior Interventions and Supports (PBIS; Sugai & Horner, 2002). PBIS relies on tiered logic and universal prevention efforts to reduce schoolwide problem behaviors and proactively support struggling students. Based on evaluations of its implementation, PBIS is considered an evidence-based model and represents a successful application of general public health frameworks to mental health in schools (Horner, Sugai, & Anderson, 2010). PBIS has been shown to reduce student externalizing and attention difficulties in primary (Horner et al., 2009) and secondary (Lane, Wehby, Robertson, & Rogers, 2007) samples, and these reductions in problem behaviors are linked to reduced suspensions (Bradshaw, Mitchell, & Leaf, 2010), and improved academic outcomes in PBIS schools (Lassen, Steele, & Sailor, 2006). However, less evidence has been established for the use of MTSS models to address internalizing difficulties, such as anxiety and depression (McIntosh, Ty, & Miller, 2014). This is perhaps due to discrepant conceptual understandings of the causes and consequences of externalizing (commonly within a socioecological framework) and internalizing concerns (individual-centric medical model) in schools (Kilgus et al., 2015), as well as the historical view that targeting internalizing problems is outside of the scope of school-based practice (Chafouleas, Kilgus, & Wallach, 2010). A

holistic approach, inclusive of both externalizing and internalizing concerns, is necessary for tackling the whole burden of youth mental health problems.

### **Boone County Schools Mental Health Coalition**

Multi-tiered systems offer a promising avenue to address youth mental health on a population level by applying public health frameworks in school settings. The Boone County Schools Mental Health Coalition (the Coalition; Reinke et al., 2018; Thompson et al., 2017) is an innovative partnership between six county school districts, as well as the School Psychology and Social Work programs at the University of Missouri, to implement a large-scale, multi-tiered approach to the prevention and intervention of youth mental health difficulties in schools. A countywide needs assessment revealed that although schools observed significant student mental health difficulties across all grade levels, they did not have tools to systematically identify youth in need of support, processes for connecting students to appropriate interventions, or the requisite professional development for staff to meet those needs (Reinke et al., 2018).

To meet these needs, the Coalition was formed in 2015, funded by a countywide sales tax established to support programs promoting the mental health of county youth. Full-time mental health practitioners support the coordination of services across every K-12 public school building and two parochial schools in the county, reviewing universal screening data, providing systems and teacher consultation, conducting professional development workshops, and providing direct, evidence-based services to students in their schools. All six school district superintendents serve on the executive board of the Coalition and advise its directors and coordinators, ensuring that the priorities of the Coalition are representative of actual school needs – this is consistent with the value of



stakeholder partnership described as essential to prevention research, including effective engagement with existing leadership structures (Kellam et al., 1999). The Coalition represents an innovative way to prevent and target a variety of youth mental health concerns on a massive scale by not only providing services in schools, but also equipping school systems to identify and ameliorate risk effectively.

### **School-based Behavioral and Emotional Screening**

As outlined previously, the developmental prevention science model (Kellam et al., 1999) suggests that the most effective prevention and early intervention programs are those that accurately identify youth who, for a variety of reasons, are at risk for developing difficulties later in life, and targeting malleable risk factors to alter the course of development. Essential to this model is the use of assessment tools to quickly identify at-risk youth from a universal sample (Zuckerbrot et al., 2007; Kessler et al., 2002); for example, the American Academy of Pediatrics recently called for universal screening for depression of all adolescent patients starting at age 12 (Zuckerbrot et al., 2018). Broadly, screening practices are commonplace in school settings for a variety of concerns, including vision and hearing (Kemper, Fant, Bruckman, & Clark, 2004), academic (Lembke, McMaster, & Stecker, 2010; Walker & Shinn, 2002), and behavioral difficulties (Severson et al., 2007).

For behavioral concerns, current common screening approaches include examining frequency of student office discipline referrals (ODRs; Sugai, Sprague, Horner, & Walker, 2000), teacher nomination of students of concern, and systematic rating scales (including teacher, student, or parent report forms). Rating scales enable schools to rely on structured, predetermined criteria for flagging students and to identify

students proactively rather than waiting for students to experience multiple office referrals. Although systematic behavioral and emotional screening is comparatively less common, (Kilgus, Chafouleas, & Riley-Tillman, 2013), around 12.6% of schools report engaging in the practice (Bruhn, Woods-Groves, & Huddle, 2014). Additionally, previous research has demonstrated that reliance on ODRs and teacher nomination can result in the under-identification of students in need of services, and the subjectivity of these methods makes them vulnerable to bias and inappropriate or disproportionate flagging of students. For instance, research has found that African-American students are disproportionately referred to the office and suspended even when controlling for observed student behavior and school-level variables (Rocque, 2010; Skiba, Michael, Nardo, & Peterson, 2002). Finally, in some cases, about one third of students identified through systematic screening procedures were previously unknown to school staff, demonstrating again the advantage of systematic screening measurement over teacher nomination or traditional referral methods (Scott et al., 2009).

Many socioemotional and behavioral screeners have been developed and evaluated for use in school settings, including the Student Risk Screening Scale (SRSS; Drummond, 1994), the Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007), the Systematic Screening for Behavior Disorders tool (SSBD; Walker & Severson, 1992), the Strengths and Difficulties Questionnaire (SDQ; Goodman, 2001), the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS; Kilgus, Chafouleas, & Riley-Tillman, 2013), and more. Glover and Albers (2007, p. 119) recommended that screeners be evaluated for “(a) their appropriateness for the intended use,” including appropriateness for initial screening and appropriate cut

scores based on representative norm samples, “(b) their technical adequacy,” including evidence of their reliability and validity, and “(c) their usability,” including cost, burden placed on raters including teachers, and length of the rating scale. Much research has been dedicated to establishing whether existing screeners meet these criteria. However, existing screeners for socioemotional concerns vary in their cost, burden placed on respondents, and degree of technical adequacy.

The Coalition relies on its own mental health screening measure: the Early Identification System (EIS-SR; Huang et al., 2018; Reinke, et al., 2018). The EIS-SR screens for mental health symptoms (internalizing and externalizing domains), overall emotion regulation, academic readiness, school engagement, and related risk factors such as bullying victimization. It is administered to all students in Coalition schools, including a teacher-report form for grades K-12 and a student-report form for grades 3-12. With the support of an assigned regional coordinator from the Coalition, school teams review reports of student emotional and behavioral risk, in which specific domains are color-coded by level of risk (green indicating little attention needed, yellow suggesting brief or less intensive intervention indicated, and red suggesting greater concern and need for individualized intervention). Within the overall Coalition model, EIS-SR administration enables early identification and ongoing surveillance of student behavioral and mental health concerns: individualized reports of student risk allow for students to be matched to appropriate evidence-based interventions, and aggregate reports of patterns at the grade and building level allow administrators to confidently select systems-level interventions to address widespread concerns. The screener is promising for its breadth of content, the usability of its associated reports, and its accessibility, as the screener is freely available

to Coalition schools. This study will examine the EIS-SR in more detail to examine its measurement properties and the translation of its results to decision-making within the Coalition framework.

### **Advanced Methodology in Mental Health Measurement**

In multi-tiered models such as the Coalition, universal screening is the primary pathway for previously unsupported students to be systematically identified and matched to evidence-based interventions. Therefore, highly accurate and efficient screening technology is necessary for identifying needs and selecting students for appropriate interventions. In the interest of developing technically sound screeners and rigorously evaluating existing measures, researchers have applied advanced measurement models, developed in other areas of education, to socioemotional assessment. The following section describes this methodology with specific attention paid to its application to mental health measurement.

#### **Latent Class and Latent Profile Analysis**

Person-centered analytic strategies, as opposed to variable-centered analysis, possess theoretical advantages in the interests of better understanding groupings of unique individuals within populations. These strategies often align well with the goals of researchers (e.g. understanding in more nuance clusters of comorbid risk factors present within participants) as well as community clinicians interested in investigating the unique needs of the population they serve. Latent class analysis (LCA) and latent profile analysis (LPA) are clustering approaches that generate groupings of individuals within a sample based on their similarities across multiple discrete dimensions, identifying latent variables (profiles) that best group participants (Walrath et al., 2004). This facilitates

more detailed understanding of the groupings of observed traits most often present in a population. Although they employ similar estimation methods and the same fit statistics, LCA generates latent groupings from categorical variables, whereas LPA determines clusters of individuals based on continuous variables (Williams & Kibowski, 2016).

In the context of the Coalition’s screening model, which relies on classifying students as “at risk” based on their continuous scores on each subscale, LPA may provide a unique advantage in understanding students’ needs at the systems level. Rather than attempting to triage student needs based on individual subscales flagged as a concern most frequently, school stakeholders can identify groupings of risk (e.g. emotion dysregulation and internalizing problems) that may facilitate more targeted and efficient matching of students to appropriate interventions. In this framework, individual students would also be assigned scores based on their probability of membership in different latent classes, which has the potential to assist stakeholders in triaging students’ complex profiles of risk and needs. These methods also enable examination of covariates of latent classes, which can enable screening and intervention planning informed even further by covariates of socioemotional risk profiles.

### **Item Response Theory**

Item response theory (IRT) is an innovative measurement framework including strategies to evaluate the measurement properties of scales and model respondents’ continuous latent traits (Embretson & Reise, 2000). Historically, IRT was developed within the field of psychometrics in an attempt to a) estimate measurement and respondent properties separately and concurrently, and b) develop tools to evaluate measures independent of their respondents. Many traditional applications of item

response theory involve cognitive or academic latent traits of interest, and this is reflected in much of the traditional language within the IRT framework.

IRT analyses involve the estimation of several parameters to describe individual item functioning, scale performance, and individual respondent latent traits, based on the assumption that all item responses are a function of individual respondent characteristics interacting with item properties. The latent trait of interest is represented by  $\theta$  and can be any trait a test is designed to measure, ranging from algebra ability to depression severity and more. The discrimination parameter,  $a$ , represents the ability of an item to reliably distinguish between respondents along different levels of the latent trait,  $\theta$ . Difficulty ( $b$ ) represents the degree of latent trait severity necessary for an individual to endorse an item or answer it correctly. For example, in modeling mathematics ability, an item with a higher difficulty parameter would require a greater amount of respondent math ability for it to be more likely than not that they answer the problem correctly. Extending this framework to mental health measurement, when modeling latent depression severity, an item asking about suicide ideation may require a greater degree of depression severity for a respondent to be highly likely to endorse it (higher  $b$  parameter), whereas an item assessing occasional disturbances in sleep may require a lower level of latent depression to be likely to be endorsed.

Traditional IRT models evaluated measures with dichotomous item structure (true/false, yes/no) using logistic functions to predict the likelihood that a respondent would endorse the correct or affirmative response option based on their location along the latent trait. However, several commonly-used IRT models have been developed to handle measures with more than two response options, including the generalized partial credit

model (GPCM; Muraki, 1992) and the graded response model (GRM; Samejima, 1970). These models still estimate  $a$  and  $b$  parameters to model latent traits ( $\theta$ ), and their interpretation remains similar. In polytomous models for items with more than one response category,  $b$  parameters are called “threshold” parameters, in which multiple  $b$  parameters represent the degree of latent trait severity necessary to endorse successive response categories. For the purposes of establishing common language among IRT researchers from education to psychology, some authors have advocated for describing  $b$  using the general descriptor “item location” rather than “difficulty” (often describing cognitive ability) or “severity” (often describing an affective trait) and describing respondents’ degree of latent trait severity as “person location” because  $\theta$  and  $b$  are measured using the same scale (de Ayala & Santiago, 2017).

When examining a given measure, candidate IRT models can be compared against one another using common model fit statistics for comparative and absolute fit to the data to determine which is the most appropriate representation of measure and item performance. Furthermore, IRT analyses yield a statistic known as information criteria, which estimates the areas along the latent trait for which a given measure (or individual item) measures most precisely. The information function for each individual item is dependent on its discrimination parameter ( $a$ ) as well as item location ( $b$ ), and test information is simply the sum of each item’s information function. Higher information values indicate greater measurement precision, although, due to its summative dependence on the number of items, it is difficult to establish absolute interpretation criteria. All these can be used to establish a measure’s precision for measuring a given latent construct and its appropriateness for use in different settings.

## **Research Applications to Youth Mental Health**

IRT methodology is not uncommon in educational assessment research, but its application to mental health assessment is relatively recent (Reise & Waller, 2009; Thomas, 2010), and there are even fewer applications to the measurement of youth mental health. In adult populations, IRT has been used to evaluate mental health scale performance (Reise & Waller, 2009), develop computerized adaptive tests to quickly measure symptoms (Sunderland, Batterham, Carragher, Calear, & Slade, 2017), and even develop measures based on IRT pilot studies of item performance (Pilkonis et al., 2011). Especially for researchers, IRT has provided great value to psychometrics and clinical measurement by providing tools to develop precise measures, link latent scores across measures, and estimate item bias across demographic groups – initial studies in youth indicate promising research applications as well (Olinio et al., 2013).

IRT methods have also been applied to reliably estimate the performance of mental health measures (such as depression screening tools) in adolescent populations (Olinio et al., 2012; Olinio et al., 2013). One group of researchers has applied IRT methodology to a school-based socioemotional screener, the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS; Kilgus et al., 2013; Kilgus, von der Embse, Chafouleas, & Riley-Tillman, 2014). One study evaluated the discrimination (*a*) and threshold (*b*) parameters of the SAEBRS to determine the performance of each item and their relationship to measuring youth mental health reliably and appropriately (von der Embse, Pendergast, Kilgus, & Eklund, 2016). Another study fit a bifactor IRT model to estimate item performance in measuring each dimension concurrent with performance measuring an overall latent trait (Kilgus et al., 2018). Similarly, researchers evaluating



the SRSS determined that most items discriminated well and measured expected ranges of the latent trait, but that all items exhibited differential item functioning (DIF, an indication of item bias) across demographic categories (Schatschneider, Lane, Oakes, & Kalberg, 2014). Together, these studies demonstrate the potential for IRT methods to be used specifically for refining school-based youth mental health screening measures.

### **Multidimensional IRT**

A key assumption of traditional IRT analyses is the unidimensionality of the scale in question; that is, in order to estimate a single continuous latent trait,  $\theta$ , the scale should only be designed to measure a single underlying construct. However, in order to handle broadband scales intended to measure multiple latent constructs, multidimensional IRT was developed to simultaneously estimate items' locations along multiple continuous traits (Reckase, 2007). Accounting for multiple dimensions within a scale results in more accurate parameter estimation as well as more detailed information about item contributions to each latent trait (Reckase, 2007). When compared to confirmatory factor analysis, multidimensional IRT yields similar information about a scale's factor structure and item correspondence to proposed latent constructs, but multidimensional IRT provides additional information about individual item functioning and contributions to the scale (Osteen, 2010).

### **Mixture IRT**

An innovative extension of IRT methodology, mixture IRT, involves concurrent modeling of respondent latent traits, item parameters, and latent classes (Rost, 1990; de Ayala & Santiago, 2017). These models yield both dimensional indicators of individuals' scores along a continuous latent trait as well as categorical indicators of membership in

qualitatively different population subgroups, as in latent profile analysis. Identified latent classes can then be examined for their mean scores on different subscales within the measure of interest as well as their associations with other proximal and distal outcomes. In the simplest form of this model, the mixture Rasch model (in which all item discrimination [ $a$ ] parameters are held fixed at 1; Rost, 1990) estimates the probability ( $P$ ) of a given individual ( $i$ ) in a given latent class ( $c$ ) endorsing a given item ( $j$ ) with discrimination ( $a$ ) and difficulty ( $b$ ) parameters using the following logistic expression:

$$p_{ijc} = \frac{\exp\{a_{ijc}(\theta_{ic} - b_{jc})\}}{1 + \exp\{a_{ijc}(\theta_{ic} - b_{jc})\}}$$

As in latent class/profile analysis, latent profiles are assumed to be both mutually exclusive and exhaustive; therefore, mixture IRT models assume that

$$\sum_{c=1}^c \pi_c = 1$$

or that the total of class proportions present in the sample is equal to 1. Finally, using the following equation based on a given respondent's vector of responses to all items on the scale, the mixture IRT yields estimates of the respondent's location along the latent trait  $\theta$  as well as the probabilities of their latent class membership:

$$p(\mathbf{x}_i) = \sum_{c=1}^c \pi_c \cdot p(\mathbf{x}_i|c) = \sum_{c=1}^c \pi_c \left[ \prod_{j=1}^L p_{ijc}^{x_{ij}} (1 - p_{ijc})^{(1-x_{ij})} \right]$$

in which the probability of the response vector ( $\mathbf{x}_i$ ) to all  $L$  items is generalized from their response to item  $j$  and based on their class membership ( $c$ ).

In one study, Finch and Pierson (2011) used mixture IRT to identify subtypes of risky behavior in a youth sample. In addition to identifying four profiles of risk present in the sample, the mixture IRT model estimated which items most efficiently distinguished

between students within each profile. Another study (Kadengye, Ceulemans, & Van den Noortgate, 2014) applied mixture IRT longitudinally to measure differential trajectories of student learning in an online course while across different levels of initial latent content knowledge. This methodology has clear implications for evaluating the EIS-SR and its use in universal screening for several types of mental health risk in large student populations.

Recent research has examined the possibility of estimating multidimensional mixture IRT analyses (Finch & Finch, 2013; Bacci, Bartolucci, Grilli, & Rampichini, 2017). For example, Finch and Finch (2013) modeled multiple academic dimensions on a single test (reading and math) and profiles of student performance along both dimensions in addition to presence or absence of student learning disabilities (this model also included a multilevel component to account for student nesting within schools). In this study, we will attempt to fit a multidimensional mixture IRT model to the EIS-SR data, along with other candidate mixture models, to generate rich information about sample characteristics as well as the measurement properties of the EIS-SR. Each respondent's continuous latent risk ( $\theta$ ) will be estimated for each modeled latent trait (an overall dimension, multiple subscales on the EIS-SR, or both) as well as probabilities of membership in distinct profiles of risk in the sample. For example, one respondent may receive an "at risk" overall risk score that may not indicate an immediate need for intensive services, but they may have high probability of membership in a latent class composed of high internalizing symptoms and low school engagement, and that profile may be associated with high rates of dropout. In this case, taking a student's dimensional

scores and profile membership together may lead to more nuanced, detailed case conceptualization and inform more appropriate intervention selection.

### **Purpose**

The purpose of the current study is threefold: to 1) evaluate the factor structure of the EIS-SR and basic measurement properties of EIS-SR, 2) to attempt to model EIS-SR item functioning with IRT and apply person-centered analyses (either LPA or mixture IRT methodology) to extend advanced methods to youth mental health measurement, and 3) to demonstrate the tenability of harnessing information generated by the mixture IRT model to support school-based decision-making. This study will rely exclusively on the student-report version of the EIS-SR, as it has been previously evaluated for its internal structural properties but not the measurement properties of individual items (Huang et al., 2019; Reinke et al., 2020). Findings are intended to bridge the research-to-practice gap from psychometrics to the provision of school-based mental health services in support of a public health approach to identifying at-risk students and matching them to appropriate mental health supports. Results would be of interest to quantitative researchers for extending advanced methodology to a large, applied sample, to school administrators interested in the measurement precision of an accessible universal screening instrument, and to school mental health professionals interested in leveraging universal screening data to make precise intervention decisions.

Again, our research questions are as follows:

1. What is the dimensionality and factor structure of the EIS-SR in a high school sample?

Previous research has demonstrated the presence of both a 6-factor model (Huang et al., 2019) in a grades 3-8 sample, and a 7-factor model (Reinke et al., 2020) for the EIS-SR in an elementary school sample; however, research has yet to evaluate the dimensionality of the EIS-SR in high school students. This represents a necessary first step in understanding the measure's performance in this population as well as determining appropriate subsequent measurement strategies (IRT, multidimensional IRT, or not appropriate for IRT).

2. Should an IRT model be appropriate for these data, when the EIS-SR is evaluated using a polytomous (and possibly multidimensional) IRT model, do any items warrant removal due to poor performance?

The next step of evaluating the EIS-SR is conducting initial diagnostic analyses to determine item performance in an IRT framework. EIS-SR items have never before been evaluated for their discriminative abilities or item locations. Items were evaluated as part of an iterative process in which they would be flagged for poor performance based on criteria including fit with factor-analytic models as well as individual item performance and evaluated for future inclusion on the EIS-SR based on the degree of poor performance as well as their theoretical relevance to the constructs assessed by the measure.

3. What classes of individuals will emerge in the school-based sample when EIS-SR data are modeled using either a multidimensional mixture IRT model or a latent profile analysis approach?

Should initial measure diagnostics support the use of IRT modeling, several mixture IRT models (including a mixture Rasch model, mixture 2-parameter logistic model, and a

multidimensional mixture IRT model) would be fitted to EIS-SR data and evaluated for comparative fit. The superior model should yield estimates of item functioning as well as profiles of socioemotional risk, which can be further described by examination of mean scores on EIS-SR subscales. Should mixture IRT be determined an inappropriate measurement strategy given the structure and properties of the EIS-SR, latent profile analysis (LPA) would be conducted independent of an IRT framework to generate profiles of risk present within the sample across EIS-SR subscales. EIS-SR profiles were evaluated for proportion of the sample in each class as well as their association with cross-sectional demographic variables (race, gender, income level) and their ability to predict distal outcomes such as office discipline referrals, attendance, or standardized test scores.

## **CHAPTER III: METHODS**

This chapter will review elements of the proposed methodology of the study, including the research design, the measure of interest (the EIS-SR), study procedures, and plans and proposed models for analyzing and organizing data.

### **Research Design**

The current study is quantitative in nature, involving the analysis of existing data collected as part of the ongoing efforts of the Coalition. Analyses will first investigate dimensionality and properties of the EIS-SR assessment used by the Coalition, then attempt to evaluate individual item performance on the EIS-SR. Further person-centered analyses will be conducted to model latent classes of student risk, and, if possible, model latent classes concurrently with statistics on item performance in the prediction of membership to the latent classes.

### **Participants**

#### **Student Respondents**

The sample for this study is drawn from participants who already completed the EIS-SR as part of normal school universal screening operations during their Fall 2018 administration. Because EIS-SR administration is universal, K-12, countywide practice, it qualifies as a “regular educational practice” that does not require additional active consent for research (American Psychological Association, 2017, p. 11). Still, a passive consent procedure was implemented in which parents were sent an informational note home about the screening procedures and given the opportunity to opt their student out. This study focuses on students in grades 9-12. Our sample included responses from 5,860 students, of which 50.56% were female, 72.93% were White, 13.18% were Black, 4.77%

were Asian, 4.01% were Latinx, 4.54% were multiracial, and 30.09% were eligible for free or reduced-price school lunches. With regard to grade level, 28.46% of students were in 9<sup>th</sup> grade ( $n = 1,668$ ), 27.58% were in 10<sup>th</sup> grade ( $n = 1,616$ ), 24.27% were in 11<sup>th</sup> grade ( $n = 1,422$ ), and 19.69% were in 12<sup>th</sup> grade ( $n = 1,154$ ). Participating students were drawn from nine county high schools.

We selected high school students for our sample because of the additional mental health risk associated with the transition to high school (McIntosh et al., 2008) as well as the importance of identifying concerns and connecting youth to accessible mental health services before they enter adulthood and the workforce.

## **Measures**

### **Early Identification System – Student Report (EIS-SR)**

The primary measure of interest for this study is the Early Identification System Student Report (EIS-SR), the universal student social behavioral and emotional screener developed for use by the Coalition. As discussed previously, the EIS-SR is the primary surveillance tool used by the Coalition to identify students in need of additional services. This current study evaluated the EIS-SR (Huang et al., 2019), administered to students in grades 9 through 12.

Items on the EIS-SR measure early signs of internalizing and externalizing concerns as well as general risk factors such as bullying victimization. Item wording was developed in collaboration between researchers and school-based practitioners working with the Coalition. In the interest of capturing student mental health risk efficiently rather than diagnosing existing concerns, mental health items include more common, lower-level symptoms (e.g. “In the past month, I have felt lonely”) in favor of more severe



symptoms (e.g. “I have wished I were dead”) that may only identify the most severe symptoms. This is consistent with previous discussions on the advantages of dimensional conceptualizations of mental health as well as the importance of identifying and addressing subthreshold mental health symptoms. All items are rated on a Likert-type scale with four response options: 0 (“never”), 1 (“sometimes”), 2 (“often”), and 3 (“always”).

Huang and colleagues (2019) found that in a sample of third through eighth grade respondents, 30 of the 41 EIS-SR items loaded successfully onto six factors: peer relations, internalizing, externalizing, relational aggression, emotion regulation, and school engagement. Similarly, Reinke and colleagues (2020) found that, in an elementary school sample, 35 EIS-SR items loaded onto seven factors (subscale descriptions matched the previous six identified subscales, with the addition of an “attention” domain). These factors are consistent with the risk and mental health domains identified during the development of the EIS-SR. Additionally, four items included in the EIS-SR are not part of any one scale but are considered stand-alone risk factors: “Other kids make fun of me at school,” “I am bullied by others,” “I get into fights with others,” and “I am late to school.” However, factor structure of the EIS-SR may vary in this study, as its factor structure has yet to be examined in a high school sample. Furthermore, the sample evaluated by Huang and colleagues (2019) was composed by a large number of students from a single parochial school, whereas the sample in the current study will be primarily composed of students from local public schools. For this study, all 41 EIS-SR items will be evaluated for possible consideration in the full scale and eliminated based on poor loading onto latent factors.

## **Demographics**

Demographic variables, including student sex, race, ethnicity, and free- and reduced-price lunch status, were drawn from existing school records. Additionally, student behavioral and academic outcome data, including total discipline referrals (office discipline referrals, in-school suspensions, and out-of-school suspensions) accrued during the 2018-2019 school year as well as standardized test scores for math and reading, were also obtained from school records.

## **Procedures**

As part of business-as-usual operations of the Coalition, the EIS-SR is administered universally by schools in Boone County three times per year: in fall, winter, and spring. EIS-SR items are administered electronically on computers at the students' schools, and data are stored on a secure server at the University of Missouri. At the close of EIS-SR administration, data are available to school administrators and school-based mental health professionals through an online dashboard that allows them to view teacher- and student-reported data reports. These data are provided to them via school and grade level aggregated data reports, as well as individual reports for students found to have risk. Typical Coalition activities associated with the EIS-SR proceeded as usual, including reviewing of EIS-SR data with teams of Coalition regional coordinators, school mental health professionals, and administrators. Once again, these data serve as the primary surveillance tool for the Coalition to identify students at risk for future socioemotional concerns, students with existing needs for mental health supports, and patterns of risk at the system level that suggest building-wide strategies.

This study uses data from the Fall 2018 EIS-SR administration, and no additional data were collected above the typical data collected as part of the Coalition. Data were cleaned and analyzed to evaluate the dimensionality and measurement precision of the overall EIS-SR for students in grades 9 through 12 and its individual items. Further analyses involved latent profiles of student risk using person-centered analyses.

### **Data Analytic Plan**

All quantitative analyses were conducted in R (R Core Team, 2018) and Mplus (Muthén & Muthén, 2013). Analyses in R relied on the *mirt* (Chalmers, 2012), *lordif* (Choi, Gibbons, & Crane, 2011), and *lavaan* (Rosseel, 2012) packages.

### **Exploratory and Confirmatory Factor Analysis**

Exploratory factor analysis (EFA) was conducted in order to determine the dimensionality of the EIS-SR and to guide selection of appropriate subsequent analyses. The sample ( $n = 5,860$ ) was randomly split into two samples ( $n$  of each = 2,930) to support separately running exploratory and confirmatory factor analysis. Due to the length of the measure (maximum of 41 items), many different factor structures may be investigated; in order to streamline the exploratory process, dimensionality of the EIS-SR was investigated by examination of a scree plot and use of Horn's parallel analysis (Horn, 1965). Based on the results of these analyses, appropriate candidate models were selected and compared by their fit indices, including the Tucker-Lewis Index (TLI), the root mean square error of approximation (RMSEA), and the Bayesian Information Criterion (BIC; Schwarz, 1978). In order to be considered a model with good fit, corresponding TLI values must exceed .90, and RMSEA values must be  $< .08$ , in accordance with fit guidelines suggested by Hu and Bentler (1995). After selection of the strongest model, an

iterative process took place in which items with poor factor loadings (e.g.  $< .35$ ) or cross-loading onto multiple factors were removed, and EFA was re-run using remaining items. Finally, after determination of the factor structure through EFA, confirmatory factor analysis (CFA) was run on the second randomly-selected subsample ( $n = 2,930$ ), specifying the model suggested by the exploratory process. Following the CFA, each factor will be treated as a subscale of the EIS-SR and will be examined for descriptive statistics and scale reliability using categorical omega.

### **Evaluating Item Response Theory Models**

Unidimensionality of scales is an assumption of traditional item response theory models, including traditional polytomous models. Although previous research supports the multidimensionality of the EIS-SR, with six (Huang et al., 2018) or seven (Reinke et al., 2020) previously identified factors, results of the CFA will guide the selection of appropriate IRT models to consider. In the case of unidimensional factor structure, polytomous (rather than dichotomous) IRT models would be fitted to the data, given that EIS-SR items have more than two response categories. These would enable us to visualize and diagnose any items with poor measurement precision or unusual performance in the measurement of overall student socioemotional risk. However, due to the possibility that our data would support the multidimensionality of the scale (as suggested by previous research), we attempted to fit bifactor IRT models will be fitted to the data as well and compared to the unidimensional models. Bifactor models generate estimates of each item's contribution to the overall socioemotional risk as well as individual dimensions of risk. Multidimensional item response theory analyses (mirt) were considered and evaluated as well; however, should the EIS-SR be best described by

a 6- or 7-factor model, current computing resources would limit the ability to accurately estimate mirt parameters (e.g. estimating polytomous mirt parameters across 7 factors concurrently).

Polytomous models such as the Graded Response Model (GRM; Samejima, 1970) and the Generalized Partial Credit Model (GPCM; Muraki, 1992) would be evaluated against one another using comparative fit indices (e.g. the Comparative Fit Index [CFI], Bentler, 1990; the BIC) within the framework suggested by the CFA (unidimensional, multidimensional, or bifactor). Should a superior and appropriate model emerge, we would examine the discrimination ( $a$ ) and difficulty ( $b$ ) parameters for each item generated by the model and visually inspect each item's item response function. Item parameters would be evaluated against the criterion established by de Ayala (2009), including discrimination parameters between .90 and 2.50 and difficulty parameters within ranges of the latent trait appropriate for the measurement method (e.g.  $b$  parameters for a screening measure should not be located 3 standard deviations above the mean). Items that fail to meet these criteria will be inspected for response patterns and contribution to scale content coverage, then considered for possible removal.

Furthermore, should an appropriate model emerge, we will examine the overall information provided by the EIS-SR and the areas along the latent trait(s) for which EIS-SR measurement is most precise. The test information function is dependent on the discrimination parameter ( $a$ ) of each item as well as the probabilities of endorsing successive response categories, which determine where along the latent trait the overall test measurement is most precise. Ranges where test information exceeds the standard error would be considered ranges for which EIS-SR measurement is precise. Should any

items fail to meet adequacy criteria and be recommended for removal at this stage, the model would be run again, at which point individual item parameters will be re-evaluated as well as the information statistic of the overall EIS-SR, with the expectation that measurement precision will increase with removal of poor performing items.

### **Person-Centered Analyses**

Following the identification of an appropriate factor model, person-centered analyses were conducted to examine groupings of latent risk among participants in the sample.

### ***Mixture IRT Modeling***

Should the identified factor model allow for the possibility of IRT modeling, a mixture IRT model (de Ayala & Santiago, 2017) would be fitted to EIS-SR data. A series of candidate models would be evaluated against one another for their comparative fit to the data, which would be evaluated using CFI and BIC (which has demonstrated superiority in estimating fit of mixture IRT models; Li, Cohen, Kim, & Cho, 2009). Possible models include the mixture Rasch model (Rost, 1990), the multidimensional mixture Rasch model (Finch & Finch, 2013), the multidimensional 2-parameter-logistic model (2PL; Finch & Pierson, 2011; Bacci et al., 2017), and traditional latent profile analysis within a factor analysis framework. There are comparative advantages and disadvantages to each model (outlined in Table 1), but for this study we would select a final model using a combination of examining fit statistics and correspondence with the previous research and intended use of the EIS-SR. For the purposes of evaluating the EIS-SR within the mixture framework, polytomous item responses would be coded as binary.

### *Latent Profile Analysis*

Alternatively, should IRT estimation prove not feasible with our current data, latent profile analysis (LPA) would be conducted within a traditional structural equation modeling (SEM) framework to generate clusters of student risk independent from IRT estimation. Total scores of subscales generated following the CFA would be used as continuous indicators for each identified domain of socioemotional risk. Through running multiple candidate models for different total numbers of latent profiles, we would aim to identify the smallest number of latent profiles that accurately account for clusters of observed student risk, as determined by both theoretical considerations and examination of relative fit statistics. Models would be compared using BIC as well as AIC (Akaike Information Criterion; Akaike, 1973), with smaller BIC and AIC values representing better model fit. Other statistics used in comparing candidate models include the entropy statistic (Ramaswamy, DeSarbo, Reibstein, & Robinson, 1993), for which values closer to 1 indicate more precise classification, and the bootstrapped parametric likelihood ratio test (BLRT; McLachlan, 1987; Nylund, Asparouhov, & Muthen, 2007), which compares nested models to determine whether one additional profile leads to significantly stronger fit. Following identification of the best-fitting model, latent profiles would be examined for possible associations with demographic variables using latent class regression analysis (Guo, Wall & Amemiya, 2006), as well as possible associations with continuous academic and behavioral outcomes using the Auxiliary function in Mplus (Muthén & Muthén, 2013).

## CHAPTER IV: RESULTS

This chapter reviews the findings of the current study, organized in order of our proposed Research Questions (RQ).

### **RQ1: Dimensionality and Diagnostics**

First, the Kaiser-Meyer-Olkin measure of sampling adequacy (KMO; Kaiser, 1974) was conducted on the overall sample to determine the suitability of the dataset for factor analysis, revealing the sample was appropriate for factor analysis (KMO = 0.94). Results of the scree plot (Figure 1) suggested a model with three or four factors may be most appropriate; however, results of Horn's parallel analysis suggested a model retaining 16 factors. Therefore, exploratory factor analysis was conducted for three possible numbers of factors: three and four factors (based on results of the scree plot) as well as seven factors (based on Reinke et al., 2020); the model with 16 factors (based on results of Horn's parallel analysis) was excluded in the interest of parsimony and ease of model convergence. Fit statistics for the three-, four-, and seven-factor EFA models examined are presented together in Table 2.

Of the candidate models, the seven-factor model demonstrated the strongest fit (RMSEA = .055, TLI = .867), and was the model with the lowest BIC value (989.21); however, overall fit was not satisfactory (TLI < .90). Because the best-fitting model aligned with the factor structure suggested by Reinke and colleagues (2020) from an elementary school sample, several items were removed from the EIS-SR because the authors believed them to be independent indicators of student risk (e.g. "I have friends to talk to at school," "Other kids make fun of me at school," "I am bullied by others"). Following removal of these seven items (leaving 37 remaining items), the seven-factor



model demonstrated fully satisfactory fit statistics (RMSEA = .048, TLI = .916) and a lower BIC value (-58.29) than previous models. Even after achieving satisfactory fit, five examined items warranted removal for failing to load sufficiently onto any factor ( $< .35$ ); these items were “I have a hard time asking for help,” “I try hard to get good grades on my work,” “I listen to my teachers,” “I complete schoolwork on time,” and “There is an adult I can talk to if I need help.”

Following removal of these items, the seven-factor model was run once more with the remaining 29 items and demonstrated better fit, RMSEA = .041, TLI = .949, BIC = -447.55. Standardized factor loadings for all items are presented in Table 3. Factor names correspond with hypothesized subscales of the EIS-SR from Reinke and colleagues (2020) and include peer relations, internalizing problems, externalizing problems, attention problems, emotion dysregulation, relational aggression, and school disengagement. Correlations between factors ranged from .12 to .47 and can be found in Table 4.

### **Confirmatory Factor Analysis**

Using ordinal confirmatory factor analysis, the seven-factor, correlated factors model developed during the exploratory analyses was replicated using the second half of the sample reserved for CFA ( $n = 2,930$ ). Again, model fit indices indicated acceptable model fit to our data (RMSEA = .044, TLI = .956, CFI = .961). Descriptive statistics, including scale reliability (categorical omega), for all seven subscales are presented in Table 4. All seven scales demonstrated acceptable reliability as measured by categorical omega (.759-.911,  $M = .816$ ).

## **RQ2: IRT Item Diagnostics**

Confirmation of a seven-factor model, unfortunately, precludes further investigation within an IRT framework. Using currently available computing resources along with our current sample size ( $n = 5,860$ ), estimating a multidimensional IRT model is impossible<sup>1</sup>. Estimation of item difficulty and discrimination parameters is possible with more traditional polytomous IRT models (including the Graded Response Model or the Generalized Partial Credit Model) if treating each subscale as its own unidimensional scale; however, this approach would lead to biased and inflated results (see Monroe & Cai, 2015). Furthermore, subscales on the EIS-SR range in length from two to nine items ( $M = 4.14$ ), limiting the utility of methods intended to increase the precision of individual subscales.

## **RQ3: Person-Centered Analysis (Latent Profile Analysis)**

Although it was not possible to conduct mixture item response theory analyses, latent profile analysis was still an available method to estimate clusters of symptoms present in the student sample. The appropriate latent profile model was selected using a combination of consideration of fit indices (AIC, BIC, entropy, and BLRT), theory, and usability (e.g. uniqueness of profiles, size of profiles). Six separate models were generated (estimating two through seven latent profiles) using total scores from the subscales generated during CFA as continuous indicators. Following removal of missing values, the sample size used for LPA analyses was  $n = 5,555$ .

Fit statistics, including AIC, sample size-adjusted BIC, entropy, and BLRT, are presented in Table 5. The five-profile model was selected as the optimal model based on

---

<sup>1</sup> Multidimensional IRT estimation was attempted in R using a 7-factor model. After 32 minutes, the model failed to generate meaningful fit statistics (TLI = -3.30, CFI = 0, RMSEA = 0.59).

offering superior fit to the two-, three-, and four-profile models (based on lower AIC and BIC values and a significant BLRT value, indicating significantly improved fit when compared to the four-factor model) as well as offering clinically meaningful, unique profiles. Although potentially offering increasingly strong fit compared to the five-factor model (also based on AIC, BIC, and BLRT), the six- and seven-profile models did not generate additional clinically unique profiles, and thus the five-profile model was selected for parsimony and usability. Table 6 presents the average probabilities of most likely latent class membership by latent profile, and Table 7 presents the estimated proportions of profile membership in the sample.

Figure 2 represents the results of the five-profile model, demonstrating mean scores on each subscale of the EIS-SR. Mean scores are presented as z-scores for ease of interpretation across subscales of varying lengths. Profile descriptions are based on mean levels on each continuous indicator. Profile 1 ( $n = 2501$ , 45.02%) involved below-average scores across all seven indicators of socioemotional risk, and is described as the “High Wellness” profile. Similarly, Profile 2 ( $n = 1962$ , 35.33%) members’ scores across all domains were near the mean ( $<0.34$  SD discrepant), so it is described as the “Average Wellness” profile. Profile 3 ( $n = 442$ , 7.95%) members’ scores exceeded the mean across all seven domains, but notably higher scores were observed on the Internalizing Problems ( $+2.14$  SD), Attention Problems ( $+1.20$  SD), and Emotion Dysregulation ( $+1.30$  SD) subscales; thus, this profile is described as the “Distress/Dysregulation” profile, as members endorsed significant anxiety and depression risk in addition to overall distractibility and difficulty managing their emotions. Likewise, Profile 4 ( $n = 612$ , 11.04%) was associated with above-average mean scores across subscales; however,

scores were especially elevated in the areas of Externalizing Problems (+2.66 SD) and Relational Aggression (+1.06 SD). Profile 4 is described as the “Externalizing Risk” profile. Finally, Profile 5 ( $n = 37$ , 0.67%) was associated with significantly elevated scores across all subscales, but especially Externalizing Problems (+8.76 SD) and Relational Aggression (+4.14 SD), and it is described as the “Very High Risk” profile. It is worth noting that the small size of this profile (0.67% of the sample) relative to the other profiles; however, a profile with these characteristics was identified in all LPA models with at least three profiles, and thus it is retained in our model.

### **Latent Class Regression**

Latent class regression analyses were examined to determine whether the observed latent classes in the five-profile model correspond to demographic covariates as well as academic performance and behavioral outcomes. Ethnicity could not be examined as a covariate as too few students in the current sample were identified as Hispanic/Latinx ( $n = 25$ ).

### ***Demographic Covariates***

Possible relationships between class membership and race, sex, grade level, disability status, and free-and-reduced price lunch status (FRL) were investigated. The High Wellness profile was used as a reference group as this profile represented the least socioemotional risk as well as 45% of the sample. When compared with the High Wellness profile, members of the Average Wellness profile were more likely to qualify for FRL (odds ratio [OR] = 1.52, confidence interval [CI]: 1.38-1.67,  $p < .001$ ), yet no further significant differences emerged. Compared with the High Wellness profile,

members of the Distress/Dysregulation profile were significantly more likely to be male (OR = 0.28, CI: 0.24-0.32,  $p < .001$ ), more likely to be in 12<sup>th</sup> grade compared with 9<sup>th</sup> grade (OR = 0.62, CI: 0.51-0.74,  $p = .001$ ), and more likely to qualify for FRL (OR = 2.74, CI: 2.32-3.15,  $p < .001$ ). Members of the Externalizing Risk profile (compared with High Wellness) were more likely to be female (OR = 2.94, CI: 2.49-3.39,  $p < .001$ ), more likely to be in 9<sup>th</sup> grade compared with 12<sup>th</sup> grade (OR = 1.67, CI: 1.41-1.94,  $p = .011$ ), and more likely to qualify for FRL (OR = 2.92, CI: 2.54-3.31,  $p < .001$ ). Finally, no statistically significant demographic covariates emerged when comparing the Very High profile with High Wellness; this may be related to low statistical power associated with the low sample size of the Very High profile ( $n = 38$ ). Across the five profiles, when using the High Wellness profile as a reference group, no statistically significant relationships were observed between profile membership and race or disability status.

### ***Academic and Behavioral Outcomes***

Members of profiles associated with higher levels of socioemotional risk went on to experience higher levels of office discipline referrals ( $\chi^2 = 146.14$ ,  $p < .001$ ), in-school suspensions ( $\chi^2 = 91.50$ ,  $p < .001$ ), and out-of-school suspensions ( $\chi^2 = 55.12$ ,  $p < .001$ ) throughout the course of their school year. Statistical significance for pairwise comparisons was evaluated at the  $p < .01$  level to protect against possible Type 1 error. With regard to office discipline referrals (ODRs), members of the High Wellness profile ( $M = 0.83$ ) had significantly fewer referrals compared with members of the Average Wellness profile ( $M = 1.29$ ,  $p = .001$ ), the Distress/Dysregulation profile ( $M = 1.68$ ,  $p < .001$ ), the Externalizing Risk profile ( $M = 5.51$ ,  $p < .001$ ), and the Very High profile ( $M = 8.31$ ,  $p < .001$ ). Differences in ODRs between the Average Wellness and

Distress/Dysregulation profiles were not significant ( $p = .129$ ); however, members of both these profiles had significantly fewer ODRs compared with both the Externalizing Risk profile ( $ps < .001$ ) and the Very High profile ( $ps < .001$ ). No significant differences in ODRs were observed between the Externalizing Risk and Very High profiles ( $p = .093$ ). With regard to in-school suspensions, a similar pattern of pairwise differences emerged, although limited significant differences were observed between members of the High Wellness and Average Wellness profiles ( $p = .028$ ). Finally, with regard to out-of-school suspensions, less significant differences emerged between the High Wellness ( $M = 0.04$ ), Average Wellness ( $M = 0.07$ ), and Distress/Dysregulation profiles ( $M = 0.08$ ; HW vs. AW  $p = .024$ , HW vs. DD  $p = .061$ , AW vs. DD  $p = .504$ ). Still, significant differences remained between the aforementioned three profiles and the Externalizing Risk profile ( $ps < .001$ ), and significant differences were observed between the two wellness profiles and the Very High profile ( $p = .004$  and  $.007$ , respectively). All means and pairwise comparisons on these outcomes can be found in Table 8.

Similarly, students in profiles associated with lower levels of socioemotional risk went on to score higher on end-of-year math ( $\chi^2 = 116.76, p < .001$ ) and reading assessments ( $\chi^2 = 89.25, p < .001$ ). In the area of math, members of the High Wellness profile ( $M = 63.38$ ) received higher scores compared with the Average Wellness profile ( $M = 57.66, p = .001$ ), the Distress/Dysregulation profile ( $M = 54.35, p = .001$ ), the Externalizing Risk profile ( $M = 42.43, p < .001$ ), and the Very High profile ( $M = 37.53, p < .001$ ). Differences in math performance between the Average Wellness and Distress/Dysregulation profiles were not significant ( $p = .251$ ), yet both profiles had higher math scores compared with both the Externalizing Risk ( $ps < .001$ ) and Very High

profiles ( $p = .001$  and  $.008$ , respectively). With regard to reading, fewer pairwise differences between profiles were considered significant. Members of the High Wellness profile ( $M = 48.40$ ), Average Wellness profile ( $M = 47.17$ ), and Distress/Dysregulation profile ( $M = 48.23$ ) all had significantly higher reading scores compared with the Externalizing Risk profile ( $M = 33.18$ ,  $ps < .001$ ). Remaining pairwise comparisons for reading did not reach the  $.01$  significance threshold. Again, all means and pairwise comparisons can be found in Table 8.

## CHAPTER V: DISCUSSION

The current study aimed to examine the dimensionality of the EIS-SR, determine its suitability for item response theory analysis, and conduct person-centered mixture analysis to examine profiles of student socioemotional risk present within the high school population.

### Factor Structure

Results of our analyses revealed a seven-factor model fit to 29 EIS-SR items represents the best possible fit of our data. The seven-factor model is consistent, both in number and in groupings of item content, with the model described by Reinke and colleagues (2020) that relied on an elementary school sample. The seven-factor model was confirmed in an independent confirmatory sample, and all seven subscales suggested by the factor structure demonstrated acceptable levels of internal consistency. However, two notable differences emerged between the factor structure in the current study and the factor structure from Reinke and colleagues (2020). First, five items that successfully loaded onto the model described by Reinke and colleagues (2020) failed to load sufficiently onto any factor in our model. These items were “I have a hard time asking for help,” “I try hard to get good grades on my work,” “I listen to my teachers,” “I complete schoolwork on time,” and “there is an adult I can talk to if I need help.”

The performance of three of these items (“I try hard to get good grades on my work,” “I listen to my teachers,” and “I complete schoolwork on time”) in the high school sample may be more reflective of developmentally normative decreasing school engagement in adolescence (Kinderman, 2007). In other words, for students in this age group, diminished interest in school may not necessarily correspond with greater risk for



socioemotional difficulties, and thus may not load well onto factors designed to assess student risk. The failure of these items to load onto the School Disengagement subscale of the EIS-SR may reflect conceptual differences between the emotional components of school engagement (see Fredricks, Blumenfeld, & Paris, 2004) measured by the School Disengagement subscale (e.g. “I enjoy coming to school”), whereas the poorly-loading items reflect behavioral components of school engagement (e.g. “I complete schoolwork on time”). In this case, perhaps the emotional components of school engagement measured by the School Disengagement subscale may be more associated with overall socioemotional risk, and thus a more meaningful construct captured by the EIS-SR. Further replication of the current factor model across samples and age groups is necessary to determine whether these items truly warrant removal from the EIS-SR.

Furthermore, two items loaded to different factors in our model from the factors onto which they loaded in the model from Reinke and colleagues (2020). Specifically, the item “I blame others for my mistakes” loaded onto the Relational Aggression subscale in our study, although it had previously loaded onto the Externalizing Problems subscale in an elementary sample. Conceptually, this item could be justifiably included on either subscale, as they measure a similar valence of symptom content. The shift of “blaming others” from Externalizing Problems (which primarily measures conflict with adult authority figures, including “I disrupt class” and “I get in trouble at school”) in the elementary sample to Relational Aggression (which primarily measures peer conflict, including “I make fun of others” and “I talk about others behind their back”) in the high school sample may be reflective of typical adolescent development. Peer relationships become increasingly salient as youth enter adolescence (Brown & Larson, 2009), and

these relationships may change quickly and be fraught with conflict (Poulin & Chan, 2010), which may contribute to high school students increasingly interpreting “blaming others” as a peer-related act of relational aggression rather than an adult-related component of conflict with authority figures.

The item “I need help with my emotions” loaded onto the Internalizing Problems subscale in the current study, whereas it previously loaded onto the Emotional Dysregulation subscale in an elementary school sample. Again, this may be reflective of changing connotations of the word “emotions” across development. Prior research has found that younger children develop a rich emotional vocabulary by age 6 that typically includes “all” emotions, including anger, happiness, sadness, and fear (Ridgeway, Waters, & Kuczaj, 1985). However, as children enter adolescence, they experience corresponding changes in neurological and hormonal functioning that result in increased self-consciousness and sensitivity to social status, as well as an increased cognitive capacity to examine and reflect on one’s own internal emotional states (Steinberg, 2005). This may result in respondents increasingly distinguishing “my emotions” as internal states (such as sadness or worry) from one’s external behaviors (such as angry outbursts), producing the observed pattern of responding in our high school sample. Internalizing problems such as depression have a typical age of onset in early adolescence (Merikangas et al., 2010), which may also contribute to increased salience of internalizing difficulties when considering emotional problems.

This is the first study that examined the factor structure of the EIS-SR in a secondary sample, and we specifically examined its performance with high-school-aged students. Although prior research has examined the consistency of factor models of self-

report measures across age groups using strategies such as measurement invariance (see Fonseca-Pedrero, Sierra-Baigrie, Lemos-Giraldes, Paino, & Muniz, 2012), many measures are designed for siloed age groups (e.g. the Youth Self-Report [YSR] is designed for respondents ages 11-18), often capturing only adolescent respondents. Use of the same EIS-SR form for students ages 8-18 may have challenges due to developmentally different interpretation of item content, but it supports simplicity of implementation and interpretation district-wide, as well as tracking student scores across age and school building transitions.

Finally, results of the EFA and CFA provide additional data by which to examine the EIS-SR under the criteria suggested by Glover and Albers (2007) for evaluating screening measures. Replication of the seven-factor model across subsamples provides additional evidence for the appropriateness of the intended usage of the EIS-SR as a broadband screener for a variety of separate socioemotional concerns. These factors were found to be correlated at levels that suggest they measure related yet discrete constructs (.12 to .47). As for technical adequacy, our results revealed strong internal consistency of each identified EIS-SR subscale (omega values all  $> .75$ ), as well as evidence for structural validity based on the content of discovered factors and their congruence with proposed EIS-SR subscales. Ultimately, the EFA also resulted in a shorter and thus more usable screener, as 29 of the original 41 candidate items loaded sufficiently onto given factors. Further research is warranted to determine whether permanent removal from the EIS-SR is appropriate; however, the 29-item EIS-SR would also meet a commonly recommended criterion for screener brevity as it has fewer than 30 items (e.g. Levitt, Saka, Romanelli, & Hoagwood, 2007).

## Latent Profile Analysis

Results of the LPA suggested a five-profile model yielded superior fit compared to two-, three-, or four-profile solutions, and examination of mean scores on EIS-SR subscales suggested greater parsimony and theoretical fit for the five-profile model when compared with the six- or seven-profile solutions. Again, this solution suggested the following five profiles: High Wellness (below-average scores on all risk domains), Average Wellness (average scores on all domains), Distress/Dysregulation (high Internalizing Problems and above-average Attention Problems and Emotion Dysregulation), Externalizing Risk (high Externalizing Problems and above-average Relational Aggression), and Very High Risk (significant elevation across risk domains). In our study, students who qualified for FRL were significantly more likely to be members of each risk profile other than High Wellness. This is consistent with prior research suggesting that adolescents with lower socioeconomic status, experiencing associated challenges, are two to three times as likely to develop mental health difficulties (Reiss, 2013), including both internalizing (Wight, Boticello, & Aneshensel, 2006) and externalizing (Astone-P'Olak et al., 2009) problems.

Together, the High Wellness and Average Wellness profiles comprised 80% of our sample. This aligns precisely with MTSS models that propose that approximately 80% of students are not currently demonstrating risk for socioemotional concerns, whereas an additional 15% of students would benefit from early intervention and a final 5% require individualized, intensive intervention (see Sugai & Horner, 2002, p. 38; Schanding & Nowell, 2013). This lends further support to the theoretical validity as well as practical utility of our current LPA model, as it is identifying appropriate proportions

of students as candidates for early intervention. Furthermore, members of the High Wellness profile, on average, had fewer office disciplinary referrals and higher end-of-year math scores compared with members of other profiles, suggesting that their mental well-being corresponds with academic and behavioral success; this lends further support to the construct validity of the Wellness profile. Members of the High Wellness profile were also significantly more likely to be male. This may be partly attributable to gender differences in the prevalence of some mental health difficulties, including anxiety and depression, that typically emerge during adolescence (Merikangas et al., 2010), as well as gender differences in survey response style (see Gilman et al., 2008).

Students in the Externalizing Risk profile report increased disruptive behaviors as well as increased school-based consequences (e.g. “I get in trouble at school”). Additionally, these students report elevated Relational Aggression, including making fun of others and blaming others for mistakes. Members of this profile report average levels of Internalizing Problems and typical levels of risk (less than one standard deviation from the mean) across the remaining socioemotional domains; therefore, this profile represents students struggling with fairly straightforward difficulties with disruptive behavior at school. In our sample, members of the Externalizing Risk profile were more likely to be female compared to members of the High Wellness profile (OR = 2.94). This is in contrast with previous research suggesting that male adolescents tend to exhibit higher levels of externalizing symptoms compared with females (McLaughlin, Hilt, & Nolen-Hoeksema, 2007). This gender effect may be reflective of the elevated relational aggression present in these students, which is typically considered more common among female adolescents (Espelage, Mebane, & Swearer, 2004), although other studies have

found no differences between males and females in rates of relational aggression (Kuppens et al., 2008). Further research is necessary to replicate these findings and further investigate whether this observed profile is generalizable to understanding gender distribution of externalizing risk in adolescents. Appropriate interventions for these students may include Check-In/Check-Out (Todd, Campbell, Meyer, & Horner, 2008), Check and Connect (Anderson, Christenson, Sinclair, & Lehr, 2004) individualized behavior intervention plans, and other environmental modifications that may address the function of their disruptive behavior at school.

The Very High Risk profile represents students with extremely elevated scores across all risk domains, especially Externalizing Problems and Relational Aggression. Comparatively, this profile offers minimal utility in terms of screening for student risk, as it accounts for only 37 students (0.67%) distributed across nine different schools. The extremely elevated scores across all EIS-SR domains suggest these students are either demonstrating a negatively biased response style (raising concerns about the validity of these scores), or these students are demonstrating such significant externalizing and aggressive symptoms that they are likely to have already been identified by school staff as in need of further intervention without the use of a screening tool.

Finally, the Distress/Dysregulation profile highlights a subpopulation of at-risk students with unique and complex intervention needs. Again, this profile involved very high levels of Internalizing Problems, as well as elevated Attention Problems and Emotion Dysregulation, and it represents an estimated 7.95% of the student population. This cluster of symptoms is not uncommon in youth, and this profile can be conceptualized clinically in several different ways.

## **Profiles of Distress and Dysregulation**

First, given the significantly higher levels of internalizing symptoms relative to the other problem domains, it is possible that youth in this profile are experiencing attention problems and emotion dysregulation as secondary consequences of anxiety or depression. In children and adolescents, irritability is often a symptom of depression, even in the absence of low or sad mood (Goldman, 2012), and overall difficulty tolerating emotional distress is a key risk factor for anxiety (Keough, Riccardi, Timpano, Mitchell, & Schmidt, 2010), both of which may be captured in items on the Emotion Dysregulation subscale (e.g. “I get mad easily,” “I get crabby and irritated easily”). Furthermore, difficulty concentrating is often a symptom of either anxiety (Nail et al., 2015) or depression (Crowe, Ward, Dunnachie, & Roberts, 2006), resulting from the increased cognitive load on minds occupied by depressive or anxious thoughts (Rose & Ebmeier, 2006; Eysenck & Calvo, 1992); this functional impairment may be captured by elevated Attention Problems (e.g. “I have trouble finishing my work,” “I have trouble focusing in school”). One study found irritability and difficulty concentrating to be the two most frequently endorsed symptoms of adolescent depression, rated more consistently across participants than feelings of sadness or anhedonia (Crowe et al., 2006). More broadly, anxiety and depression are both associated with significant functional impairment in youth, even when symptoms occur at subthreshold levels (Balasz et al., 2013). Given this conceptualization, the EIS-SR may be identifying students with significant levels of undetected internalizing symptoms that are already reaching impairing levels. For these students, referrals to outside providers or access to school-based cognitive-behavioral treatments may be most appropriate.

Alternatively, this profile may be reflective of comorbidity across several possible diagnostic categories. For example, children with Attention-Deficit/Hyperactivity Disorder (ADHD) often experience academic, behavioral, and social impairment from a young age (Wilens, Biederman, & Spencer, 2002), resulting in frequent negative feedback from adults and experiences of failure, often leading to subsequent negative self-concept and possible development of depressive symptoms (Cole, Jacquez, & Maschman, 2001; Daviss, 2008). In this case, depressive symptoms may develop secondary to concerns with inattention and poor executive functioning that inhibit task performance (Herman et al., 2007; Herman et al., 2020). Indeed, estimated rates of comorbidity between ADHD and major depressive disorder range from 12 to 50% (Angold, Costello, & Erkanli, 1999). Similarly, ADHD symptoms may contribute to feelings of anxiety, as individuals with ADHD are often easily overwhelmed due to their difficulty cognitively organizing and breaking down large, complex tasks (Brown, 2001; Schatz & Rostain, 2006). This, combined with the competency-feedback pathway present in the ADHD/depression model, may contribute to increased risk for anxiety, and in fact, estimated comorbidity between ADHD and anxiety disorders ranges from 15 to 35% (Pliszka, Carlson, & Swanson, 1999). For these students, appropriate intervention would primarily involve environmental accommodations to support their academic and behavioral performance, as well as possible supplementary short-term cognitive-behavioral treatment to target their negative self-concept.

Finally, students with this profile of socioemotional risk may also be conceptualized primarily as experiencing underlying emotion dysregulation and distress tolerance difficulties, which may contribute to increased internalizing symptoms (Keough



et al., 2010) and broad distractibility (Eysenck & Calvo, 1992). Temperamental difficulties with emotion regulation are considered risk factors for the development of further psychopathology, including anxiety and depression (Keenan, 2000; McLaughlin, Hatzenbuehler, Mennin, & Nolen-Hoeksema, 2011). Emotion dysregulation is also considered a core, oft-overlooked feature of ADHD for many children, reflective of emotional impulsivity (i.e. low frustration tolerance) and corresponding deficits in behavioral inhibition (Barkley, 2011). These children may benefit from a range of interventions, including cognitive-behavioral training in coping strategies, environmental modifications and routines to minimize frustration, and distress tolerance strategies borrowed from dialectical behavior therapy (DBT; Linehan, 1993), which directly target emotion regulation and have demonstrated effectiveness in reducing oppositional behaviors among irritable and dysregulated adolescents (Nelson-Gray et al., 2006).

Further research and replication is necessary to investigate the relationship between this cluster of symptoms and gender. In our study, males were significantly more likely than females to be members of this profile. However, previous research suggests females exhibit higher rates of internalizing symptoms beginning in adolescence (Merikangas et al., 2010), in addition to greater levels of overall emotional dysregulation and corresponding anxiety (Neumann, van Lier, Gratz, & Koot, 2010; Bender, Reinholdt-Dunne, Esbjorn, & Pons, 2012). With regard to attention difficulties, although ADHD is more commonly diagnosed in males (Robison, Skaer, Sclar, & Galin, 2002), preliminary research suggests females with ADHD have lower self-efficacy and fewer appropriate coping strategies, whereas their male counterparts are more likely to experience higher levels of externalizing symptoms (Rucklidge & Tannock, 2001). Again, additional

research is necessary to clarify the relationship between these symptoms and the pathways through which they develop.

Given the multitude of possible clinical conceptualizations, idiographic follow-up assessment is necessary to determine primary sources of functional impairment and corresponding appropriate interventions for students in this profile. The emergence of this profile highlights an advantage of using LPA to examine schoolwide screening results, as it supports the identification of students with comorbid concerns that may otherwise be overlooked.

### **Academic and Behavioral Outcomes**

In our sample, members of profiles associated with increasing levels of overall socioemotional risk (all profiles except High Wellness), particularly increasing externalizing symptoms, experienced increasing rates of ODRs, in-school suspensions, and out-of-school suspensions. This pattern of self-reported externalizing symptoms predicting end-of-year behavioral outcomes lends support to the validity of EIS-SR student report scores about externalizing symptoms. Students in the Externalizing Risk and Very High profiles both reported higher levels of these symptoms and went on to experience corresponding high rates of exclusionary discipline at school.

Across behavioral outcomes, members of the Average Wellness profile did not differ significantly from members of the Distress/Dysregulation profile (e.g. ODRs,  $M_s = 1.29$  and  $1.68$ , respectively,  $p = .129$ ), despite members of the Distress/Dysregulation profile reporting greater difficulty with attention problems and emotion dysregulation, which could lead to disciplinary issues at school. This suggests that relying on traditional disciplinary outcomes such as ODRs to screen for students in need of support may result

in overlooking students reporting significant socioemotional needs across multiple domains, as “average” students are virtually indistinguishable from members of this Distress/Dysregulation profile on this metric. This highlights the value of systematic mental health screening over reliance on disciplinary metrics to identify students in need of mental health services (Scott et al., 2009; Sugai et al., 2000; Rocque, 2010). Further, the individuals making disciplinary referrals (i.e. teachers or administrators) may have difficulty detecting the difficulties reported by members of the Distress/Dysregulation profile, as evidenced by their similarity to students with Average risk profiles. On one hand, this data supports the continued use of self-report measures to identify students with risk of internalizing symptoms. On the other hand, this may suggest a negative cognitive bias in students with significant internalizing difficulties that leads to overestimation of their own attention problems and emotion dysregulation, which may explain why they receive typical levels of disciplinary referrals (Rheingold, Herbert, & Franklin, 2003; Lee, Hankin, & Mermelstein, 2010). Additional research is necessary to distinguish the reason these students’ reports of elevated emotion dysregulation and attention difficulties do not correspond with notably elevated disciplinary referrals.

With regard to academic outcomes, again, students experiencing lower levels of overall socioemotional risk and lower levels of externalizing symptoms at their fall EIS-SR screening achieved higher math scores on end-of-year assessments, and students with lower rates of externalizing symptoms achieved higher reading scores compared with students in high-externalizing profiles (Externalizing Risk and Very High). This is consistent with prior studies that demonstrate that even lower levels of mental health symptoms can interfere with academic functioning and contribute to academic

impairment (Nail et al., 2015; Reinke et al., 2008). Again, this lends support to the strength and clinical utility of the chosen five-profile solution, as clinically significant differences emerged on meaningful outcomes among all profiles.

### **Latent Profile Analysis in School Screening Research**

Several other studies have examined widely-used school socioemotional screening measures using latent class or latent profile analysis. Kilgus and colleagues (2015) used LPA to demonstrate the utility of the SABRS in distinguishing between students at-risk and not at-risk specifically for academic as well as behavioral difficulties, rather than distinguishing students as generally at-risk on an overall factor. However, in this study, approximately 45% of participants were members of the “at-risk” profiles, and thus the authors advocated for the use of established cut scores rather than LPA results for use in screening decisions. In another study, researchers examined the BESS using LCA to examine distribution student risk across four socioemotional factors (Kim & Kamphaus, 2013); a four-class solution was selected, and classes were described as comprised of students with minimal risk across all factors, average risk across factors, and medium or high risk on Personal Adjustment alone. LPA has also been conducted to examine trajectories of risk longitudinally (e.g. Felix et al., 2019). One advantage of the current study is its utilization of a measure spanning seven factors, enabling detailed modeling of patterns of student socioemotional concerns within meaningful latent profiles.

### **Multidimensional IRT in Screening Research**

Based on the results of the EFA and CFA, which determined a seven-factor model best described the EIS-SR, IRT methods were not available to analyze our data, as we did

not have access to the computing capabilities or extremely large sample required to concurrently estimate polytomous parameters for 29 items across seven subscales.

As discussed in Chapter 2, IRT methodology offers a number of meaningful and largely unexplored applications to youth mental health measurement research and practice (Reise & Waller, 2009; Thomas, 2010). However, even given a relatively large sample size ( $n = 5,860$ ), IRT analyses were not possible. Our study highlights a number of challenges present in advancing the use of advanced measurement strategies in education and psychology research. First, large sample sizes are typically needed to conduct IRT, and significant computing power may be necessary for estimating multidimensional models with larger samples (Yao & Boughton, 2007, p. 90), requiring active partnerships between school district partners (with access to large samples of respondents) and research institutions (with access to statistical expertise and hardware and software resources). Additionally, measures selected for further IRT research must have appropriate dimensionality (e.g. fewer factors or bifactor structure) to facilitate the feasible estimation of IRT parameters. Practically, this demonstrates a challenging tension for the advancement of applied IRT research in mental health screening: candidate measures must have sufficiently few dimensions to enable IRT estimation, as well as sufficiently broad target constructs to support justification of schoolwide screening efforts to administrators and collection of large samples of data.

This roadblock is certainly surmountable yet requires careful planning across academic disciplines and thoughtful collaboration with school stakeholders. Several available school mental health screening measures have been identified as having four or fewer factors, including the BESS (Harrell-Williams, Raines, Kamphaus, & Dever,

2015), the SAEBRS (von der Embse et al., 2016) and the SRSS (Fredrick, Drevon, & Jervinsky, 2019). Further, many schools have identified narrow-band mental health issues they wish to target with universal screening efforts, such as district-wide suicide risk screening (Aseltine & DeMartino, 2004; Hallfors et al., 2006), although this is not without controversy (Scherff, Eckert, & Miller, 2005; Fox, Eisenberg, McMorris, Pettingell, & Borowsky, 2013). Researchers interested in advancing applications of IRT to youth mental health screening may be tasked with demonstrating the potential value of IRT methods to potential collaborators, identifying community partners committed to high-quality universal screening for mental health concerns, and selecting dimensionally-simple measures that still meet stakeholders' needs. Although the goal of this project was to demonstrate exciting potential uses of IRT analyses, the ultimate goal of applied research in school mental health should be supporting students and the school professionals tasked with identifying and helping them; statistical analyses selected should follow from this goal, and not vice versa.

### **Limitations**

Several limitations to the current study are worth noting. With regard to the factor analysis, a significant number of items (five) were found to fit poorly with our seven-factor model that had previously demonstrated good fit in an elementary school sample, and two items loaded onto different factors from those demonstrated in a prior study (Reinke et al., 2020). Although our CFA in an independent sample confirmed the fit of a seven-factor model with only 29 items, the dimensionality of the EIS-SR and corresponding item fit requires further replication across additional independent samples. Additionally, as this is the first study that examined the performance of the EIS-SR in a

high school sample, further research on the EIS-SR across age groups is necessary to distill which changes in item performance are attributable to the “true” overall factor structure of the EIS-SR across age groups rather than developmental differences in student response patterns.

With regard to the chosen solution to our LPA, the entropy value of the five-profile suggested the least precise fit among the examined solutions from two to seven possible profiles. This solution was selected over lower-profile solutions because the AIC and BIC values suggested stronger fit compared with the two-, three-, and four-profile solutions, and the BLRT suggested significantly greater fit compared with the lower nested model (four profiles). Nylund and colleagues (2007) suggest that BIC and BLRT indices are the most reliable indicators of latent class fit, and thus we considered the five-profile solution better-fitting compared with lower-profile solutions. However, the six- and seven-profile solutions also had lower AIC and BIC values compared with the five-profile solution, and the BLRT values suggested they each contributed additional information when compared with the lower-profile nested solutions (including our chosen five-profile solution). The five-profile solution was ultimately considered a better theoretical fit compared with the six- or seven-profile solutions. The additional profile generated by the six-factor solution appeared to demonstrate the same pattern as the Externalizing Risk profile found in the five-profile solution with slightly lower mean values on Externalizing Problems (+2.87 SD) and Relational Aggression (+1.83 SD), suggesting that this profile captured similar students as the Externalizing Risk profile in our five-profile solution with slightly less severe symptoms. The seven-profile solution generated one additional profile with average mean scores on all indicators except Peer

Relations (+4.02 SD) and School Difficulties (+2.04 SD); however, this profile was considered less usable due to small sample size ( $n = 56$ , 1.01%). Therefore, selection of the five-profile solution was ultimately based on usability and theoretical considerations as well as parsimony, but it is worth noting that the lower entropy value would suggest the least precise fit of the candidate models.

Difficulty selecting an appropriate LPA solution due to conflicting fit indices may also limit the successful use of this methodology in applied settings such as school-based teams, as stakeholders may be hesitant to make intervention decisions based on profile solutions with mixed or confusing justification. On the other hand, this illustrates the importance of conducting screening efforts within a broader problem-solving framework that includes both problem identification and subsequent problem analysis, in which identified students undergo more detailed assessment before being matched to interventions (Tilly, 2008; Walker et al., 2005). Establishing a thorough protocol for problem analysis following universal screening may relieve some of the burden to justify selection of a profile solution, as efficient identification of students should result in idiographic assessment for intervention planning regardless.

### **Future Directions**

#### **Applications of Person-Centered Analyses to School-Based Screening**

The current study represents a demonstration of applied uses for person-centered modeling, suggesting several immediately available applications. Person-centered analytic strategies, such as latent profile analysis, are likely more useful to school administrators in planning for intervention decisions at the system level when compared with typical descriptive statistics. Ideally, administrators reviewing a school's mental



health screening data will also be tasked with using that data to plan for intervention efforts for identified students, including planning for intervention selection, staff and resource allocation, and formation of small groups of similarly at-risk students. This is no simple task, especially when utilizing a broadband screening measure such as the EIS-SR which attempts to screen for a range of socioemotional concerns that correspond with a wide range of potentially appropriate interventions. Traditionally, when reviewing these data, administrators generate or are provided with descriptive statistics, including school- or grade-level means for each risk domain. Ultimately, however, administrators must make intervention decisions about *individuals*, including whom to assign to what interventions, rather than indicators. Even if administrators select priority domains they hope to target with interventions, or if they hope to use data to triage areas of highest need, examining each domain in isolation can obscure meaningful comorbidities and groupings of symptoms that might suggest more effective interventions for each student.

For these reasons, running an analysis such as LPA at the district level or school level may support decision-makers in organizing intervention efforts around patterns of identified student needs. This may also support efforts to triage primary systems-level needs: rather than having to choose between a host of elevated mean scores, LPA can support distinguishing between students who are mostly well and three or four remaining profiles of socioemotional risk. This aligns with the overall goal of selecting quantitative methodologies in school mental health research that lend themselves to effective intervention decision-making. Previous research has identified that data-based decision making can improve student academic (Carlson, Borman, & Robinson, 2011) and behavioral outcomes (Bradshaw et al., 2010), and models that generate richer detail about

student risk may support more efficient problem analysis and more appropriate intervention selection. Again, this analytic effort would need to take place within a broader public health system, such as the Coalition, with the resources to collect and analyzes screening data, draw meaningful conclusions, conduct further problem analysis, and provide interventions.

Further research is necessary to determine the extent to which applications of advanced methodologies can be translated effectively into applied settings to impact students. Schools participating in extensive universal screening procedures may be reluctant to engage in additional problem analysis before starting intervention. Person-centered modeling of efficiently-collected universal screening data may provide rich details to minimize the amount of follow-up assessment necessary to inform more appropriate intervention choices.

### **Data Use and Applied Interpretation of Person-Centered Research**

For administrators, strategic partnerships with researchers may be necessary to manage and interpret modeling of screening data, and counselors and other school mental health professionals may require additional training in evidence-based interventions appropriate to meet needs identified through screening. Access to screening data does not necessarily impact student outcomes or even inform intervention decisions; thus, data *access* should not be conflated with data-based *decision-making* (Spillane, 2012). Taylor, Kilgus, and Huang (2017) found that providing schools with systematic universal screening data alone not only failed to significantly impact academic and behavioral outcomes, but it also had minimal impact on intervention decision-making compared with traditional identification methods such as ODRs. Students identified, whether through

screening or ODRs, were similarly placed in pre-determined Tier 2 intervention groups regardless of the nature or function of students' difficulties. Teachers and other school professionals report a lack of familiarity with a range of Tier 2 interventions (Stormont & Reinke, 2013), resulting in the placement of identified students into familiar protocols rather than conducting more detailed follow-up assessments to inform more appropriate intervention choices (Stormont & Reinke, 2013). Advanced quantitative methods such as latent profile analysis are not typically a part of the required training for even doctoral-level school administrators, suggesting a gap between current educational research and practice. Collaboration with external researchers may support schools in accessing the applications of these methods efficiently, rather than prescribing additional statistical training to school professionals.

Researchers collaborating with school decision-makers still need to consider strategic ways to communicate results of person-centered analyses. Currently, the Coalition model assigns a regional coordinator to support schools in accessing and interpreting their screening data, in addition to generating reports for schools at the individual, grade, and school levels of analysis regarding descriptive statistics across each EIS-SR domain. Future research may involve piloting new sample screening reports, involving visualizing LPA profiles, to support communicating findings and to determine whether access to information about profiles results in differential decision-making about student interventions compared to business-as-usual efforts. This may identify additional gaps in practice, including knowledge of appropriate interventions to support profiles of comorbid concerns.

### **Integration of Additional Informants**

Finally, future research, both on the EIS-SR and on person-centered analyses in screening, should consider the use of additional informants in screening protocols, including teachers and parents. The Coalition currently administers a teacher-report form of the EIS (EIS-TR) alongside the EIS-SR during its three screening windows. In participating schools, each student grades K-12 has at least one primary-subject classroom teacher complete the EIS-TR regarding their behavior and emotions. Data from the EIS-TR are typically presented to school teams and incorporated in decision-making for interventions; however, the EIS-TR has yet to be evaluated for dimensionality and item performance. Further research is necessary to determine the utility and strength of this measure and its contribution to the understanding of student risk in the population served by the Coalition.

Broadly, integration of multiple informants into screening protocols is not uncommon, and screening efforts are often most successful when examining discrepant informants' reports together as separate yet valuable data points (Achenbach, McConaughy, & Howell, 1987; Humphrey & Wigelsworth, 2016). De Los Reyes and Kazdin (2005) describe informant discrepancies as evidence that informants vary systematically in the contexts in which they observe student behavior, and youth may express symptoms differently across different contexts, and thus discrepancies between ratings may reflect true differences in symptom expression.

Generally, teachers and parents are considered more reliable reporters of student externalizing behavior (Loeber, Green, Lahey, & Stouthamer-Loeber, 1991), whereas students are considered more reliable reporters of their own internalizing symptoms (Smith, 2007; Youngstrom, Loeber, & Stouthamer-Loeber, 2000; Loeber et al., 1991),

especially as they enter adolescence (Dowdy & Kim, 2012). External raters often exhibit difficulty reliably detecting and differentiating between youth internalizing problems (Lonigan, Carey, & Finch, 1994), and thus researchers have found low correspondence between youth and adult ratings of youth internalizing problems ( $r$ s between .20-.40, Achenbach et al., 1987; De Los Reyes et al., 2013). Prior research shows teachers typically report significantly fewer student internalizing problems than their students report for themselves (Herman et al., 2018; Youngstrom et al., 2000), perhaps representing low teacher sensitivity to these symptoms. Informant discrepancies typically increase in adolescence, and by high school, self-report data is considered most valuable in assessing student risk (Dowdy & Kim, 2012). A holistic, broadband screening effort is likely best served by examining data from multiple informants, although further research is necessary to determine the unique contributions of each informant for different domains of socioemotional risk. Additionally, further research is necessary to determine systems for weighing the value of each informant's contribution to support effective use of the data (Humphrey & Wigelsworth, 2016; von der Embse, Kim, Kilgus, Dedrick, & Sanchez, 2019).

In addition to examining the correspondence between scores, future research could examine the use of person-centered analytic strategies in understanding profiles of student risk as rated by both the student and their teachers. For example, there may emerge separate groups of students who rate high levels of internalizing symptoms that either go undetected or noticed by their teachers; examination of demographic correlations may reveal systematic under-detection of certain groups of students, and examination of outcomes may reveal these groups responding differentially to

intervention. Further research into the role of teachers as informants may support administrators in identifying faculty “blind spots” and developing targeted professional development to address these needs.

### **Conclusion**

The current study aimed to connect the power of advanced quantitative methodology with the growing needs of youth for mental health services. Our results represent a proof-of-concept that person-centered analyses can support the efforts to larger public health systems working to efficiently provide evidence-based services to children in schools. Researchers should continue to attend to developments in methodology, search for innovative applications, and pursue collaborative partnerships with school leaders to reduce the population burden of mental health concerns and minimize the research-to-practice gap.

## References

- Abrams, L., Flood, J., & Phelps, L. (2006). Psychopharmacology in the schools. *Psychology in the Schools, 43*, 493–501. doi:10.1002/pits.20163
- Achenbach, T.M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. *Psychological Monographs, 80*, 1–37.  
<http://dx.doi.org/10.1037/h0093906>
- Achenbach, T.M., & Edelbrock, C. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. *Psychological Bulletin, 85*(6), 1275–1301. <http://dx.doi.org/10.1037/0033-2909.85.6.1275>
- Achenbach, T.M., Krukowski, R.A., Dumenci, L., & Ivanova, M.Y. (2005). Assessment of adult psychopathology: Meta-analyses and implications of cross-informant correlations. *Psychological Bulletin, 131*, 361–382.  
<http://dx.doi.org/10.1037/0033-2909.131.3.361>
- Achenbach, T.M., McConaughy, S.H., & Howell, C.T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*(2), 213–232.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory* (B. Petrov and F. Csaki, eds.) 267-281. Akademiai Kiado, Budapest.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

- 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>.
- Amone-P'Olak, K., Burger, H., Ormel, J., Huisman, M., Verhulst, F.C., & Oldehinkel, A.J. (2009). Socioeconomic position and mental health problems in pre-and early-adolescents. *Social Psychiatry and Psychiatric Epidemiology*, *44*(3), 231-238.
- Anderson, A.R., Christenson, S.L., Sinclair, M.F., & Lehr, C.A. (2004). Check & Connect: The importance of relationships for promoting engagement with school. *Journal of School Psychology*, *42*(2), 95-113.
- Angold, A., Costello, E.J., & Erkanli, A. (1999). Comorbidity. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *40*(1), 57-87.
- Angold, A., Erkanli, A., Farmer, E.M., Fairbank, J.A., Burns, B.J., Keeler, G., & Costello, E.J. (2002). Psychiatric disorder, impairment, and service use in rural African American and white youth. *Archives of General Psychiatry*, *59*(10), 893-901.
- Aseltine, R.H., & DeMartino, R. (2004). An outcome evaluation of the SOS suicide prevention program. *American Journal of Public Health*, *94*(3), 446-451.
- Bacci, S., Bartolucci, F., Grilli, L., & Rampichini, C. (2017). Evaluation of student performance through a multidimensional finite mixture IRT model. *Multivariate Behavioral Research*, *52*(6), 732-746.
- Backenstrass, M., Frank, A., Joest, K., Hingman, S., Mundt, C., Kronmuller, K.T. (2006). A comparative study of nonspecific depressive symptoms and minor depression



- regarding functional impairment and associated characteristics in primary care. *Computational Psychiatry*, 47, 35-41.
- Balázs, J., Miklósi, M., Keresztény, Á., Hoven, C.W., Carli, V., Wasserman, C., ... & Cotter, P. (2013). Adolescent subthreshold-depression and anxiety: Psychopathology, functional impairment and increased suicide risk. *Journal of child psychology and psychiatry*, 54(6), 670-677.
- Barkley, R. A. (2011). The importance of emotion in ADHD. *Journal of ADHD and Related Disorders*, 1(2), 5-37.
- Bender, P.K., Reinholdt-Dunne, M.L., Esbjørn, B.H., & Pons, F. (2012). Emotion dysregulation and anxiety in children and adolescents: Gender differences. *Personality and Individual Differences*, 53(3), 284–288. <https://doi.org/10.1016/j.paid.2012.03.027>
- Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107(2), 238.
- Blanco, C., Wall, M.M., He, J.P., Krueger, R.F., Olfson, M., Jin, C.J., . . . Merikangas, K.R. (2015). The space of common psychiatric disorders in adolescents: Comorbidity structure and individual latent liabilities. *Journal of the American Academy of Child and Adolescent Psychiatry*, 54, 45–52. <http://dx.doi.org/10.1016/j.jaac.2014.10.007>
- Boudreaux, E.D., De Beurs, D.P., Nguyen, T.H., Haskins, B.L., Larkin, C., & Barton, B. (2018). Applying computer adaptive testing methods to suicide risk screening in the emergency department. *Suicide and Life-Threatening Behavior*.

- Bradshaw, C.P., Mitchell, M.M., & Leaf, P.J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions*, *12*(3), 133-148.
- Bradshaw, C.P., Reinke, W.M., Brown, L.D., Bevans, K.B., & Leaf, P.J. (2008). Examining the process of Positive Behavioral Interventions and Supports implementation. *Education & Treatment of Children*, *31*, 1–26.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77-101.
- Brown B.B., & Larson, J. (2009). Peer relationships in adolescence. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (3rd ed., pp. 74–103) Hoboken, NJ: Wiley.
- Brown, T.E. (2000). *Attention-deficit disorders and comorbidities in children adolescents, and adults*. Washington, DC: American Psychiatric Press.
- 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*(4), 611-634. <http://dx.doi.org/10.1353/etc.2014.0039>
- Burns, B.J., Costello, E.J., Angold, A., Tweed, D., ... & Erkanli, A. (1995). Children's mental health service use across service sectors. *Health Affairs*, *14*(3), 147-159.
- Calear, A.L., & Christensen, H. (2010). Systematic review of school-based prevention and early intervention programs for depression. *Journal of Adolescence*, *33*(3), 429-438.

- Carlson, D., Borman, G.D., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis*, 33(3), 378-398.
- Carrellas, N.W., Biederman, J., & Uchida, M. (2017). Review article: How prevalent and morbid are subthreshold manifestations of major depression in adolescents? A literature review. *Journal of Affective Disorders*, 210, 166-173.  
doi:10.1016/j.jad.2016.12.037
- Catalano, R.F., Fagan, A.A., Gavin, L.E., Greenberg, M.T., ... & Shek, D.T. (2012). Worldwide application of prevention science in adolescent health. *The Lancet*, 379(9826), 1653-1664.
- Chafouleas, S.M., Kilgus, S.P., & Wallach, N. (2010). Ethical dilemmas in school-based behavioral screening. *Assessment for Effective Intervention*, 35, 245–252.  
<http://dx.doi.org/10.1177/1534508410379002>
- Chalmers, R.P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6).
- Chapman, R.L., Buckley, L., Sheehan, M., & Shochet, I. (2013). School-based programs for increasing connectedness and reducing risk behavior: A systematic review. *Educational Psychology Review*, 25(1), 95-114.
- Choi, S.W, Gibbons, L.E., & Crane, P.K. (2011). lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, 39(8).

- Christensen, H., Pallister, E., Smale, S., Hickie, I.B., & Callear, A.L. (2010). Community-based prevention programs for anxiety and depression in youth: a systematic review. *The Journal of Primary Prevention, 31*(3), 139-170.
- Coie, J.D., Watt, N.F., West, S.G., Hawkins, J.D., ... & Long, B. (1993). The science of prevention: A conceptual framework and some directions for a national research program. *American Psychologist, 48*(10), 1013.
- Cole, D.A., Jacquez, F.M., & Maschman, T.L. (2001). Social origins of depressive cognitions: A longitudinal study of self-perceived competence in children. *Cognitive Therapy and Research, 25*(4), 377-395.
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.
- Crowe, M., Ward, N., Dunnachie, B., & Roberts, M. (2006). Characteristics of adolescent depression. *International Journal of Mental Health Nursing, 15*(1), 10-18.
- Davis, N.J. (2002). The promotion of mental health and the prevention of mental and behavioral disorders: Surely the time is right. *International Journal of Emergency Mental Health, 4*(1), 3-29.
- Daviss, W.B. (2008). A review of co-morbid depression in pediatric ADHD: Etiologies, phenomenology, and treatment. *Journal of Child and Adolescent Psychopharmacology, 18*(6), 565-571. <https://doi-org.proxy.mul.missouri.edu/10.1089/cap.2008.032>
- De Ayala, R., & Santiago, S. (2017). An introduction to mixture item response theory models. *Journal of School Psychology, 60*(2), 25-40.  
doi:10.1016/j.jsp.2016.01.002

- De Ayala, R.J. (2009). *The Theory and Practice of Item Response Theory*. New York: Guilford Press. doi:10.1002/pits.20016
- De Los Reyes, A., Ehrlich, K.B., Swan, A.J., Luo, T.J., Van Wie, M., & Pabón, S.C. (2013). An experimental test of whether informants can report about child and family behavior based on settings of behavioral expression. *Journal of Child and Family Studies*, 22, 177–191. <https://doi.org/10.1007/s10826-012-9567-3>.
- De Los Reyes, A., & Kazdin, A.E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, 131, 483–509. doi:10.1037/0033-2909.131.4.483
- Dowdy, E., & Kim, E. (2012). Choosing informants when conducting a universal screening for behavioral and emotional risk. *School Psychology Forum*, 4(4), 1–10.
- Drummond, T. (1994). *The student risk screening scale (SRSS)*. Grants Pass, OR: Josephine County Mental Health Program.
- Educational Testing Service. (2010). *About the GRE general test*. Retrieved from <http://www.ets.org/gre/general/about>.
- Embretson, S.E., & Reise, S.P. (2000). *Item Response Theory for Psychologists*. Mahwah, NJ: Erlbaum.
- Ennett, S.T., Tobler, N.S., Ringwalt, C.L., & Flewelling, R.L. (1994). How effective is drug abuse resistance education? A meta-analysis of Project DARE outcome evaluations. *American Journal of Public Health*, 84(9), 1394-1401.

- Espelage, D.L., Mebane, S.E., & Swearer, S.M. (2004). Gender differences in bullying: Moving beyond mean level differences. In *Bullying in American Schools* (pp. 37-58). Routledge.
- Felix, E.D., Binmoeller, C., Sharkey, J.D., Dowdy, E., Furlong, M.J., & Latham, N. (2019). The influence of different longitudinal patterns of peer victimization on psychosocial adjustment. *Journal of School Violence, 18*(4), 483-497.
- Finch, W.H., & Hernández Finch, M.E. (2013). Investigation of specific learning disability and testing accommodations based differential item functioning using a multilevel multidimensional mixture item response theory model. *Educational and Psychological Measurement, 73*(6), 973-993.
- Finch, W., & Pierson, E. (2011). A mixture IRT analysis of risky youth behavior. *Frontiers in Psychology, 2*, doi:10.3389/fpsyg.2011.00098
- Flay, B.R., Biglan, A., Boruch, R.F., Castro, F.G., ... & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*(3), 151-175.
- Forbush, K.T., & Watson, D. (2013). The structure of common and uncommon mental disorders. *Psychological Medicine, 43*, 97–108.  
<http://dx.doi.org/10.1017/S0033291712001092>
- Fonseca-Pedrero, E., Sierra-Baigrie, S., Lemos-Giráldez, S., Paino, M., & Muñiz, J. (2012). Dimensional structure and measurement invariance of the Youth Self-Report across gender and age. *Journal of Adolescent Health, 50*(2), 148-153.
- Fox, C.K., Eisenberg, M.E., McMorris, B.J., Pettingell, S.L., & Borowsky, I.W. (2013). Survey of Minnesota parent attitudes regarding school-based depression and

suicide screening and education. *Maternal and child health journal*, 17(3), 456-462.

Fredrick, S.S., Drevon, D.D., & Jervinsky, M. (2019). Measurement invariance of the Student Risk Screening Scale across time and gender. *School Psychology*, 34(2), 159–167. <https://doi-org.proxy.mul.missouri.edu/10.1037/spq0000278>

Fredricks, J.A., Blumenfeld, P.C., & Paris, A.H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74, 59–109.

Gilman, R., Huebner, E.S., Tian, L., Park, N., ... & Langknecht, H. (2008). Cross-national adolescent multidimensional life satisfaction reports: Analyses of mean scores and response style differences. *Journal of Youth and Adolescence*, 37(2), 142-154.

Glover, T.A., & Albers, C.A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, 45(Universal Screening for Enhanced Educational and Mental Health Outcomes), 117-135.

doi:10.1016/j.jsp.2006.05.005

Goldman, S. (2012). Developmental epidemiology of depressive disorders. *Child and Adolescent Psychiatry Clinics of North America*, 21, 217–235.

doi:10.1016/j.chc.2011.12.002.

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

Greenwald, P., & Cullen, J.W. (1984). The scientific approach to cancer control.

*Canadian-American Journal for Clinicians*, 34, 328-332.

- Guo, J., Wall, M., & Amemiya, Y. (2006). Latent class regression on latent factors. *Biostatistics*, 7(1), 145-163.
- Hallfors, D., Brodish, P.H., Khatapoush, S., Sanchez, V., Cho, H., & Steckler, A. (2006). Feasibility of screening adolescents for suicide risk in “real-world” high school settings. *American Journal of Public Health*, 96(2), 282-287.
- Harrell-Williams, L.M., Raines, T.C., Kamphaus, R.W., & Dever, B.V. (2015). Psychometric analysis of the BASC–2 Behavioral and Emotional Screening System (BESS) student form: Results from high school student samples. *Psychological Assessment*, 27(2), 738–743. <https://doi.org/10.1037/pas0000079>
- Haslam, N., Holland, E., & Kuppens, P. (2012). Categories versus dimensions in personality and psychopathology: A quantitative review of taxometric research. *Psychological Medicine*, 42, 903–920. <http://dx.doi.org/10.1017/S0033291711001966>
- Hawkins, J.D., Catalano, R.F., & Arthur, M.W. (2002). Promoting science-based prevention in communities. *Addictive Behaviors*, 27(6), 951-976.
- Herman, K.C., Cohen, D., Reinke, W.M., Ostrander, R., Burrell, L., McFarlane, E., & Duggan, A.K. (2018). Using latent profile and transition analyses to understand patterns of informant ratings of child depressive symptoms. *Journal of School Psychology*, 69, 84-99.
- Herman, K.C., Hodgson, C.G., Eddy, C.L., Cohen, D.R., Reinke, W.M., Burrell, L., ... & 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*, 265–274.
- Herman, K.C., Merrell, K.W., Reinke, W.M., & Tucker, C.M. (2004). The role of school psychology in preventing depression. *Psychology in the Schools, 41*(7), 763–775.  
<https://doi.org/10.1002/pits.20016>
- Hill, C.E., Thompson, B.J., & Williams, E.N. (1997). A guide to conducting consensual qualitative research. *The Counseling Psychologist, 25*(4), 517-572.
- Hoogland, I., Schildkamp, K., Van der Kleij, F., Heitink, M., ... & Dijkstra, A.M. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. *Teaching and Teacher Education, 60*, 377-386.
- Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179-185.
- Horner, R.H., Sugai, G., & Anderson, C.M. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptional Children, 42*(8).
- Horner, R.H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A.W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions, 11*(3), 133-144.

- Hu, T., & Bentler, P.M. (1995). Evaluating model fit. In *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage Publications, Inc.
- Huang, F., Reinke, W., Thompson, A., Herman, K. & County School Mental Health Coalition (2019). An investigation of the psychometric properties of the Early Identification System–Student Report. *Journal of Psychoeducational Assessment*, 37, 473-485. doi:10.1177/0734282918758791
- Humphrey, N., & Wigelsworth, M. (2016). Making the case for universal school-based mental health screening. *Emotional & Behavioural Difficulties*, 21(1), 22–42. <https://doi.org/10.1080/13632752.2015.1120051>
- Jimerson, S.R., Burns, M.K., & VanDerHeyden, A.M. (Eds.). (2007). *The handbook of response to intervention: Science and practice of assessment and intervention* (1st ed.). New York, NY: Springer Science. <http://dx.doi.org/10.1007/978-0-387-49053-3>
- Kadengye, D.T., Ceulemans, E., & Van den Noortgate, W. (2014). A generalized longitudinal mixture IRT model for measuring differential growth in learning environments. *Behavior Research Methods*, 46(3), 823-840.
- Kaiser, M.O. (1974). Kaiser-Meyer-Olkin measure for identity correlation matrix. *Journal of the Royal Statistical Society*, 52, 296-298.
- Kamphaus, R.W. (2012). Screening for behavioral and emotional risk: Constructs and practicalities. *School Psychology Forum*, 6, 89–97.

- Kamphaus, R.W., & Reynolds, C.R. (2007). *Behavior Assessment System for Children—Second Edition (BASC–2): Behavioral and Emotional Screening System (BESS)*.  
Bloomington, MN: Pearson.
- Kataoka, S.H., Zhang, L., & Wells, K.B. (2002). Unmet need for mental health care among US children: Variation by ethnicity and insurance status. *American Journal of Psychiatry*, *159*(9), 1548-1555.
- Keenan, K. (2000). Emotion dysregulation as a risk factor for child psychopathology. *Clinical Psychology: Science and Practice*, *7*(4), 418-434.
- Kellam, S.G., & Langevin, D.J. (2003). A framework for understanding “evidence” in prevention research and programs. *Prevention Science*, *4*(3), 137-153.
- Kellam, S.G., Koretz, D., & Mościcki, E.K. (1999). Core elements of developmental epidemiologically based prevention research. *American Journal of Community Psychology*, *27*(4), 463-482.
- Kemper, A.R., Fant, K.E., Bruckman, D., & Clark, S.J. (2004). Hearing and vision screening program for school-aged children. *American Journal of Preventive Medicine*, *26*(2), 141-146.
- Keough, M.E., Riccardi, C.J., Timpano, K.R., Mitchell, M.A., & Schmidt, N.B. (2010). Anxiety symptomatology: The association with distress tolerance and anxiety sensitivity. *Behavior Therapy*, *41*(4), 567-574.
- Kessler, R.C., Andrews, G., Colpe, L.J., Hiripi, E., Mroczek, D.K., ... & Zaslavsky, A.M. (2002). Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychological Medicine*, *32*(6), 959-976.

- Kilgus, S.P., Bonifay, W.E., von der Embse, N.P., Allen, A.N., & Eklund, K. (2018). Evidence for the interpretation of Social, Academic, and Emotional Behavior Risk Screener (SAEBRS) scores: An argument-based approach to screener validation. *Journal of School Psychology, 68*(129-141). doi:10.1016/j.jsp.2018.03.002
- 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*(3), 210.
- Kilgus, S.P., Reinke, W.M., & Jimerson, S.R. (2015). Understanding mental health intervention and assessment within a multi-tiered framework: Contemporary science, practice, and policy. *School Psychology Quarterly, 30*(2), 159–65. <https://doi.org/10.1037/spq0000118>
- Kilgus, S.P., von der Embse, N.P., Chafouleas, S.M., & Riley-Tillman, T.C. (2014). Social, academic, and emotional behavior risk screener—teacher rating scale (SAEBRS-TRS)(Unpublished document).
- Kim, J., Kamphaus, R.W., & Society for Research on Educational Effectiveness (SREE). (2013). Latent Class Analysis of Youth Behavioral and Emotional Risk: Associations with Demographic Characteristics. In *Society for Research on Educational Effectiveness*.
- Kindermann, T.A. (2007). Effects of naturally existing peer groups on changes in academic engagement in a cohort of sixth graders. *Child Development, 78*, 1186–1203.
- Kotov, R. (2016). The quantitative classification of mental illness: Emerging solution to boundary problems. In E. Bromet (Ed.), *Long-term outcomes in psychopathology*

*research: Rethinking the scientific agenda* (pp. 140 –157). New York, NY:  
Oxford University Press.

Kotov, R., Krueger, R.F., Watson, D., Achenbach, T.M., Althoff, R.R., Bagby, R.M., . . .

Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology, 126*(4), 454-477. <http://dx.doi.org/10.1037/abn0000258>

Krueger, R.F., & Markon, K.E. (2011). A dimensional-spectrum model of psychopathology: Progress and opportunities. *Archives of General Psychiatry, 68*, 10 –11. <http://dx.doi.org/10.1001/archgenpsychiatry.2010.188>

Kuppens, S., Grietens, H., Onghena, P., Michiels, D., & Subramanian, S. V. (2008).

Individual and classroom variables associated with relational aggression in elementary-school aged children: A multilevel analysis. *Journal of School Psychology, 46*(6), 639-660.

Lane, K.L., Wehby, J.H., Robertson, E.J., & Rogers, L.A. (2007). How do different types

of high school students respond to schoolwide positive behavior support programs? Characteristics and responsiveness of teacher-identified students. *Journal of Emotional and Behavioral Disorders, 15*(1), 3-20.

Lassen, S.R., Steele, M.M., & Sailor, W. (2006). The relationship of school-wide positive

behavior support to academic achievement in an urban middle school. *Psychology in the Schools, 43*(6), 701-712.

- Lee, S., Aos, S., Drake, E., Pennucci, A., Miller, M., & Anderson, L. (2012). Return on investment: Evidence-based options to improve statewide outcomes. *Olympia: Washington State Institute for Public Policy*.
- Lee, A., Hankin, B.L., & Mermelstein, R.J. (2010). Perceived social competence, negative social interactions, and negative cognitive style predict depressive symptoms during adolescence. *Journal of Clinical Child & Adolescent Psychology, 39*(5), 603-615.
- Lembke, E.S., McMaster, K.L., & Stecker, P.M. (2010). The prevention science of reading research within a response-to-intervention model. *Psychology in the Schools, 47*(1), 22– 35.
- Levitt, V.H., & Merrell, K.W. (2009). Linking assessment to intervention for internalizing problems of children and adolescents. *School Psychology Forum: Research to Practice, 3*, 13-26.
- Levitt, J., Saka, N., Romanelli, L.H., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. *Journal of School Psychology, 45*, 163–191. doi:10.1016/j.jsp.2006.11.005
- Li, F., Cohen, A.S., Kim, S.H., & Cho, S.J. (2009). Model selection methods for mixture dichotomous IRT models. *Applied Psychological Measurement, 33*(5), 353-373.
- Linehan, M.M. (1993). *Skills training manual for treating borderline personality disorder*. Guilford Press.

- Loeber, R., Green, S.M., Lahey, B.B., & Stouthamer-Loeber, M. (1991). Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. *Journal of Abnormal Child Psychology*, *19*, 75-95.
- Lonigan, C.J., Carey, M.P., & Finch, A.J. (1994). Anxiety and depression in children and adolescents: negative affectivity and the utility of self-reports. *Journal of Consulting and Clinical Psychology*, *62*(5), 1000.
- Markon, K.E., Chmielewski, M., & Miller, C.J. (2011). The reliability and validity of discrete and continuous measures of psychopathology: A quantitative review. *Psychological Bulletin*, *137*, 856 – 879. <http://dx.doi.org/10.1037/a0023678>
- Masia-Warner, C., Nangle, D.W., & Hansen, D.J. (2006). Bringing evidence-based child mental health services to the schools: General issues and specific populations. *Education and Treatment of Children*, *29*(2), 165-172.
- McIntosh, K., Flannery, K.B., Sugai, G., Braun, D.H., & Cochrane, K.L. (2008). Relationships between academics and problem behavior in the transition from middle school to high school. *Journal of Positive Behavior Interventions*, *10*(4), 243-255.
- McIntosh, K., Ty, S.V., & Miller, L.D. (2014). Effects of school-wide positive behavioral interventions and supports on internalizing problems: Current evidence and future directions. *Journal of Positive Behavior Interventions*, *16*, 209 –218.  
<http://dx.doi.org/10.1177/1098300713491980>
- McLachlan, G.J. (1987). On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture. *Annals of Statistics*, *36*, 318–324.

- 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.
- McLaughlin, K.A., Hilt, L.M., & Nolen-Hoeksema, S. (2007). Racial/ethnic differences in internalizing and externalizing symptoms in adolescents. *Journal of Abnormal Child Psychology*, *35*(5), 801-816.
- Merikangas, K.R., He, J., Burstein, M., Swanson, S.A., Avenevoli, S., ... Swendsen, J. (2010). Lifetime prevalence of mental disorders in US adolescents: Results from the National Comorbidity Study-Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*, 980-989. doi: 10.1016/j.jaac.2010.05.017
- Merrell, K.W., & Buchanan, R. (2006). Intervention selection in school-based practice: Using public health models to enhance systems capacity of schools. *School Psychology Review*, *35*, 167-180.
- Monroe, S., & Cai, L. (2015). Examining the reliability of student growth percentiles using multidimensional IRT. *Educational Measurement: Issues and Practice*, *34*(4), 21-30.
- Mrazek, P.J., & Haggerty, R.J. (Eds.). (1994). *Reducing risks for mental disorders: Frontiers for preventive intervention research*. Washington, DC: National Academy Press.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *ETS Research Report Series*, *1992*(1), i-30.



- Murthy, V.H. (2015). Surgeon General's perspectives. *Public Health Reports* (Washington, D.C. : 1974), 130(3), 193–195.
- Muthén, L.K. and Muthén, B.O. (2013). Mplus User's Guide. Seventh Edition. Los Angeles, CA: Muthén & Muthén.
- Nail, J.E., Christofferson, J., Ginsburg, G.S., Drake, K., Kendall, P.C., McCracken, J.T., & Sakolsky, D. (2015). Academic impairment and impact of treatments among youth with anxiety disorders. *Child & Youth Care Forum*, 44(3), 327-342.  
doi:10.1007/s10566-014-9290-x
- Neil, A.L., & Christensen, H. (2009). Efficacy and effectiveness of school-based prevention and early intervention programs for anxiety. *Clinical Psychology Review*, 29, 208–215.
- Nelson-Gray, R.O., Keane, S.P., Hurst, R.M., Mitchell, J.T., Warburton, J.B., Chok, J.T., & Cobb, A.R. (2006). A modified DBT skills training program for oppositional defiant adolescents: promising preliminary findings. *Behaviour Research and Therapy*, 44(12), 1811–1820.
- Neumann, A., van Lier, P.A.C., Gratz, K.L., & Koot, H.M. (2010). Multidimensional assessment of emotion regulation difficulties in adolescents using the difficulties in emotion regulation scale. *Assessment*, 17(1), 138–149.
- Nylund, K.L., Asparouhov, T., & Muthén, B.O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural equation modeling: A multidisciplinary Journal*, 14(4), 535-569.

- Osteen, P. (2010). An introduction to using multidimensional item response theory to assess latent factor structures. *Journal of the Society for Social Work and Research, 1*(2), 66-82.
- Owens, P.L., Hoagwood, K., Horwitz, S.M., Leaf, P.J., ... & Ialongo, N.S. (2002). Barriers to children's mental health services. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*(6), 731-738.
- Paschall, M.J., & Bersamin, M. (2018). School-based mental health services, suicide risk and substance use among at-risk adolescents in Oregon. *Preventive Medicine, 106*209-215. doi:10.1016/j.ypmed.2017.11.004
- Perou, R., Bitsko, R.H., Blumberg, S.J., Pastor, P., ... & Parks, S.E. (2013). Mental health surveillance among children—United States, 2005–2011. *MMWR Surveill Summ, 62*(Suppl 2), 1-35.
- Pilkonis, P.A., Choi, S.W., Reise, S.P., Stover, A.M., ... & PROMIS Cooperative Group. (2011). Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS®): depression, anxiety, and anger. *Assessment, 18*(3), 263-283.
- Pliszka, S.R., Carlson, C., & Swanson, J.M. (1999). *ADHD with comorbid disorders: Clinical assessment and management*. New York: Guilford.
- Poulin, F., & Chan, A. (2010). Friendship stability and change in childhood and adolescence. *Developmental Review, 30*, 257–272.

- R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ramaswamy, V., DeSarbo, W.S., Reibstein, D.J., & Robinson, W.T. (1993). An empirical pooling approach for estimating marketing mix elasticities with PIMS data. *Marketing Science*, 12(1), 103-124.
- Reckase, M.D. (2007). *Multidimensional item response theory*. New York, NY: Springer.
- Reinke, W.M., Herman, K.C., Huang, F., McCall, C., Holmes, S., & Thompson, A. (2020) Examining the Factor Structure and Predictive Validity of the Early Identification System – Student Report in an Elementary School Sample. Manuscript in preparation.
- 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(5), 759-770.
- Reinke, W.M., Thompson, A., Herman, K.C., Holmes, S., ... & 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., & Waller, N.G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, 5, 27-48.
- Reiss, F. (2013). Socioeconomic inequalities and mental health problems in children and adolescents: a systematic review. *Social Science & Medicine*, 90, 24-31.

- Rheingold, A.A., Herbert, J.D., & Franklin, M.E. (2003). Cognitive bias in adolescents with social anxiety disorder. *Cognitive Therapy and Research*, 27(6), 639-655.
- Ridgeway, D., Waters, E., & Kuczaj, S.A. (1985). Acquisition of emotion-descriptive language: Receptive and productive vocabulary norms for ages 18 months to 6 years. *Developmental Psychology*, 21(5), 901.
- Robison, L.M., Skaer, T.L., Sclar, D.A., & Galin, R.S. (2002). Is attention deficit hyperactivity disorder increasing among girls in the US? *CNS Drugs*, 16(2), 129-137.
- Rocque, M. (2010). Office discipline and student behavior: Does race matter?. *American Journal of Education*, 116(4), 557-581.
- Rones, M. & Hoagwood, K. (2000). School-based mental health services: A research review. *Clinical Child and Family Psychological Review*, 3(4), 223-241.
- Rose, E.J., & Ebmeier, K.P. (2006). Pattern of impaired working memory during major depression. *Journal of Affective Disorders*, 90(2), 149–161. <https://doi-org.proxy.mul.missouri.edu/10.1016/j.jad.2005.11.003>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. <http://www.jstatsoft.org/v48/i02/>.
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14(3), 271-282.

- Rucklidge, J.J., & Tannock, R. (2001). Psychiatric, psychosocial, and cognitive functioning of female adolescents with ADHD. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(5), 530-540.
- Rudner, L.M. (1998). *An on-line, interactive, computer adaptive testing tutorial*. Retrieved February 6, 2018 from <http://echo.edres.org:8080/scripts/cat/catdemo.htm>
- Samejima, F. (1970). Estimation of latent ability using a response pattern of graded scores. *Psychometrika, 35*(1), 139.
- Schanding, G.T., & Nowell, K.J. (2013). Universal screening for emotional and behavioral problems: Fitting a population-based model. *Journal of Applied School Psychology, 29*, 104 –119. doi:10.1080/15377903.2013.751479
- Schatschneider, C., Lane, K.L., Oakes, W.P., & Kalberg, J.R. (2014). The student risk screening scale: Exploring dimensionality and differential item functioning. *Educational Assessment, 19*(3), 185-203.
- Schatz, D.B., & Rostain, A.L. (2006). ADHD with Comorbid Anxiety: A Review of the Current Literature. *Journal of Attention Disorders, 10*(2), 141–149.
- Scherff, A.R., Eckert, T.L., & Miller, D.N. (2005). Youth suicide prevention: a survey of public school superintendents' acceptability of school-based programs. *Suicide and Life-Threatening Behavior, 35*(2), 154-169.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461-464.

- Scott, M.A., Wilcox, H.C., Schonfeld, I.S., Davies, M., ... & Shaffer, D. (2009). School-based screening to identify at-risk students not already known to school professionals: the Columbia suicide screen. *American Journal of Public Health, 99*(2), 334-339.
- Severson, H.H., Walker, H.M., Hope-Doolittle, J., Kratochwill, T.R., & Gresham, F.M. (2007). Proactive, early screening to detect behaviorally at-risk students: Issues, approaches, emerging innovations, and professional practices. *Journal of School Psychology, 45*, 193–223. doi:10.1016/j.jsp.2006.11.003
- Shankman, S.A., Lewinsohn, P.M., Klein, D.N., Small, J.W., Seeley, J.R., & Altman, S.E. (2009). Subthreshold conditions as precursors for full syndrome disorders: A 15-year longitudinal study of multiple diagnostic classes. *Journal of Child Psychology and Psychiatry, 50*, 1485–1494.
- Skiba, R.J., Michael, R.S., Nardo, A.C., & Peterson, R.L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review, 34*(4), 317-342.
- Smith, S.R. (2007). Making sense of multiple informants in child and adolescent psychopathology: A guide for clinicians. *Journal of Psychoeducational Assessment, 25*(2), 139-149.
- Spillane, J.P. (2012). Data in practice: Conceptualizing the data-based decision-making phenomena. *American Journal of Education, 118*(2), 113-141.
- Spoth, R., Trudeau, L., Goyll, M., Shin, C., & Redmond, C. (2009). Universal intervention effects on substance use among young adults mediated by delayed

- adolescent substance initiation. *Journal of Consulting and Clinical Psychology*, 77(4), 620.
- Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences*, 9(2), 69-74.
- Stormont, M., & Reinke, W. M. (2013). Implementing Tier 2 social behavioral interventions: Current issues, challenges, and promising approaches. *Journal of Applied School Psychology*, 29, 121–125.  
<http://dx.doi.org/10.1080/15377903.2013.778769>.
- Stormont, M., & Reinke, W. M. (2014). Providing performance feedback for teachers to increased treatment fidelity. *Intervention in School and Clinic*, 49, 219–224
- Stormont, M., Reinke, W.M., & Herman, K.C. (2010). Introduction to the special issue: Using prevention science to address mental health issues in schools. *Psychology in the Schools*, 47(1), 1-4.
- Sugai, G. Sprague, J.R., Horner, R.H., & Walker, H.M. (2000). Preventing school violence: The use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of Emotional and Behavioral Disorders*, 8, 94-101. doi: 10.1177/1063426600000800205
- Sugai, G., & Horner, R. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child & Family Behavior Therapy*, 24(1-2), 23-50.
- Sunderland, M., Batterham, P., Carragher, N., Calear, A., & Slade, T. (2017). Developing and validating a computerized adaptive test to measure broad and specific factors of internalizing in a community sample. *Assessment*, 1073191117707817.

- Tadić, A., Helmreich, I., Mergl, R., Hautzinger, ... & Hegerl, U. (2010). Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. *Journal of Affective Disorders, 120*(1-3), 86-93.
- Taylor, C.N., Kilgus, S.P., & Huang, F. (2017). Treatment Utility of Universal Screening for Behavioral Risk: A Manipulated Assessment Study. *Journal of Applied School Psychology, 0*(0), 1–17. <https://doi.org/10.1080/15377903.2017.1394949>
- Tegethoff, M., Stalujanis, E., Belardi, A., & Meinschmidt, G. (2014). School mental health services: Signpost for out-of-school service utilization in adolescents with mental disorders? A nationally representative United States Cohort. *PLoS One, 9*, e99675.
- Todd, A.W., Campbell, A.L., Meyer, G.G., & Horner, R.H. (2008). The effects of a targeted intervention to reduce problem behaviors: Elementary school implementation of check in—check out. *Journal of Positive Behavior Interventions, 10*(1), 46-55.
- Thomas, M.L. (2010). The value of item response theory in clinical assessment: a review. *Assessment, 18*(3), 291-307.
- Thompson, A.M., Reinke, W., Holmes, S., Danforth, L., & Herman, K. (2017). County Schools Mental Health Coalition: A Model for a Systematic Approach to Supporting Youths. *Children & Schools, 39*(4), 209-218.
- Tilly, W.D. (2008). The evolution of school psychology to a science-based practice: Problem solving and the three-tiered model. In J. Grimes & A. Thomas (Eds.),



*Best practices in school psychology V* (pp. 17–36). Bethesda, MD: National Association of School Psychologists.

- Tram, J.M., & Cole, D.A. (2006). A multimethod examination of the stability of depressive symptoms in childhood and adolescence. *Journal of Abnormal Psychology, 115*, 674–686.
- von der Embse, N., Kim, E.S., Kilgus, S., Dedrick, R., & Sanchez, A. (2019). Multi-informant universal screening: Evaluation of rater, item, and construct variance using a trifactor model. *Journal of School Psychology, 77*, 52-66.
- von der Embse, N.P., Pendergast, L.L., Kilgus, S.P., & Eklund, K.R. (2016). Evaluating the applied use of a mental health screener: Structural validity of the Social, Academic, and Emotional Behavior Risk Screener. *Psychological Assessment, 28*(10), 1265.
- Walker, B., Cheney, D., Stage, S., Blum, C., & Horner, R.H. (2005). Schoolwide screening and positive behavior supports identifying and supporting students at risk for school failure. *Journal of Positive Behavior Interventions, 7*, 194–204. <http://dx.doi.org/10.1177/10983007050070040101>
- Walker, H.M., & Shinn, M.R. (2002). Structuring school-based interventions to achieve integrated primary, secondary, and tertiary prevention goals for safe and effective schools. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventative and remedial approaches* (pp. 1-25). Bethesda, MD: NASP.
- Walker, H.M., Horner, R.H., Sugai, G., Bullis, M., ... & Kaufman, M.J. (1996). Integrated approaches to preventing antisocial behavior patterns among school-

- age children and youth. *Journal of Emotional and Behavioral Disorders*, 4(4), 194-209
- Walker, H.M., Nishioka, V.M., Zeller, R., Severson, H.H., & Feil, E.G. (2000). Causal factors and potential solutions for the persistent under-identification of students having emotional or behavioral disorders in the context of schooling. *Assessment for Effective Intervention*, 26, 29-40.
- Walrath, C., Petras, H., Mandell, D.S., Stephens, R.L., Holden, E.W., & Leaf, P.J. (2004). Gender differences in patterns of risk factors among children receiving mental health services: Latent class analyses. *The Journal of Behavioral Health Services & Research*, 31, 297–311. doi:10.1097/00075484-200407000-00006
- Walton, K.E., Ormel, J., & Krueger, R.F. (2011). The dimensional nature of externalizing behaviors in adolescence: Evidence from a direct comparison of categorical, dimensional, and hybrid models. *Journal of Abnormal Child Psychology*, 39, 553–561. <http://dx.doi.org/10.1007/s10802-010-9478-y>
- Wight, R.G., Botticello, A.L., & Aneshensel, C.S. (2006). Socioeconomic context, social support, and adolescent mental health: A multilevel investigation. *Journal of Youth and Adolescence*, 35(1), 109.
- Williams, G.A., & Kibowski, F. (2016). Latent class analysis and latent profile analysis. *Handbook of methodological approaches to community-based research: Qualitative, quantitative, and mixed methods*, 143-151.
- Wilens, T. E., Biederman, J., & Spencer, T. J. (2002). Attention Deficit/Hyperactivity Disorder across the lifespan. *Annual Review of Medicine*, 53, 113–131.

- Wilson, S.J., & Lipsey, M.W. (2007). School-based interventions for aggressive and disruptive behavior: Update of a meta-analysis. *American Journal of Preventive Medicine*, 33(2), S130-S143.
- Wolchik, S.A., Sandler, I.N., Millsap, R.E., Plummer, B.A., ... & Haine, R.A. (2002). Six-year follow-up of preventive interventions for children of divorce: A randomized controlled trial. *The Journal of the American Medicine Association*, 288(15), 1874-1881.
- Wu, P., Hoven, C.W., Cohen, P., Liu, X., Moore, R.E., Tiet, Q., ... & Bird, H.R. (2001). Factors associated with use of mental health services for depression by children and adolescents. *Psychiatric Services*, 52(2), 189-195.
- Yao, L., & Boughton, K.A. (2007). A Multidimensional Item Response Modeling Approach for Improving Subscale Proficiency Estimation and Classification. *Applied Psychological Measurement*, 31(2), 83–105.
- Zuckerbrot, R.A., Cheung, A., Jensen, P.S., Stein, R.E., Laraque, D., & GLAD-PC STEERING GROUP. (2018). Guidelines for adolescent depression in primary care (GLAD-PC): Part I. Practice preparation, identification, assessment, and initial management. *Pediatrics*, 120, e20174081.
- Zuckerbrot, R.A., Maxon, L., Pagar, D., Davies, M., Fisher, P.W., & Shaffer, D. (2007). Adolescent depression screening in primary care: feasibility and acceptability. *Pediatrics*, 119(1), 101-8.

Table 1  
*Comparison of candidate mixture models for estimating EIS item properties and latent profiles of risk concurrently.*

	Estimates item location (difficulty) along latent risk trait	Estimates item discrimination ( <i>a</i> )	Accounts for dimensionality of EIS	Estimation difficulty
Mixture Rasch	Yes	No	No	Intermediate
Multidimensional Mixture Rasch	Yes	No	Yes	Difficult
Multidimensional Mixture 2PL	Yes	Yes	Yes	Difficult
Traditional LPA and CFA	No	No	Yes	Easy

Table 2

*Fit statistics for exploratory and confirmatory factor analyses of the EIS-SR.*

	$\chi^2$	df	RMSEA	TLI	BIC
Exploratory sample ( $n = 2,930$ in 9 schools)					
3 factors	13294	700	.079	.727	7706.03
4 factors	10936	662	.073	.765	5651.38
7 factors (initial)	54962	554	.055	.867	989.21
7 factors (modified)	1341	224	.041	.949	-447.55
Confirmatory sample ( $n = 2,930$ in 9 schools)					
7 correlated factor (modified)	3223	406	.044	.956	CFI .961

Table 3

*Standardized factor loadings for the exploratory (confirmatory) factor analysis of the EIS-SR.*

Item	Question	PR	EX	IN	ATT	ED	RA	SD
4	I am a good friend*	.54 (.76)						
5	I cooperate with others*	.85 (.89)						
6	I work well with my classmates*	.80 (.90)						
18	I get in trouble at school		.74 (.82)					
19	I am sent out of class for bad behavior		.68 (.88)					
20	I disrupt class		.56 (.84)					
21	I get into fights with others		.49 (.89)					
22	My friends get in trouble at school		.39 (.70)					
9	I like myself*			.57 (.75)				
12	In the past month, I felt sad			.72 (.87)				
13	In the past month, I felt fearful			.62 (.74)				
14	In the past month, I felt lonely			.80 (.89)				
15	In the past month, I felt worried			.62 (.77)				
16	In the past month, I felt like I did not matter			.84 (.93)				
17	In the past month, I felt hopeless			.82 (.92)				
34	I feel left out by others			.59 (.80)				
39	I need help with my emotions			.51 (.78)				
23	I have trouble sitting still at school				.66 (.79)			
24	I have trouble finishing my work				.54 (.80)			
25	I have trouble paying attention				.88 (.91)			
26	I get mad easily					.88 (.90)		
27	I have a hard time controlling my temper					.77 (.90)		
33	I get crabby and irritated easily					.65 (.87)		
10	I am mean to others						.47 (.84)	
32	I blame others for my mistakes						.47 (.73)	
35	I talk about people behind their back						.65 (.75)	
36	I make fun of others						.72 (.83)	
28	I look forward to learning new things at school*							.78 (.75)
37	I enjoy coming to school*							.80 (.93)

*Notes.* \* = reverse-coded. PR = peer relations, EX = externalizing problems, IN = internalizing problems, ATT = attention problems, ED = emotion dysregulation, RA = relational aggression, SD = school disengagement.

Table 4

*Subscale correlations, reliabilities, and descriptive statistics based on factors of the EIS-SR.*

	PR	IN	EX	ATT	ED	RA	SD
PR	1.00						
IN	.31	1.00					
EX	.20	.12	1.00				
ATT	.29	.46	.28	1.00			
ED	.31	.47	.33	.47	1.00		
RA	.26	.32	.36	.33	.40	1.00	
SD	.47	.24	.13	.33	.21	.24	1.00
Reliabilities (Omega)	.828	.911	.781	.809	.853	.759	.771
<i>M</i>	1.92	4.95	1.13	2.48	1.86	1.20	2.79
<i>SD</i>	1.76	5.14	1.69	2.08	2.00	1.60	1.69
Range	0-9	0-27	0-15	0-9	0-9	0-12	0-6

*Note.* PR = peer relations, EX = externalizing problems, IN = internalizing problems, ATT = attention problems, ED = emotion dysregulation, RA = relational aggression, SD = school disengagement.

Table 5.  
*Fit indices for each latent profile solution evaluated.*

Solution	AIC	BIC (Sample size-adjusted)	Entropy	BLRT
Two-profile	160598.33	160674.11	.826	.00
Three-profile	158337.57	158440.91	.873	.00
Four-profile	156653.59	156784.49	.864	.00
Five-profile	155295.46	155453.92	.819	.00
Six-profile	154401.21	154587.23	.822	.00
Seven-profile	153583.58	153797.15	.844	.00

*Note.* BLRT =  $p$  values associated with the bootstrapped likelihood ratio test comparing fit between nested models.



Table 6.

*Probability of most likely class membership (row) by latent class (column) for the five-profile LPA solution.*

	1	2	3	4	5
1	0.894	0.000	0.080	0.026	0.000
2	0.000	0.903	0.096	0.000	0.000
3	0.024	0.119	0.832	0.026	0.000
4	0.024	0.003	0.070	0.902	0.001
5	0.001	0.000	0.000	0.019	0.980

Table 7.

*Proportion of likely class membership in the five-profile solution.*

	Estimated $n$ (total = 5,555)	Estimated proportion
Profile 1 (High Wellness)	2500.93	.45
Profile 2 (Average Wellness)	1962.48	.35
Profile 3 (Distress/Dysregulation)	441.75	.08
Profile 4 (Externalizing Risk)	612.97	.11
Profile 5 (Very High)	36.86	.01

Table 8.  
*Outcome means, SE, and equality tests across profiles of student risk (n = 5553).*

	Profile 1: High Wellness (n = 2518)	Profile 2: Average Wellness (n = 1921)	Profile 3: Distress/ Dysregulation (n = 439)	Profile 4: Externalizing Risk (n = 637)	Profile 5: Very High (n = 38)	Overall test of significance	Significant class comparisons
ODR	0.83 (0.08)	1.29 (0.11)	1.68 (0.23)	5.51 (0.42)	8.31 (1.61)	146.14***	1 vs. 2** 1 vs. 3*** 1 vs. 4*** 1 vs. 5*** 2 vs. 4*** 2 vs. 5*** 3 vs. 4*** 3 vs. 5***
ISS	0.15 (0.02)	0.24 (0.03)	0.36 (0.07)	1.45 (0.15)	2.61 (0.79)	91.50***	1 vs. 3** 1 vs. 4*** 1 vs. 5** 2 vs. 4*** 2 vs. 5** 3 vs. 4*** 3 vs. 5**
OSS	0.04 (0.01)	0.07 (0.01)	0.08 (0.02)	0.31 (0.04)	0.51 (0.16)	55.12***	1 vs. 4*** 1 vs. 5** 2 vs. 4*** 2 vs. 5** 3 vs. 4***
Math	63.38 (1.01)	57.66 (1.20)	54.35 (2.54)	42.43 (1.84)	37.53 (5.75)	116.76***	1 vs. 2** 1 vs. 3** 1 vs. 4*** 1 vs. 5*** 2 vs. 4*** 2 vs. 5** 3 vs. 4*** 3 vs. 5**
Read -ing	48.40 (0.85)	47.17 (1.02)	48.23 (2.01)	33.18 (1.63)	34.75 (5.46)	89.25***	1 vs. 4*** 2 vs. 4*** 3 vs. 4***

*Note.* ODR, ISS, and OSS = total number of office discipline referrals, in-school suspensions, and out-of-school suspensions for the school year, respectively. Math and Reading = scores on statewide end-of-year math and reading exams.  $\chi^2$  *p* values: \* *p* < .05, \*\* *p* < .01, \*\*\* *p* < .001. Class comparisons were only examined if *p* < .01 to protect against possible Type 1 error.

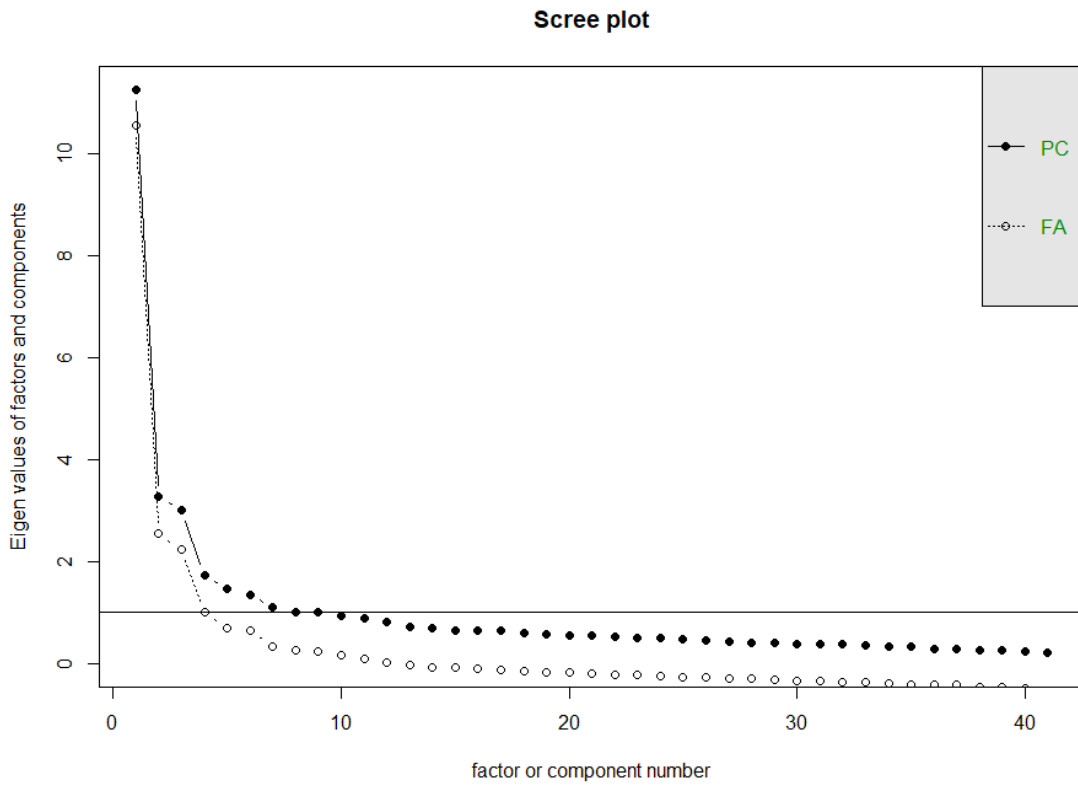


Figure 1. Scree plot of the EIS-SR.

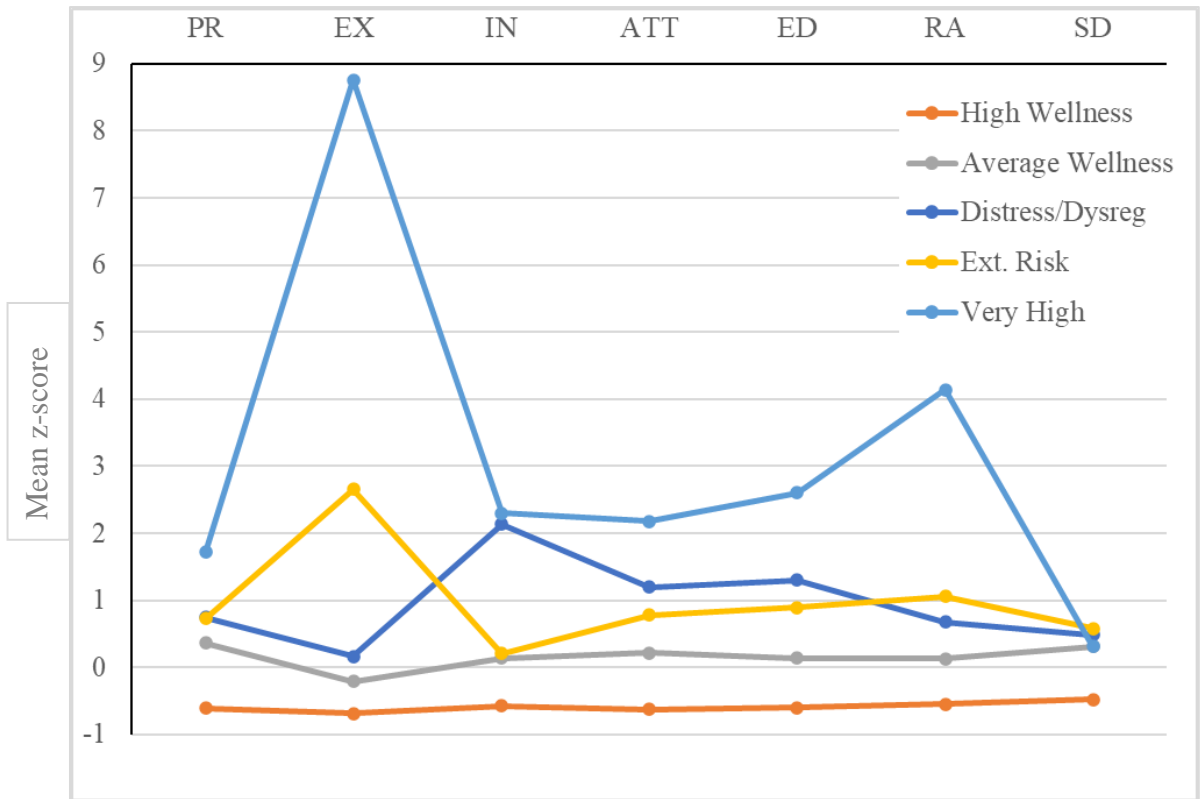


Figure 2. Mean z-scores of the five-profile LPA solution across EIS-SR indicators.

## VITA

Caroline Hodgson was born in Miami, Florida on May 31, 1993 to David and Karen Grossman. After graduating from Miami Palmetto Senior High School in 2011, Caroline attended Washington University in St. Louis, where she earned a bachelor's degree in Philosophy-Neuroscience-Psychology and minored in Religious Studies and Public Health. During this time, she had the opportunity to learn about the overlapping systems that impact children's well-being, including the public health burden of mental health concerns, educational inequity, and other potential barriers to their flourishing.

Following her graduation in 2015, Caroline entered the doctoral program in School Psychology at the University of Missouri in August 2015. She earned her master's degree in August 2017 and married Tony Hodgson in October 2017. Caroline completed her predoctoral internship in Child Clinical Psychology at Nationwide Children's Hospital in Columbus, Ohio. She secured a postdoctoral fellowship at Nationwide Children's Hospital in Pediatric Acute Treatment Psychology beginning in August 2020. Her research and clinical interests center around developing innovative and efficient ways to deliver the highest-quality care to target internalizing concerns in youth, whether through population-level approaches, prevention activities, or data science.