March 2023

Identifying Contributors to Disproportionality: The Influence of Perception on Student Social, Emotional, and Academic Behavior Ratings

Chelsea Salvatore
University of South Florida

Identifying Contributors to Disproportionality:

The Influence of Perception on Student Social, Emotional, and Academic Behavior Ratings

by

Chelsea Salvatore

A thesis submitted in partial fulfillment
of the requirements for the degree of
Education Specialist
with a concentration in School Psychology
Department of Educational and Psychological Studies
College of Education
University of South Florida

Major Professor: Nathaniel von der Embse, Ph.D.
Eunsook Kim, Ph.D.
Jose Castillo, Ph.D.

Date of Approval:
March 28, 2023

Keywords: utility, perspective, universal screening, behavioral risk

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Successful implementation of the MTSS framework involves equitable assessment and decision-making. This calls for researchers to investigate the processes utilized to identify students for early behavioral support, such as multi-informant universal social-emotional behavioral screening. The current study aimed to investigate this by first examining the usability of the Social, Academic, and Emotional Behavior Subscale (SAEBRS) and the student self-report version (mySAEBRS) across Black and White students in grades seven through nine. After examining these tools for measurement invariance, the trifactor model was employed with race as a predictor variable to examine the mean difference of Black and White students' scores in terms of the common factor, and unique perspective factors. The SAEBRS and mySAEBRS Academic and Emotional Behavior Subscales were invariant across Black and White students in the sample, implying that these subscales may be culturally responsive to these populations. The SAEBRS and mySAEBRS Social Behavior Subscale, however, did not show measurement invariance across Black and White students, and thus, could not be included in the trifactor model evaluation. Black students scored lower than White students similarly across all three factors on the Academic Behavior Subscales. On the Emotional Behavior Subscales, Black students were rated higher than White students in terms of the student perspective factor but were significantly rated lower than White students in terms of the teacher perspective factor. These findings may imply an influence of implicit racial bias on student SAEBRS scores, but this could not be distinguished from the influence of other teacher or school characteristics.

CHAPTER 1: INTRODUCTION

Statement of the Problem

Students who are racially minoritized experience substantially worse school outcomes. Public high school graduation rates in the United States show significant differences when disaggregated by student race/ethnicity (National Center for Education Statistics [NCES], 2021). For example, data from the 2018-2019 school year demonstrate that 86% of students graduate with a regular diploma within four years of beginning 9th grade. However, when examined by student race/ethnicity, White and Asian/Pacific Islander students have graduation rates of 89% and 93%, whereas graduation rates for Hispanic (82%), Black (80%), and American Indian/ Alaska Native (74%) students fell below the average (NCES, 2021). One factor contributing to differences in timely graduation is the unaddressed behavioral needs of particular student subgroups with challenging classroom behavior (e.g., inattentiveness) being associated with worse academic outcomes (Rabiner et al., 2016). Exclusionary discipline practices are one such factor that is negatively related to student academic outcomes (Morris & Perry, 2016); students with one suspension are more likely to experience later suspensions, failing course grades, chronic absenteeism, and dropout (Balfanz et al., 2014). This may be especially concerning for Black students since this subpopulation receives significantly more punitive disciplinary sanctions than students from other races (Fabelo et al., 2011). This disparity remains even when controlling for socioeconomic status (Skiba et al., 20002) and considering differences in the number of misbehaviors by student race (Huang, 2018).

In alignment with these findings, Black students are more than twice as likely as White peers to receive disciplinary action, including office discipline referrals (ODR) (Young & Butler, 2018). This is especially likely to be the case when subjective ODRs, such as those for disrespect or defiance, are used compared to objective ODRs, such as those for fighting or theft (Smolkowski et al., 2016). This is particularly significant since schools have traditionally used ODRs to identify students for mental health or behavioral services (Girvan et al., 2017). Given prior research controlling for variables such as socioeconomic status (Skiba et al., 2002) and the number of behavior incidents (Huang, 2018), implicit racial bias is another mechanism that should be examined for contributions to the existing disparities.

Several empirical studies have demonstrated the potential influence of educators' implicit racial bias (Okonofua & Eberhardt, 2015). In one study, K-12 teachers viewed an ODR record belonging to a student with either a stereotypical Black or White name and reported feeling similarly across incidents involving White students but reported feeling more troubled and more likely to use a punitive disciplinary action for Black students following the second incident (Okonofua & Eberhardt, 2015). The potential influence of implicit racial bias in ODRs leads to the examination of contemporary approaches.

With the increasing adoption of Multi-Tiered Systems of Support (MTSS) in schools across the United States, more schools are likely to engage in universal social, emotional, and behavioral (SEB) screening to identify students for mental health and behavioral services. However, these assessment tools also must be investigated for the potential influence of implicit racial bias. Before examining rater bias, the assessment tool must be tested for bias. Measurement invariance studies explore the usability of assessment tools across subpopulations (Pendergast et al., 2017). On the BASC-2 Behavioral and Emotional Screening System (BESS)-

Teacher Form- Child/Adolescent (Kamphaus & Reynolds, 2007), elementary school teachers rated male and Black students at significantly higher risk for behavioral and emotional challenges compared to female, Hispanic, and White peers (Splett et al., 2018). Teachers may have differential expectations for students based on race. Due to this, it may be essential to disaggregate ratings by race when examining teacher ratings of student behavior. It is best practice to gather multiple informant reports in the psychopathological assessment of youth (Mash & Hunsley, 2005), and reports will likely vary between informants (De Los Reyes, 2011). Therefore, there is also value in investigating the extent to which teacher or student perspectives influence SEB ratings. Identifying unique perspective factors may provide insight into the extent to which teachers and students perceive Black students to have worse social, emotional, or behavioral functioning than White students. Rater bias is the presence of systematic error in ratings caused by rater attitudes, beliefs, or experiences (Hoyt, 2000). While the influence of unique informant perspectives on student scores across racial subgroups is not an established measure of rater implicit racial bias, Bauer et al. (2013) suggest informant perspectives are equivalent to rater bias, and research is needed to investigate it as such.

Defining Key Terms

Social, Emotional, and Behavioral (SEB) Functioning

From a dual-factor approach, SEB functioning encompasses both strengths and challenges (Suldo & Shaffer, 2008). Traditional methods take a deficit approach and assume mental health in the absence of symptoms of psychopathology (Suldo & Shaffer, 2008). The dual-factor approach, however, considers the presence of problem behaviors as well as indicators of subjective well-being (Suldo & Shaffer, 2008). This allows schools to provide preventative programs based on student strengths while also identifying students needing support (Kim & Choe, 2022).

Universal Social, Emotional, and Behavioral (SEB) Screening

Universal SEB screening, as opposed to universal academic screening, entails a behavior rating or nomination process, completed by one or more informants, that considers all students at the school for additional support (Dineen et al., 2022). The universal screening tool used within the current study is the Social, Academic, and Emotional Behavioral Risk Screener (SAEBRS; Kilgus & von der Embse, 2014).

Implicit Racial Bias

Implicit racial bias is the unconscious associations related to race, formed through personal experiences and exposure to media, that influence our understanding and behavior involuntarily and do not necessarily align with our explicit beliefs (Staats, 2014).

Disproportionality

Disproportionality is observed within data when a percentage of a group is identified at a larger or smaller percentage than that group represents within the population (Gabel et al., 2009). Researchers use risk ratios to examine disproportionality by comparing the percentage of a minority group to the percentage a comparison group makes up (Farkas & Morgan, 2018). Groups experiencing disproportionality are likely to be underserved or overrepresented among adverse outcomes.

Measurement Invariance

Measurement invariance is a key aspect of scale development, and testing for this determines the appropriateness of use across subpopulations (Meredith, 1993). A tool is concluded to hold measurement invariance if the underlying construct is measured the same across different groups or time points (Pendergast et al., 2017). Race was the focus of the measurement invariance analysis.

Purpose of Current Study

The proposed study intended to evaluate the potential influence of implicit racial bias on student SEB scores. Previously, von der Embse et al. (2021) conducted a study in which student self-ratings were compared to teacher ratings of student behavior and examined how student demographics relate to student outcomes. This study identified male, non-White, and special education students as more likely to be rated by themselves and their teacher as having worse behavioral functioning. Additionally, significant incongruence between teacher and student scores indicates a need for schools to consider both teacher and student self-report scores when identifying students in need of services. von der Embse et al. (2019) used student and teacher-matched data to examine SAEBRS measurement invariance across grade levels, gender, and special education status, as well as used a trifactor model separate the variance across the common factor, which represents the shared perspective between informants, the informant perspective factors, which represents the unique perspectives across informants, and the item-specific factors, which are related directly to the uniqueness of the item itself. von der Embse (2019) examined age, gender, and special education status as predictors to explain the relative variance of each of these factors. The proposed study intended to expand on these findings by examining the SAEBRS and mySAEBRS for measurement invariance across race and using a trifactor model to examine student race as a predictor variable to explain the variance of each factor. It was hypothesized that Black students would be rated consistently lower on each of the three subscales in terms of the informant perspective factors.

The specific research questions to be answered in this study are the following:

- Do the SAEBRS and mySAEBRS subscales display measurement invariance across Black and White students in grades seven through nine?
- Does student race predict how Black and White students in grades seven through nine are scored on the SAEBRS and mySAEBRS subscales?
    - Do Black students receive systematically lower scores compared to White students in terms of the teacher perspective factor?
    - Do Black students receive systematically lower scores compared to White students in terms of the student perspective factor?
    - Is there any difference between Black and White students in terms of the common factor?

Contributions to the Literature

This study intended to enhance the research on the utility of the SAEBRS, specifically by examining differences across racial groups. While current research investigates disproportionality by comparing the percentage of those identified from one subgroup to the percentage of the population that group makes up, the present study took this a step further by considering if informant perspectives contribute to this disproportionality. Given that there is little known about the influence of informant perspectives on student SEB screening scores, the findings from the proposed study can advance the literature in this area. This is especially valuable because informant perspectives have been suggested to be equivalent to rater bias (Bauer et al., 2013). Through the separation of the common factor, item-specific factor, and informant perspective factors, the present study showed the extent to which informant perspectives influence how students are rated, which may be indicative of rater implicit racial bias.

CHAPTER 2: LITERATURE REVIEW

This chapter is a review of the literature leading up to the proposed research questions. First, it will be necessary to understand the Multi-Tiered System of Supports (MTSS) framework utilized in many schools and how universal screening fits into this framework. Next, traditional methods for identifying students at risk of behavioral or mental health challenges will be reviewed before presenting current methods to proactively identify students. The potential influence of implicit racial bias on each method will be discussed, and the SEB assessment tool used in the current study will be reviewed. Finally, this chapter will discuss the current study's approach to examining the utility of the SAEBRS across racial subgroups.

Multi-Tiered System of Supports

As schools continue to discover ways to effectively support their students, many have adopted an MTSS framework. MTSS is a framework for continuous improvement and is defined as an evidence-based model that utilizes problem-solving techniques to develop appropriate instruction and interventions that meet students' academic and behavioral needs (Gamm et al., 2012). Before the development of MTSS, schools typically employed other problem-solving frameworks, such as Response to Intervention (RtI; Brown-Chidsey & Steege; 2010) and Positive Behavioral Interventions and Supports (PBIS; Sugai & Horner, 2009). Schools often implement RtI for academic support and PBIS for behavioral support. MTSS provides a system in which these PBIS and RtI are essentially integrated. There are many similarities between RtI and PBIS, such as using data-based decision-making and instruction as a preventative service for all students. However, there are differences, such as in terminology and procedures (McIntosh &

Goodman, 2016). Since academic and behavioral difficulties coincide and influence one another, an integrated system is likely more effective and feasible than implementing two separate initiatives that each require financial resources and educator time (Eagle et al., 2015).

The MTSS framework comprises three Tiers of service. Tier 1 consists of services such as core instruction and universal behavior support. All students at the school are offered these services, with about 80% being successfully supported at this tier of service. To improve Tier 1, environmental factors, such as instructional curricula, and the implementation of universal programming, such as a social-emotional learning curriculum, should be considered. For example, a school with students exhibiting a high number of behavior challenges may examine the buy-in, feasibility, and fidelity of implementation for the currently existing Tier 1 services. It is essential that schools strengthen this Tier to prevent resource expenditure at the other tiers due to high need. Tier 2 consists of supplemental instruction/support in addition to universal instruction and support students continue to receive through Tier 1 services. This tier of service is ideally provided to about 10-15% of students. These services consist of extra materials provided individually or in a small group format by the teacher or other school personnel if available. This tier offers additional general education materials to assist with students' learning rather than the materials being personalized to the student. About 5% of students receive intensive, individualized interventions at Tier 3. Students at this tier receive individualized support while continuing to receive universal instruction/ support at Tier 1 and supplemental support through Tier 2 (Tilly, 2008).

One way schools can identify students who are not responding to Tier 1 and may benefit from Tier 2 or Tier 3 support is through universal screening (Marsh & Mathur, 2020). This is a method by which all students are rated by an informant on their social-emotional, academic, and/or behavioral functioning (Kilgus & Eklund, 2016). Although the research literature supports the use of universal screening within an MTSS framework (Eagle et al., 2015), Briesch et al. (2020) examined the universal screening implementation guidelines provided by the state departments of education across the United States. Many states do not provide guidelines on how schools should implement universal SEB screening (Briesch et al., 2020). Of these guidelines, 82% of states had some mention of universal SEB screening, and 14% of states reported universal screening as an essential component of MTSS but did not provide further detail on what screening should entail. Also, 43% of states provided some information on SEB screening with an MTSS framework, and out of these states, 31% provided information on the implementation of an SEB screener with varying levels of specificity, ranging from general information to specific recommendations. Only one state mandates universal screening within an MTSS context (Briesch et al., 2020). Along with insufficient guidelines on universal SEB screening, school psychologists have reported feeling unprepared to conduct universal screening in school (Burns & Rapee, 2021).

Traditional Methods for Student Identification

Traditionally, many schools have used office discipline referrals (ODRs) to identify students at risk for social-emotional and behavioral difficulties (Predy et al., 2014). ODRs are standardized school forms in which a teacher can record a problem behavior displayed by a student. These forms typically include a section to describe the behavior, time, location, and date it occurred. Previous research has supported using ODRs to identify students for behavioral support (McIntosh et al., 2009). ODRs may provide information on the root causes of ODRs (McIntosh et al., 2018), which is more objective than personal discretion (Fenning & Jenkins, 2017). Discovering the root cause of ODRs is worthwhile because students disciplined with exclusionary practices for the same behavior as other students are less likely to acquire SEB skills valued by the school (Anyon et al., 2014; Reyes et al., 2013). Middle school students with at least one ODR during the Fall months significantly predicted between two to six ODRs by the end of the school year (Predy et al., 2014). Although a few students were noted as false positives, meaning they did not end the year with moderate or high levels of ODRs, the overall use of ODRs to identify students with future behavioral problems was supported (Predy et al., 2014). Similarly, elementary students with two or more ODRs by October were likely to end the school year with at least 6 ODRs (McIntosh et al., 2010). Based on these findings, children who receive ODRs early on are likely to receive more as the year progresses and may be good candidates to receive social-emotional or behavioral support.

ODR use for this purpose may be particularly beneficial within a School-Wide Information System (SWIS; May et al., 2000). This data management system allows for ODRs to be systematically administered and tracked. This system's operational definitions for ODRs may be beneficial in minimizing the subjective use of ODRs. The system can also monitor ODR usage at a school-wide, classroom, or individual level. Additional benefits of this system include the early identification of problem behaviors and the capacity to generate reports that provide details on ODRs given at certain locations of the school and at various times of the day (Irvin et al., 2006). Across four participating school districts, between 62% and 77% of elementary school staff reported using SWIS ODR data on a weekly to monthly basis to support decision-making, and between 88% to 99% of middle school staff reported the same frequency of use (Irvin et al., 2006). Implementing the SWIS has been associated with decreasing ODRs for some students (Hawkens, 2018).

Critique of Traditional Screening Methods

Using ODRs to identify students for support may lead to significant amounts of students being overlooked (McIntosh et al., 2010). ODRs also may not capture the needs of students with internalizing concerns, as these may be less noticeable and disruptive and, in turn, less likely to warrant an ODR by the teacher (Splett et al., 2018). This means that students with internalizing concerns are less likely to be referred for social-emotional or behavioral support if ODRs are used to determine who receives these services (McIntosh et al., 2009). Some groups of students may be overidentified due to the different types of ODRs used at schools. Schools have reported using primarily subjective ODRs, for behaviors such as disrespect or defiance rather than objective ODRs for behaviors such as theft or fighting. Specifically, students receive one objective ODR for every 7.2 subjective ODRs (Smolkowski et al., 2016). This is concerning since Black students are significantly more likely to receive subjective ODRs than White peers (Smolkowski et al., 2016).

Further, ODR data show significant disproportionality (Mizel et al., 2016), which is present when members of a group are identified at a higher percentage than they represent within the population (Nowicki, 2018). Black students are significantly disproportionally represented within ODR data, being twice as likely than White peers to receive an ODR at the elementary level and four times as likely than White peers to receive an ODR at the middle school level (Skiba et al., 2011). Even when school teams are provided with SWIS reports to review, disciplinary data may not be connected to meaningful change regarding disciplinary equity (McIntosh et al., 2020). This is particularly problematic since some research has supported using ODRs to identify students for behavioral support (Irvan et al., 2006; McIntosh et al., 2009).

Implicit bias is a substantial factor influencing racial disproportionality among ODRs (Girvan et al., 2017).

Implicit Racial Bias in Traditional Screening Methods

Implicit bias is "the attitudes and associations made subconsciously that often do not reflect 'actual,' or explicit, beliefs" (Beachum & Gullo, 2020). An attitude is a belief that predisposes one to think, feel and respond in a particular way to an object, which is a person, behavior, or event (Eagly & Chaiken, 1993). Mental processing of an object occurs along two distinct paths, with implicit attitudes being spontaneous and automatic associations in the presence of the object, such as automatically driving in the presence of a green light, and explicit attitudes being conscious and controlled processing, such as selecting a destination to drive to (Kahneman, 2011). An implicit attitude, or bias, is formed through experiences, such as exposure to media, parental views, and common stereotypes that shape views of oneself and the world (Harro, 2000). Since implicit biases are typically held in our unconscious mind, we are often unaware of them, which is why someone who believes in human equality may unknowingly behave in a way that does not facilitate this (Staats, 2014). Implicit racial bias refers to the attitudes developed through the cognitive process of assigning characteristics to others based on appearance or race (Russell-Brown, 2018).

Black students may be at heightened risk of experiencing automatic responses related to implicit racial bias due to the negative stereotypes promoted within the media. With the average American watching more than five hours of television a day (Koblin, 2016), individuals are likely to be exposed to negative stereotypes that support the development of implicit racial bias. Black people are overrepresented in television news as criminal suspects, and viewing this overrepresentation is positively associated with the perception of Black people as violent (Dixon, 2008). While teachers' roles position them to promote equitable outcomes for their students, teachers' levels of implicit and explicit bias are similar to those of nonteachers (Starck et al., 2020). Implicit bias can be difficult to detect due to its subconscious nature. However, it can be observed in student disciplinary data.

Disproportionality in student disciplinary data is often represented within patterns at two key points in the discipline decision-making process: 1) the differential selection of racially minoritized students for ODRs and 2) the differential assignment of severe disciplinary actions for racially minoritized students, such as suspension and expulsion (Anyon et al., 2014). In addition to Black students being significantly more likely to receive subjective ODRs than White peers, the majority of the variance in the total disproportionality can be explained by subjective ODRs (Girvan et al., 2017), indicating that disproportionality is likely related to teachers' differential perceptions and can in part be attributed to implicit bias. Black students are significantly more likely than White peers to receive ODRs even when controlling for students' level of teacher-rated behavior problems, teacher ethnicity, and other classroom factors (Bradshaw et al., 2010). Similar to the differential selection of racially minoritized students for ODRs, differential processing of these students occurs, meaning there is disproportionality in student disciplinary actions. For example, while Black students make up about 15.5% of public school students, they disproportionately represent about 39% of students suspended from school (Nowicki, 2018).

Further, Black students with disabilities represent about 19% of K-12 students in the United States, but this group makes up about 36% of students with disabilities who have been suspended from school (Nowicki, 2018). Students who are boys, are Black and whose parents have less education are at increased risk of suspension/expulsion (Mitzel et al., 2016). At the elementary, middle, and high school levels, Black boys are significantly overrepresented in suspensions, and Black girls are suspended at higher rates than White and Hispanic girls (Mendez & Knoff, 2003). Even when controlling for other student variables, such as socioeconomic status and base rates, Black students still receive significantly higher rates of suspensions (Anyon et al., 2014). Additionally, while 5% of suspensions are due to violent behavior, 51% result from disruptive behavior (Skiba & Rausch, 2006).

Implicit racial bias may influence the selection of students for ODRs and student disciplinary actions. This may be seen within teacher expectations for behavior challenges. Students held to lower expectations by their teachers are likely to receive more ODRs than students rated highly by their teachers, regardless of race (Santiago-Rosario et al., 2021). When accounting for race, Black students in this sample received more ODRs than white peers, and teacher expectations explained about 21% of the difference (Santiago-Rosario et al., 2021). Teachers may watch Black students more closely for misbehavior due to implicit bias (Gilliam et al., 2016), and teachers may be more likely to discipline Black students when a similar behavior is displayed by a White student (Scott et al., 2019). Also, teachers report being more worried about misbehavior and are more likely to view the misbehavior as a pattern when performed by a Black student than a White student (Okonofua & Eberhardt, 2015). Regarding student disciplinary actions, administrator implicit biases account for about 19% of the differences in the severity of subjective discipline for Black students (Gullo & Beachum, 2020). Additionally, high levels of racial biases within the community are associated with racial disproportionality of both ODRs and out-of-school suspensions (Girvan et al., 2021). Disproportionality in discipline data is likely related to varying teacher expectations for student behavior across settings and varying thresholds for behaviors that constitute the need for an ODR or exclusionary discipline (Wang et al., 2018). These data suggest that ODRs are not the most effective or equitable method for identifying students at risk of future behavioral challenges.

Contemporary Methods for Student Identification

Universal screening, however, has been deemed an effective method for identifying students with behavioral, emotional, or academic concerns. Universal Social-Emotional and Behavioral (SEB) screening provides all students with the opportunity to be assessed on their social-emotional and behavioral functioning to identify students who may be at risk of mental health or behavioral challenges. This is a proactive approach compared to traditional methods that follow the "wait-to-fail" model, for which students are provided support only once they are identified from a referral due to significant problems (Glover & Albers, 2007). Typically, SEB screening consists of either a brief behavior rating scale completed by an informant (i.e., teacher, parent, student self-report) or teacher nominations along with a risk assessment before selecting an appropriate intervention (Dineen et al., 2022). This method provides a data-driven approach for assessing student and school-level functioning (Siceloff et al., 2017). For example, universal SEB screening serves as a "temperature check" on the percentage of students at risk across different areas of the school (i.e., class-wide, by grade level, school-wide) to inform the selection of evidence-based interventions that best match student needs. Regarding school-level functioning, universal screening data can inform the effectiveness of Tier 1 interventions. Specifically, if many students (>20%) across the school are identified as "at-risk," the Tier 1 supports currently in place can be evaluated for their fidelity of implementation and efficacy for use with the present student population. It will be necessary for key partners, such as the teachers, administration, and parents, to be bought into the idea of universal SEB screening. It is especially crucial for the informant, which is typically the student's teacher, to value this process.

Informants can be provided with logistical information to increase the perceived feasibility of screening. Also, sharing the benefits of this process may make teachers feel less burdened to complete the screener for their students (Siceloff et al., 2017), which may, in turn, be related to more thoughtful and thorough ratings of students. There are several existing universal SEB screeners. Schools can evaluate their needs and select an appropriate assessment tool while considering technical adequacy, cost, and usability, such as feasibility and acceptability (Glover & Albers, 2007).

To understand the effectiveness of universal screening for identifying students, Splett et al. (2018) implemented universal screening to investigate how many additional students would be identified as being at risk for mental health concerns. Out of 3,744 students across six elementary schools, 679 students were newly identified as in need of mental health support. This is a 180.1% increase in students identified from the 377 students previously identified within the sample (Splett et al., 2018). It is also notable that many of the students identified from this had higher academic grades, were less likely to be male, and were likely to have fewer ODRs compared to peers who were not identified (Splett et al., 2018). These and similar findings by Eklund and Dowdy (2014) highlight the wide range of students who might be overlooked without using universal mental health screening. Additionally, Splett et al. (2018) highlight that using ODRs to identify students may provide less useful information. Students identified as at risk for challenges on an SEB screener at the beginning of the year are more likely to struggle academically at the end of the year compared to students identified by ODRs (Naser et al., 2018). This indicates that SEB screeners may be better than ODRs at predicting students in need of academic support. Although the results support the use of universal screening in schools, this research also indicates for schools to be aware of the resources available prior to screening due to the high number that may be identified as needing additional support.

While using ODRs to identify students strictly focuses on behavioral concerns, many universal SEB screeners require respondents to report on both student challenges and strengths. This approach follows the dual-factor model in which mental health is comprised of two separate but related constructs, namely psychopathology and subjective well-being (Suldo & Shaffer, 2008). Psychopathology involves the presence of internalizing disorders (e.g., anxiety) or externalizing disorders (e.g., oppositional defiance disorder). At the same time, subjective well-being is comprised of life satisfaction, positive affect, and negative affect (Suldo & Shaffer, 2008). Those with high levels of subjective well-being might report overall positive feelings towards their quality of life and report experiencing positive affect more frequently than negative (Suldo & Shaffer, 2008). The dual-factor model highlights the importance of examining wellness rather than only illness. The dual-factor approach is incorporated into universal SEB screening by including items assessing for both SEB challenges and competence. For example, assessing social behavior might include items looking at temper outbursts and pro-social behavior. While deficit-oriented approaches are likely to identify students at risk for psychopathology, students experiencing low levels of well-being are likely to be overlooked (Antaramian et al., 2010). The dual-factor approach, on the other hand, can capture the needs of students who are experiencing low levels of psychopathology but are also experiencing low levels of well-being (Antaramian et al., 2010). This approach can assist schools in identifying the most at-risk students who are experiencing both low levels of psychopathology and subjective well-being (Antaramian et al., 2010). Additionally, the dual-factor approach within universal SEB screening creates the opportunity for educators to implement interventions that capitalize on student strengths (Humphrey & Wigelsworth, 2016).

Universal SEB screening also provides the opportunity to incorporate data from multiple informants. Teachers are typically selected for this role due to the extensive time spent with the student each school day. A parent or guardian may also be asked to report on their child's behavior, or the student may be asked to self-report on their behavioral functioning. Having only one informant can save time and valuable resources that can be used to support additional students (Taylor et al., 2023). On the contrary, it may also be beneficial to view differences in ratings among informants (De Los Reyes, 2011).

Critique of Contemporary Screening Methods

Universal screening is a core component of MTSS, yet, many schools do not engage in this method of identifying students (Wood & McDaniel, 2020). Universal screening for academic concerns has been well-established in schools due to policies such as the No Child Left Behind Act of 2001, which uses language directly indicating for schools to universally screen students to identify those at risk of school failure (Cook et al., 2010). However, schools have been slower to adopt universal SEB screening (Briesch et al., 2020). From examining 245 schools, 98.8% of principals reported that they do not actively use universal mental health screening (Wood & McDaniel, 2020). While the primary reason for this was lack of access or funds (Wood & McDaniel, 2020), the stark percentage from this study raises concern. Additional research surveying superintendents across 12,470 districts found that the most used method of identifying students for social-emotional or behavioral support was an internal referral (54.70%), followed by an external referral (12.11%), with school-wide screening reported the least (5.54%) (Dineen et al., 2022).

In addition to schools' limited use of universal SEB screening (Dineen et al., 2022), limited research exists on the usability of SEB assessments (Brann et al., 2022). Usability is comprised of multiple components, including feasibility, acceptability, treatment utility, cost analysis, cultural responsiveness, accommodation considerations, and support for tiered problem-solving (Brann et al., 2022). Cultural responsiveness involves examining a measure for use with marginalized populations, such as the extent to which it reduces disparities in access to SEB supports or investigating screening data for disproportionality. Because of this, the lack of research on cultural responsiveness among universal SEB screeners is of particular concern to the current study (Brann et al., 2022). Studies showing support for cultural responsiveness showed insignificant levels of differential item functioning (DIF) across racial and ethnic groups (Brann et al., 2022). That is, the items on the culturally responsive screeners measure the construct the same way across race and ethnicity. DIF is a method of examining measurement invariance, or the extent to which the psychometric properties of the observed indicators are generalizable across groups or over time (Jeong & Lee, 2019). The lack of research investigating measurement invariance in universal SEB screening indicates the potential for bias in the construct validity of these tools, meaning a variable score may be systematically over or under-estimated for a particular group and, thus, impact their overall score for that trait (Reynolds & Ramsay, 2003). A primary aim of the current study will be to address the usability of the SAEBRS across racial groups by first testing the model's fit before testing for measurement invariance across subgroups.

Measurement Invariance

Measurement invariance in universal SEB screening tools is necessary for making accurate predictions across groups (Millsap & Kwok, 2004). Specifically, students from a particular group may be over- or under-identified. On a universal SEB screener, students are assessed for SEB risk. Unobservable constructs, such as SEB risk, are called latent variables or factors. Since these are unobservable, inferences are made about students' level of SEB risk based on indicators, such as responses on a universal SEB screener. Measurement invariance testing investigates the observed indicators for systematic differences between particular groups of students and is crucial to the selection of students for SEB support. Without measurement invariance, "fairness and equity cannot exist. Thus, when the purpose of test use is selection of individuals, measurement invariance is a necessary condition for fair selection procedures" (Borsboom et al., 2006, p. 179).

Measurement invariance, or more specifically DIF, has been investigated within universal SEB screeners such as the Emotional and Behavioral Screener (EBS; Cullinan & Epstein, 2013). The degree of DIF identified among White, Black, and Hispanic students showed that EBS scores were invariant across racial/ethnic groups (Lambert et al., 2018). The EBS has been reported as demonstrating adequate evidence of bias evaluation, but many universal SEB rating scales have not evaluated any potential bias in administration or scoring procedures (Houri & Miller, 2020). The Behavior Assessment System for Children, Second Edition (BASC-2) Behavioral and Emotional Screening System Teacher Form (BESS; Kamphaus & Reynolds, 2007) has also been examined for measurement invariance. Responses to items on the Student Form were measurement invariant across ethnicities, English language proficiency classification, and socioeconomic status (Harrell-Williams et al., 2015), while the Teacher Form showed measurement invariance across English language proficiency classification (Dowdy et al., 2011). Likert-type items with less than five levels, such as those commonly used in universal SEB screening (e.g., never, sometimes, often, almost always), can be treated as continuous data with the maximum likelihood estimation (Rhemtulla et al., 2012), and is best to be conducted with Multigroup Confirmatory Factor Analysis (MG-CFA) (Pendergast et al., 2017). The Behavior Assessment System for Children, Third Edition (BASC-3) BESS (Kamphaus & Reynolds, 2007) has been assessed for measurement invariance using MG-CFA. This SEB screener holds measurement invariance across elementary-aged children. However, this screener has yet to be investigated for measurement invariance across race as is intended by the current study. Establishing measurement invariance is meant to limit assessment bias. At the same time, measurement invariance is separate from selection invariance (Pendergast et al., 2017), meaning there may be bias in which students are selected for SEB supports.

Rater Discrepancies

Beyond investigating the generalizability of a tool's psychometric properties across groups, it is important to investigate rater discrepancies within multi-informant screening data. It is best practice to gather multiple informant reports when assessing youth psychopathology (Mash & Hunsley, 2005), and reports will likely vary between informants (De Los Reyes, 2011). For example, students may have a better awareness of their functioning, as implied by significant discrepancies between student and parent ratings of social competence (Renk & Phares, 2004), social anxiety (Deros et al., 2018), and symptoms of depression (Makol & Polo, 2018).

The Needs-to-Goals Gap framework addresses rater discrepancies (De Los Reyes et al., 2022). This framework is intended to reduce the research-to-practice gap regarding how multi-informant data are gathered and interpreted to better connect individual students' needs to the goals of services they receive (De Los Reyes et al., 2022). This framework is directly informed by the Operations Triad Model (OTM; De Los Reyes et al., 2013), which has three key components for identifying and understanding patterns in multi-informant reports. Converging

Operations characterize these patterns when findings are interpreted based on the extent to which various informants' reports point to the same conclusion (De Los Reyes et al., 2013). A biased conclusion may be drawn if only one informant is considered, whereas examining data across multiple informants provides a complete picture of the student and minimizes the risk of bias (De Los Reyes et al., 2013).

Another component of the OTM, Diverging Operations, guides the interpretation of informant discrepancy patterns in muti-informant reports (De Los Reyes et al., 2013). This concept considers differences in ratings as providing a meaningful variation of the behavior and considers how discrepancies are related to the influence of context on a child's behavior and the variability in individuals' perception of the behavior (De Los Reyes et al., 2013). This indicates that rater discrepancies may be at least partially attributable to the differences in behavior displayed and observed across various social contexts rather than informant bias (De Los Reyes et al., 2013). For example, Diverging Operations have been reflected in informant discrepancies between parent and teacher ratings of a child's aggressive behavior (Hartley et al., 2011). Similarities between environments were related to increased correspondence among informants. In contrast, differences in ratings were related to individual perceptions of the environmental cues for aggression, such as negative peer interactions or adult demands placed on the child (Hartley et al., 2011). The research on Diverging Operations implies that differences between informant reports may partly be due to the meaningful differences in the environments that the behavior is observed, which is referred to as domain-relevant information (De Los Reyes et al., 2022). If decision-making relies on only one informant, valuable domain-relevant information may be excluded. If discrepancies are not found to provide meaningful information about behavior, then the third component of the OTM can be investigated. Examining Compensating

Operation entails running further statistical analyses to determine if the observations are best explained by measurement error (De Los Reyes et al., 2013). Most multi-informant research over the past few decades has utilized Converging Operations to interpret consensus across raters as evidence of the tool functioning the same way across each student and to inform decision-making (von der Embse & De Los Reyes, 2022, *in press*). Similar to how teacher perspective may influence which students receive ODRs, teacher and student perspectives may influence student SEB scores and, ultimately, which students receive SEB support. Teacher or student bias may best explain the meaningful differences in multi-informant ratings. Rater difference does not guarantee the influence of bias. However, using a trifactor model to separate the item-specific factor, common factor, and informant perspective factors allows one to examine whether student race predicts variance in unique teacher or student perspective factors. This is suggested to be equivalent to rater bias (Bauer et al., 2013) and, more specifically, may be considered to be implicit racial bias due to the focus on student race.

Current Study Approach to SEB Screening

The Needs-to-Goals Gap framework (De Los Reyes et al., 2022) underlies the current study. The Social, Academic, and Emotional Behavior Risk Screener (SAEBRS; Kilgus & von der Embse, 2014) is a universal SEB screening tool that provides the option for multiple informant ratings by including a teacher, parent, and student self-report version. While many existing universal assessment tools screen for mental health concerns or academic challenges, the SAEBRS acknowledges that these domains are interrelated. The SAEBRS teacher rating scale (SAEBRS-TRS) is a 19-item measure for which teachers use a 4-point Likert scale (1= never, 4= almost always) to rate their students on several items for each domain. Students receive a score for each of the three domains to determine risk for behavioral or mental health concerns, as well as a total score indicative of overall student functioning (Kilgus & von der Embse, 2014).

This screening tool has been examined for appropriateness, technical adequacy, and usability, which are the three essential considerations when determining if an instrument should be used (Glover & Albers, 2007). In terms of usability, the SAEBRS-TRS was designed to be feasible for teachers as it is expected that they are able to complete ratings in approximately one to three minutes per student. In addition, the usability of this tool can be improved further by schools by focusing on what is needed. For example, schools can select to assess only social and emotional behavior to limit the number of items the teacher must complete. The SAEBRS-TRS includes both adaptive and maladaptive characteristics. Per the dual-factor model of mental health, it is necessary to assess both psychological functioning and subjective well-being (Suldo & Shaffer, 2008). Complete mental health, or low levels of psychological functioning along with high levels of subjective well-being, best predicts positive student outcomes (Suldo & Shaffer, 2008). This supports using the SAEBRS-TRS for predicting student functioning as this measure provides information on both continuums.

The technical adequacy of this tool has also been examined. In one study of 346 students in grades 3 to 5, teachers completed the SAEBRS-TRS and two criterion measures to investigate the technical adequacy of the SAEBRS-TRS (Kilgus et al., 2016a). This research supports using the SAEBRS-TRS for assessing students within each of the three domains. High internal consistency was found for all three subscales and the total scale. Also, statistically significant correlations were seen between the SAEBRS-TRS and both criterion scales, indicating concurrent validity. In addition, student scores were moderately to highly correlated with academic outcomes (Kilgus et al., 2016a). This finding supports the need to screen students for early behavior challenges in each of the three domains (social, academic, and emotional) as they are interrelated. Further investigations of the SAEBRS-TRS have supported its use. A sample of 567 elementary and 297 middle school students who received SAEBRS-TRS ratings from their teachers showed support for internal consistency, concurrent validity, and diagnostic adequacy (Kilgus et al., 2016b).

The proposed study also examined student self-ratings via the Social Academic and Emotional Behavior Risk Screener-Student Rating Scale (SAEBRS-SRS; von der Embse et al., 2017a). This is a 20-item measure that was developed based on the SAEBRS-TRS to strengthen the reliability of SAEBRS screening scores. Scores on the SAEBRS-TRS can be compared to the SAEBRS-SRS to determine if there is an agreement between the two. The SAEBRS-SRS has strong diagnostic accuracy, particularly with the Total Behavior scale, based on comparing scores with a well-established SEB measure (von der Embse et al., 2017b). Moderate predictive validity was found for the SAEBRS-SRS social, emotional, and academic subscales, while strong predictive validity was found for the SAEBRS-SRS Total Behavior scale (von der Embse et al., 2017b). Also, measurement invariance has been supported across grade clusters of lower elementary, upper elementary, middle, and high school students at the configural and metric levels, with less support at the scalar level (Kilgus et al., 2021). For the SAEBRS-TRS, measurement invariance has been established across male and female students at the configural, metric, and scalar levels (von der Embse et al., 2017a).

In addition to measurement invariance, which exists when certain groups of students are not systematically overestimated or underestimated on the item in question, it is essential to consider the influence of unique informant perspectives on an SEB screener, which may be indicative of bias (Bauer et al., 2013). Within naturalistically collected universal SEB risk assessment data, there have been systematic differences between teacher ratings, with 20.5% of the total variance related to rater differences (Smith-Millman et al., 2017). Given the existence of rater differences, multi-informant data were essential to investigate the extent to which teacher or student perspectives influenced student ratings on the SAEBRS. A relationship between informant perspectives and screening scores has previously been established for the SAEBRS (von der Embse et al., 2019). Gender and age have been examined as predictors for informant perspectives on the SAEBRS, and there is a need to extend this work to examine race as a predictor (von der Embse et al., 2019).

Universal screening is considered an equitable method of identifying students who may benefit from additional services because all students have the opportunity of being identified, compared with traditional methods that only allow for the identification of select students, such as those with ODRs. However, universal screening may not be as equitable as initially hoped due to the potential for bias to influence how students are rated. The proposed study intended to examine the SAEBRS for assessment bias prior to investigating student race as a predictor for unique teacher and student perspectives in teacher ratings and student self-report ratings of social, emotional, and academic behavioral functioning.

CHAPTER 3: METHODS

The proposed study analyzed a large sample of pre-existing SAEBRS-TRS and mySAEBRS data. This matched teacher and student self-report data was gathered through a partnership with an assessment company (Fastbridge Learning, fastbridge.org), where the universal screener is housed. Included schools had at least 80% of students screened to ensure the screener was used as a universal assessment rather than to target students. The entire sample consisted of 24,094 students across grades kindergarten through twelfth grade. However, to limit confounding variables, the current study utilized a sample of 2,948 students across seventh through ninth grade. Students from these grades were selected due to these grades being identified as more likely to belong to the high teacher perspective class than the congruent class (Kim & von der Embse, 2021). Broken down by grade, 1,306 students (44.3%) were in seventh grade, 1,366 (46.3%) were in eighth grade, and 279 (9.4%) were in ninth grade (see Table 1). In this study, race referred to the racial identity of the participants selected on the questionnaire. Participants could select from the following: American Indian or Alaska Native, Asian, Black or African American, Hispanic, Native Hawaiian or Other Pacific Islander, White, multiracial, or other. Of the sample, 891 (30%) identified as Black, and 2,057 (70%) identified as White. All participants in the sample were in grades seven, eight, or nine and racially identified as Black or White (see Table 1). Teacher demographic information was not reported.

Table 1
Student Demographics

	Number of Students	Black Students	White Students
7th Grade	1,306	369	937
8th Grade	1,366	493	873
9th Grade	279	29	247
Total	2,948	891	2,057

Note. Only demographic information for the current study sample.

Approach of the Current Study

In multi-informant data, such as the current dataset, discrepancies are likely to occur (De Los Reyes, 2011). These informant discrepancies may be attributed to unique perspectives from observing the behavior across various settings and, therefore, provide valuable information (De Los Reyes, 2013). A trifactor model can be employed to identify uniqueness and commonalities between informants. Bauer et al. (2013) suggest a trifactor model that allows for separate examination of the common factor, which represents the shared perspective between informants, the informant perspective factors, which represent the unique perspectives across informants, and the item-specific factors, which are related directly to the uniqueness of the item itself. The informant perspective factors may provide useful information on the various perspectives held; however, Bauer et al. (2013) suggest these factors could be equivalent to rater bias. Additionally, to make meaningful comparisons across racial groups, it was crucial to first establish measurement invariance across race (Kim et al., 2017).

The current study first examined measurement invariance before conducting analyses with a trifactor model. SAEBRS factor analytic research has identified a bifactor structure with each item relating to the broad Total Behavior scale, as well as to one of the following three subscales: Social Behavior (six items), Academic Behavior (six items), and Emotional Behavior (seven items) (Kilgus et al., 2015; von der Embse et al., 2016). While a bifactor model examines the factors simultaneously, a trifactor model allows for separate examination of the common factor, informant factors, and item-specific factor. Examination of each factor individually was essential for the current study to investigate the relationship between teacher and student informant perspectives and screening scores. Also, the conditional trifactor model accounted for student race as a predictor to explain the relative variance of each factor. A trifactor model is a natural extension to the bifactor model, and its use has been supported with multi-informant SEB screening data.

Previously, a trifactor model has been employed to differentiate the variance accounted for by each factor and showed support for an association between informant perspectives and screening scores (von der Embse et al., 2019). Further, demographic variables predicted common and informant perspectives differently across the subscales (e.g., gender predicted the student perspective factor for only the Emotional Behavior subscale). This highlights the value in observing total and subscale scores individually for both student and teacher informants (von der Embse et al., 2019), which was facilitated by the trifactor model in the current study. The purpose of this study was to add to the research on the equitable performance of Universal SEB screeners in facilitating decision-making. Specifically, the goal was to assess the SAEBRS for measurement invariance across race, examine race as a predictor for unique teacher and student perspectives in SAEBRS scores in terms of the informant factor, and to assess for any difference in Black and White students' scores in terms of the common factor.

Research Questions

1) Do the SAEBRS and mySAEBRS subscales display measurement invariance across Black and White students in grades seven through nine?

2) Does student race predict how students in grades seven through nine are scored on the SAEBRS and mySAEBRS subscales?
    a. Do Black students receive systematically lower scores compared to White students in terms of the teacher perspective factor?
    b. Do Black students receive systematically lower scores compared to White students in terms of the student perspective factor?
    c. Is there any difference between Black and White students in terms of common factor?

Measures

Social, Academic, Emotional Behavior Risk Screener-Teacher Rating Scale (SAEBRS-TRS; Kilgus & von der Embse, 2014). The SAEBRS-TRS is a 19-item scale that assesses student total behavior risk, as well as provides a score in each of the following behavior risk domains: social (e.g. "arguing," 6 items), academic (e.g. "preparedness for instruction," 6 items), and emotional (e.g. "sadness," 7 items). Items utilize a 4-point Likert scale (1= never, 4= almost always), with higher scores indicative of overall more positive behavioral functioning. Teachers are encouraged to think about the student's behavior within the last 30 days. Support for strong psychometric defensibility, such as reliability, validity, and diagnostic accuracy, has been found for this tool across various studies (von der Embse et al., 2019; von der Embse et al., 2016; Kilgus & Eklund, 2016; Whitley & Cuenca-Carlino, 2020; Kilgus et al., 2016b).

Social, Academic, Emotional Behavior Risk Screener-Student Rating Scale (mySAEBRS; von der Embse et al., 2017). The mySAEBRS measure is a 20-item scale that assesses social, academic, and emotional behavior risk. Students draw upon their experiences within the past month to rate each item. Students receive a total score (20 items) indicative of overall student functioning, as well as a score within each domain; social (7 items), academic (6 items), and emotional (7 items). Students rate items (e.g., "I get along with my peers.") on a Likert scale (1= never, 4= almost always). Higher scores indicate more positive behavioral functioning. This tool has internal consistency and good predictive validity (von der Embse et al., 2017a; von der Embse et al., 2017b).

Data Analysis

Measurement Invariance

To assess measurement invariance across race, a structural equation model (SEM) framework was used to run a multigroup confirmatory factor analysis (MG-CFA) using *Mplus* version 7.11 (Muthén & Muthén, 1998–2013) (Millsap & Everson, 1993). This would provide information on the usability of SAEBRS scores to identify students for social, emotional, or academic behavioral support across students of different races (i.e., Black and White students). This would support the validity of comparing SAEBRS scores across Black and White students (Borsboom, 2006). Measurement invariance analyses are necessary for examining test bias, which may impact which students are selected for services. As a prerequisite to running the MG-CFA, individual CFAs were conducted for the Black and White race groups for each subscale to ensure each model has acceptable model fit statistics. In a CFA, items composing a construct load on an unobserved or latent factor. For the SAEBRS, inferences are made about a student's social, academic, and emotional behavior based on responses to the items that make up that latent construct. To run a CFA, a minimum sample of 200 participants per group is required (MacCallum et al., 1999). The proposed study exceeded this requirement with a sample of 891 Black students and 2,057 White students.

The general CFA was evaluated with a set of model fit indices and their cut-off scores, as suggested by Hu and Bentler (1999). Model fit indices include Comparative Fit Index (CFI), which compares the observed model to a null model. If high CFIs are discovered, the observed model is a better fit than the null model. The cut-off score for CFI is .95, which means a score above this indicates a good fit. Another model fit index is the Root Mean Square Error of Approximation (RMSEA), which investigates the difference between the model that has been hypothesized and the model based on the data. Low RMSEA scores support a good model fit. RMSEA has a cut-off score of .08, and scores lower than this indicate a good model fit. Lastly, Standardized Root Mean Square Residual (SRMR) also examines the difference between the hypothesized and sample models. Low SRMR scores are better in terms of model fit. Since SRMR has a cut-off score of .08, scores lower than this indicate a good model fit.

Once the model fit was examined, an MG-CFA was run for each subscale to assess the SAEBRS for measurement invariance across race. Four steps were utilized: configural, metric, scalar, and strict invariance. The configural invariance was investigated first (Pendergast et al., 2017). Configural invariance examines the equivalence of the model across groups and is supported if the same items load onto the same factors across the groups. If configural invariance held, then metric invariance was examined next. Metric invariance tests the equivalence of factor loadings across groups. The metric model fit was compared with the configural model fit, and metric invariance was determined if the two models were not significantly different. If metric invariance held, then the next step was to establish scalar invariance, which considers the equivalence of the item intercepts. If the scalar model fit and metric model fit were not significantly different, then scalar invariance held, and the data were supported as holding measurement invariance across race. Strict invariance testing was also conducted to provide information on the variance of the items and the variance of errors across groups. The strict invariance model was compared to the scalar model, and if there were no significant differences, then strict invariance held.

When comparing models to assess for measurement invariance, the Likelihood Ratio Test (LRT) and differences in CFI and RMSEA are typically employed (Kim et al., 2017). The LRT compares the fit of two models by comparing the log-likelihood of the models (Woolf, 1957). If there is a statistically significant difference, the model with more variables or parameters is considered the better fit for the data (Woolf, 1957). A major limitation of LRT is that it is influenced by sample size. If the sample is large, it will likely show a significant Chi-squared even if the scale holds measurement invariance. (Cheung & Rensvold, 2002). Changes in CFI, RMSEA, and SRMR across racial groups were examined as well. Measurement invariance held across groups if the following criteria was met: a change in CFI across the groups was $\leq .01$, a change in RMSEA of $\leq .015$, $< .030$ change in SRMR for metric invariance and $< .015$ change in SRMR for scalar and residual invariance (Chen, 2007). These steps examined the SAEBRS for measurement invariance across race for Black and White students in grades seven through nine.

Preliminary Analysis

Descriptive statistics by race were run before employing the trifactor model. While not directly addressing the research questions, this provided information on whether teachers rated Black or White students overall higher or lower, as well as whether students provided overall higher or lower ratings for students from one racial group. This is valuable because disproportionality is defined by the higher or lower percentage of identification from one subgroup compared to the percentage of the population made up of that group (Gabel et al., 2009). The trifactor model will take this a step further by examining if this disproportionality might be connected to a unique student or teacher perspective. This will provide information on how using different raters plays a role in scores. While there is currently no established way to measure bias, Bauer et al. (2013) suggest that perspective factors are equivalent to rater bias.

Unconditional Trifactor Model

A trifactor model was arranged for each subscale (i.e., social behavior, academic behavior, and emotional behavior) with a focus on the subgroups of Black and White students. The trifactor model allowed for distinguishing multi-informant variance across the common factor, or the shared view of the target behavior, the perspective factors, or the unique perspectives of the student and teacher raters, and item-specific factor, or the variance among each shared item (Bauer et al., 2013). An unconditional trifactor model must be built and evaluated prior to adding a predictor variable.

Before addressing the research question, the trifactor model was evaluated with the same set of model fit indices as the CFA and their cut-off scores, as suggested by Hu and Bentler (1999). As a reminder, CFI compares the observed model to a null model and has a cut-off score of .95, which means a score above this indicates a good fit. RMSEA investigates the difference between the model that has been hypothesized and the model based on the data. RMSEA has a cut-off score of .08, and scores lower than this indicate a good model fit. SRMR also examines the difference between the hypothesized model and the sample model and has a cut-off score of .08, meaning scores lower than this indicate a good model fit. The magnitude of the teacher unique perspective factor, student unique perspective factor and the common factor were then observed and interpreted.

Conditional Trifactor Model

A conditional trifactor model was utilized to examine race as a predictor in terms of the informant factors and common factor. This analysis directly responded to research question two. The observed variable of student race was included as a predictor of the teacher informant factor and as a predictor of the student informant factor to examine the difference between Black and White students in terms of the informant perspective factor. This predictor explained the mean difference between Black and White student scores in terms of the unique teacher and student perspective factors. Additionally, student race was included as a predictor of the common factor to examine any difference in Black and White student scores in terms of this factor. Race was dummy coded with 0 for White students. Thus, higher positive values would indicate higher scores for Black students compared to White peers, whereas lower negative scores would indicate lower scores for Black students compared to White peers.

CHAPTER 4: RESULTS

Missing Data Analysis and Treatment

While each participating student completed the mySAEBRS on their own behavior, teachers did not complete the SAEBRS for every student. Specifically, 2,948 students completed the mySAEBRS, and 1,521 teachers completed the SAEBRS for participating students (48% missing). The missing data varied across variables but were not related to any teacher or student variables. The mechanism for the missing data required qualitative investigation and understanding of the data collection procedures. This is challenging with a secondary dataset, such as the one in the present study. However, it is reasonable to assume for this dataset that missing SAEBRS data happened simply because some teachers did not participate in the survey while their students did (von der Embse et al., 2021). Thus, Missing at Random (MAR) was assumed. Listwise deletion could treat this data by removing the missing cases. However, this would minimize the sample size, could lead to biased estimates (Donner, 1982), and ultimately results in an unnecessary loss of data. Therefore, these cases were kept in the sample and full information maximum likelihood (FIML) was used to treat this data. FIML has shown unbiased and more efficient estimates than other data treatment methods (Enders & Bandalos, 2001).

Preliminary Analysis: Descriptive Statistics

The descriptive statistics are displayed for the SAEBRS in Table 2, Table 3, and Table 4. These data show that the average scores for Black students were consistently lower, indicative of worse behavioral functioning, than for White students across each item on all three subscales of the teacher-completed SAEBRS. The descriptive statistics are displayed for the mySAEBRS in

Table 5, Table 6, and Table 7. These data show that Black students, on average, scored themselves lower than White students across each item on the Academic Behavior Subscale. This was also true for the Social Behavior Subscale aside from item number 5 (i.e., "I am respectful"), for which Black students' average self-score was higher than White peers'. The Emotional Behavior Subscale, however, shows that Black students rated themselves higher than White students for about half of the items.

Table 2

Descriptive Statistics: SAEBRS Academic Behavior Subscale

	Black Students				White Students			
Item	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis
1	2.048	0.880	-0.654	-0.304	2.551	0.757	-1.795	2.747
2	1.751	0.880	-0.381	-0.468	2.291	0.790	-0.906	0.188
3	1.912	0.854	-0.150	-0.990	2.434	0.756	-1.048	0.019
4	1.775	0.857	-0.035	-0.870	2.308	0.764	-0.795	-0.173
5	1.823	0.842	-0.118	-0.796	2.343	0.767	-0.821	-0.346
6	1.921	0.893	-0.313	-0.844	2.437	0.772	-1.192	0.571

Table 3

Descriptive Statistics: SAEBRS Social Behavior Subscale

	Black Students				White Students			
Item	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis
1	2.194	0.856	-0.809	-0.144	2.725	0.559	-2.190	5.006
2	1.894	0.836	-0.043	-1.081	2.430	0.766	-1.069	0.084
3	2.052	0.875	-0.586	-0.461	2.631	0.640	-1.791	2.977
4	2.146	0.884	-0.743	-0.331	2.571	0.680	-1.660	2.604
5	2.054	0.851	-0.381	-0.890	2.551	0.711	-1.501	1.540
6	2.430	0.742	-1.114	0.510	2.863	0.465	-4.047	17.900

Table 4

Descriptive Statistics: SAEBRS Emotional Behavior Subscale

	Black Students				White Students			
Item	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis
1	1.743	0.918	-0.024	-1.017	2.313	0.842	-0.991	0.061
2	2.204	0.834	-0.789	-0.121	2.593	0.733	-1.965	3.476
3	2.784	0.473	-2.200	4.704	2.853	0.401	-2.889	8.810
4	1.931	0.837	-0.194	-0.890	2.447	0.734	-1.177	0.744
5	2.632	0.600	-1.574	2.201	2.800	0.460	-2.469	6.798
6	2.468	0.738	-1.350	1.386	2.660	0.634	-1.976	3.739
7	2.486	0.637	-1.000	0.583	2.604	0.613	-1.464	1.818

Table 5

Descriptive Statistics: mySAEBRS Academic Behavior Subscale

	Black Students				White Students			
Item	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis
1	2.055	0.889	-0.778	-0.052	2.196	0.840	-0.973	0.483
2	1.932	0.915	-0.593	-0.420	2.075	0.795	-0.778	0.456
3	1.914	0.846	-0.224	-0.825	2.228	0.794	-0.805	-0.225
4	1.278	1.029	0.321	-1.034	1.490	0.998	0.026	-1.056
5	2.026	0.871	-0.365	-0.919	2.203	0.804	-0.663	-0.375
6	1.892	0.931	-0.353	-0.853	2.294	0.798	-0.920	0.168

Table 6

Descriptive Statistics: mySAEBRS Social Behavior Subscale

	Black Students				White Students			
Item	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis
1	1.679	0.727	-0.760	0.387	1.928	0.632	-0.817	1.849
2	1.857	0.856	-0.034	-1.060	2.032	0.888	-0.436	-0.833
3	2.239	0.734	-0.918	0.993	2.486	0.644	-1.179	1.549
4	2.185	0.914	-0.947	0.019	2.388	0.747	-1.138	0.940
5	2.210	0.820	-0.733	-0.261	2.145	0.809	-0.683	-0.102
6	2.190	0.840	-0.596	-0.741	2.370	0.801	-1.158	0.688
7	1.752	0.977	-0.427	-0.786	2.260	0.762	-0.949	-0.748

Table 7

Descriptive Statistics: mySAEBRS Emotional Behavior Subscale

	Black Students				White Students			
Item	Mean	SD	Skewness	Kurtosis	Mean	SD	Skewness	Kurtosis
1	1.935	0.930	-0.348	-0.937	1.813	0.865	-0.150	-0.812
2	1.672	0.947	-0.346	-0.766	1.857	0.871	-0.581	-0.210
3	2.141	0.821	-0.814	0.247	2.006	0.769	-0.653	0.387
4	2.109	0.912	-0.607	-0.717	2.167	0.835	-0.642	-0.469
5	2.352	0.828	-1.278	1.094	2.164	0.854	-0.904	0.260
6	1.818	0.965	-0.536	-0.631	1.822	0.903	-0.570	-0.354
7	2.203	0.815	-0.885	0.323	1.978	0.814	-0.635	0.095

Correlations between items on each subscale are presented next. Correlations represent a stronger relationship as they move closer to 1, with values of 0.70 and higher representing a strong relationship, values less than 0.70 but at least 0.30 representing a moderate relationship, and values less than 0.30 indicating a weak relationship (Dancey & Reidy, 2007). Table 8 and Table 9 present the correlation matrices for the items on the SAEBRS Academic Behavior

Subscale for each group (i.e., Black and White students) and show correlation coefficients ranged from 0.519 to 0.823 and from 0.518 to 0.806 for Black and White students, respectively. Table 10 and Table 11 display correlation matrices for the SAEBRS Social Behavior Scale and show that correlation coefficients for Black students ranged from 0.474 to 0.773 and ranged from 0.297 to 0.701 for White students. Table 12 and Table 13 display this for the SAEBRS Emotional Behavior Subscale and show correlation coefficients ranged from 0.212 to 0.590 for Black students and from 0.283 to 0.600 for White students. While these correlations were variable, they did not show statistically different values across Black and White students.

Table 8

Correlation Matrix: Black Students on SAEBRS Academic Behavior Items

	1	2	3	4	5	6
1	1.000*					
2	0.647**	1.000*				
3	0.592**	0.550**	1.000*			
4	0.527**	0.519**	0.754*	1.000*		
5	0.572**	0.564**	0.780*	0.823*	1.000*	
6	0.566**	0.549**	0.787*	0.718*	0.721*	1.000*

Note. * = $r \geq 0.70$. ** = $0.30 \leq r < 0.70$. *** = $r < .30$.

Table 9

Correlation Matrix: White Students on SAEBRS Academic Behavior Items

	1	2	3	4	5	6
1	1.000*					
2	0.575**	1.000*				
3	0.580**	0.612**	1.000*			
4	0.518**	0.592**	0.690**	1.000*		
5	0.530**	0.628**	0.739*	0.779*	1.000*	
6	0.543**	0.640**	0.806*	0.707*	0.736*	1.000*

Note. * = $r \geq 0.70$. ** = $0.30 \leq r < 0.70$. *** = $r < .30$.

Table 10

Correlation Matrix: Black Students on SAEBRS Social Behavior Items

	1	2	3	4	5	6
1	1.000*					
2	0.531**	1.000*				
3	0.725*	0.494**	1.000*			
4	0.633**	0.474**	0.711*	1.000*		
5	0.673**	0.687**	0.598**	0.523**	1.000*	
6	0.773*	0.500**	0.629**	0.620**	0.609**	1.000*

Note. * = r ≥ 0.70. ** = 0.30 ≤ r < 0.70. *** = r < .30.

Table 11

Correlation Matrix: White Students on SAEBRS Social Behavior Items

	1	2	3	4	5	6
1	1.000*					
2	0.502**	1.000*				
3	0.658**	0.406**	1.000*			
4	0.600**	0.399**	0.701*	1.000*		
5	0.591**	0.645**	0.518**	0.486**	1.000*	
6	0.529**	0.297***	0.451**	0.435**	0.353**	1.000*

Note. * = r ≥ 0.70. ** = 0.30 ≤ r < 0.70. *** = r < .30.

Table 12

Correlation Matrix: Black Students on SAEBRS Emotional Behavior Items

	1	2	3	4	5	6	7
1	1.000*						
2	0.473**	1.000*					
3	0.286***	0.357**	1.000*				
4	0.590**	0.480**	0.230***	1.000*			
5	0.301**	0.447**	0.564**	0.355**	1.000*		
6	0.365**	0.453**	0.422**	0.408**	0.576**	1.000*	
7	0.252***	0.374**	0.555**	0.212***	0.564**	0.437**	1.000*

Note. * = r ≥ 0.70. ** = 0.30 ≤ r < 0.70. *** = r < .30.

Table 13

Correlation Matrix for White Students on SAEBRS Emotional Behavior Items

	1	2	3	4	5	6	7
1	1.000*						
2	0.351**	1.000*					
3	0.319**	0.269***	1.000*				
4	0.600**	0.390**	0.284***	1.000*			
5	0.379**	0.345**	0.503**	0.444**	1.000*		
6	0.356**	0.366**	0.425**	0.510**	0.552**	1.000*	
7	0.340**	0.318**	0.570**	0.283***	0.523**	0.404**	1.000*

Note. * = r ≥ 0.70. ** = 0.30 ≤ r < 0.70. *** = r < .30.

Table 14 and Table 15 present the correlation matrices for the items on the mySAEBRS Academic Behavior Subscale for each group (i.e., Black and White students) and show correlation coefficients ranged from 0.019 to 0.416 for Black students and 0.079 to 0.562 for White students. Table 16 and Table 17 display correlation matrices for the mySAEBRS Social Behavior Scale and show correlation coefficients ranged from -0.029 to 0.416 for Black students and 0.016 to 0.422. Table 18 shows the mySAEBRS Emotional Behavior Subscale ranged from -.105 to 0.521 for Black students, and Table 19 shows this subscale ranged from 0.096 to 0.538 for White students. While correlation coefficients were also variable for the mySAEBRS subscales, they showed a statistically similar pattern across Black and White students.

Table 14

Correlation Matrix for Black Students on mySAEBRS Academic Behavior Items

	1	2	3	4	5	6
1	1.000*					
2	0.348**	1.000*				
3	0.090***	0.233***	1.000*			
4	0.019***	0.137***	0.212**	1.000*		
5	0.078***	0.127***	0.347**	0.226***	1.000*	
6	0.083***	0.217***	0.416**	0.406**	0.329**	1.000*

Note. * = r ≥ 0.70. ** = 0.30 ≤ r < 0.70. *** = r < .30.

Table 15

Correlation Matrix for White Students on mySAEBRS Academic Behavior Item

	1	2	3	4	5	6
1	1.000*					
2	0.381**	1.000*				
3	0.171***	0.264***	1.000*			
4	0.079***	0.180***	0.349**	1.000*		
5	0.102***	0.156***	0.406**	0.297***	1.000*	
6	0.125***	0.229***	0.562**	0.383**	0.364**	1.000*

Note. * = r ≥ 0.70. ** = 0.30 ≤ r < 0.70. *** = r < .30.

Table 16

Correlation Matrix for Black Students on mySAEBRS Social Behavior Items

	1	2	3	4	5	6	7
1	1.000*						
2	0.111***	1.000*					
3	0.298**	0.106***	1.000*				
4	0.230***	0.094***	0.307***	1.000*			
5	0.015***	0.293**	0.086***	-0.029***	1.000*		
6	0.192***	0.317**	0.296**	0.154***	0.256**	1.000*	
7	0.416**	0.189***	0.285**	0.288***	0.110***	0.286***	1.000*

Note. * = r ≥ 0.70. ** = 0.30 ≤ r < 0.70. *** = r < .30.

Table 17

Correlation Matrix for White Students on mySAEBRS Social Behavior Items

	1	2	3	4	5	6	7
1	1.000*						
2	0.016***	1.000*					
3	0.300**	0.122***	1.000*				
4	0.214***	0.096***	0.225***	1.000*			
5	0.127***	0.350**	0.175***	0.110***	1.000*		
6	0.169***	0.324**	0.288**	0.182***	0.422**	1.000*	
7	0.420**	0.147***	0.290**	0.253***	0.217***	0.228***	1.000*

Note. * = r ≥ 0.70. ** = 0.30 ≤ r < 0.70. *** = r < .30.

Table 18

Correlation Matrix for Black Students on mySAEBRS Emotional Behavior Items

	1	2	3	4	5	6	7
1	1.000*						
2	-0.105***	1.000*					
3	-0.035***	0.259**	1.000*				
4	0.374**	0.009***	0.096***	1.000*			
5	0.078***	0.273**	0.521**	0.287***	1.000*		
6	0.068***	0.109***	0.161***	0.136***	0.192***	1.000*	
7	0.068***	0.252**	0.432**	0.070***	0.436**	0.157***	1.000

Note. * = r ≥ 0.70. ** = 0.30 ≤ r < 0.70. *** = r < .30.

Table 19
Correlation Matrix for White Students on mySAEBRS Emotional Behavior Items

	1	2	3	4	5	6	7
1	1.000*						
2	0.096***	1.000*					
3	0.109***	0.413**	1.000*				
4	0.409**	0.292***	0.314***	1.000*			
5	0.251***	0.403**	0.499**	0.441**	1.000*		
6	0.134***	0.309***	0.278***	0.309***	0.289***	1.000*	
7	0.188***	0.362**	0.527**	0.231***	0.538**	0.232***	1.000*

Note. * = r ≥ 0.70. ** = 0.30 ≤ r < 0.70. *** = r < .30.

Research Question 1:

Do the SAEBRS and mySAEBRS subscales display measurement invariance across Black and White students in grades seven through nine?

Confirmatory Factor Analysis

Prior to testing for measurement invariance, CFA was run to examine the fit of the model of each subscale. Model fit must be confirmed before examining the model for measurement invariance. Each subscale was assessed for measurement invariance. Evaluating the subscales independently was valuable when employing the trifactor model because it could identify if demographic variables predicted common and informant perspectives differently across the subscales, such as was observed in previous research (e.g., gender predicted the student perspective factor for only the Emotional Behavior subscale) (von der Embse et al., 2019). The proposed model for each subscale of the SAEBRS and mySAEBRS is displayed in Figure 1.

Examinations of Assumptions for Analytical Approaches

Measurement models must meet several rules. First, the counting rule states that the model's degrees of freedom must be equal to or more than zero (Kline, 2015). Also, latent variables must be scaled by identifying a unit for each latent variable in the model. This process

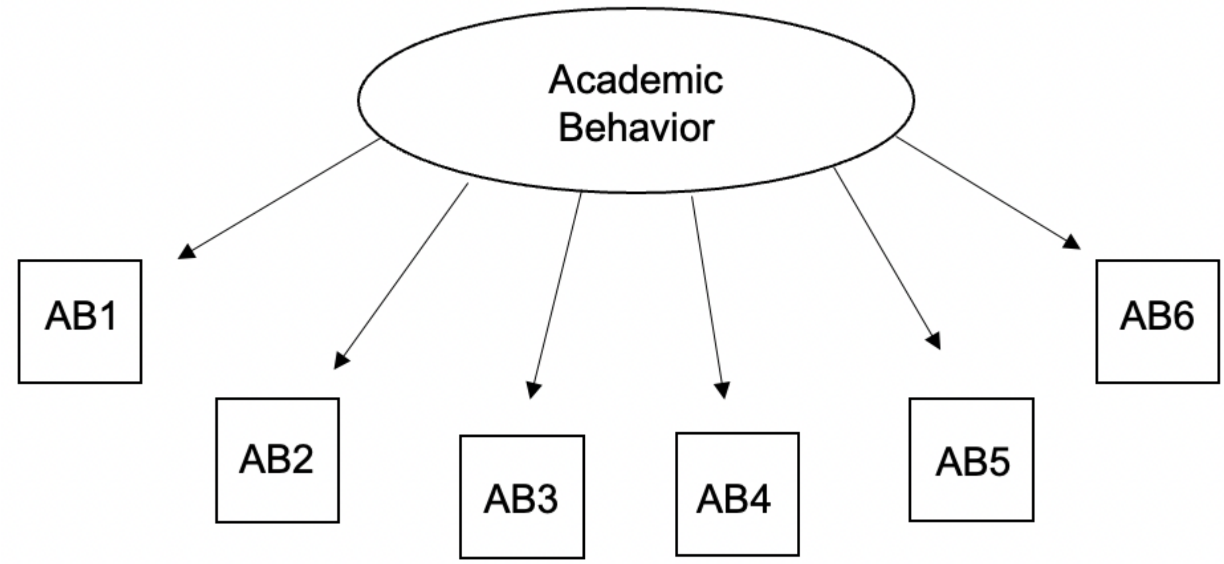

Figure 1
Academic Behavior Subscale Proposed Model

occurs automatically in *MPlus* through unit loading identification which fixes one of the factor loadings of the observed variables that load on a target factor at one. The three indicator rule states that each factor must have at least three indicators (Kline, 2015). Each subscale model has at least three indicators, thus satisfying this rule. To evaluate the model fit of each subscale to the data sample, the following lenient model fit criteria were employed: CFI $\geq$ 0.95 as good and $\geq$ 0.90 as acceptable (Bentler, 1990), RMSEA $\leq$ 0.08, and SRMR $\leq$ 0.08 (Hu & Bentler, 1999). If CFI is identified as good or acceptable, this must be in combination with one of the two absolute fit indices (i.e., RMSEA or SRMR) above their cut-off to be deemed overall good or acceptable model fit to the data (Hu & Bentler, 1999).

Model Evaluation: SAEBRS Academic Behavior Subscale

The original model fit to the total sample showed χ^2=142.572 (df=8), p=0.000, RMSEA=0.150, SRMR=0.065, and CFI=0.957, indicating poor model fit. To improve model fit, an error correlation was included for item 1 (“Interest in academic topics”) with item 2

(“Preparedness for instruction”) because if a student receives a high rating for item 1, it is reasonable to expect the student to also receive a high rating for item 2. Table 20 displays the results of the new model fit to the total sample, which shows χ^2=169.982 (*df*=8), *p*=0.000, RMSEA=0.115, SRMR=0.016, and CFI=0.977, and the model fit indices indicate that the SAEBRS Academic Behavior Subscale had good model fit to the entire sample. Since the current study focused on group comparison, a CFA was run for each group in addition to the whole sample. For Black students in the sample, χ^2 was 69.581 with a p-value of 0.0000 (df=8), which is less than .05 and indicates poor model fit in terms of exact fit. The CFI was 0.978, which is more than 0.95 and indicates good fit. The RMSEA was 0.111, which is over .08 and indicates poor model fit, while the SRMR was 0.018, which is under the cut-off score of .08 and indicates acceptable model fit. Overall, the model fit indices indicate that the SAEBRS Academic Behavior Subscale had acceptable model fit to Black students in the sample. For White students in the sample, χ^2 was 109.049 with a p-value of 0.000 (df=8), which is less than .05 and indicates poor model fit in terms of exact fit. The CFI was 0.973, which is more than 0.95 and indicates good model fit. The RMSEA was 0.119, which is over .08, indicating poor model fit, while the SRMR was 0.019, which is under the cut-off score of .08 and indicates acceptable model fit. The model fit indices overall indicate that the SAEBRS Academic Behavior Subscale had acceptable model fit to White students in the sample.

Table 20

Confirmatory Factor Analysis: SAEBRS Academic Behavior Subscale

	χ^2 (df)	CFI	RMSEA	SRMR
Total sample	169.982 (8) *p* =0.000	0.977	0.115	0.016
Black Students	69.581 (8) *p* =0.000	0.978	0.111	0.018
White Students	109.049 (8) *p* = 0.000	0.973	0.119	0.019

Model Evaluation: SAEBRS Social Behavior Subscale

Table 21 reveals the results for which the total sample was included and shows χ^2=469.345(*df*=9), *p*=0.000, RMSEA=0.183, SRMR=0.046, and CFI=0.919. The model fit indices indicate that the SAEBRS Social Behavior Subscale had acceptable model fit to the entire sample. Next, CFA by group was investigated. For Black students in the sample, χ^2 was 212.173 with a p-value of 0.000 (df=9), which indicates poor model fit in terms of exact fit. The CFI was 0.916 for Black students, which indicates acceptable model fit. The RMSEA was 0.189, which is over .08 and indicates poor model fit, while the SRMR was 0.047, indicating acceptable model fit. Overall, the model fit indices indicate that the SAEBRS Social Behavior Subscale had acceptable model fit to Black students in the sample. For White students in the sample, χ^2 was 259.799 with a p-value of 0.000 (df=9), which indicates poor model fit in terms of exact fit. The CFI was .897 for White students, which is close to acceptable model fit. The RMSEA was 0.177, which indicates poor model fit, while the SRMR was 0.055, which indicates acceptable model fit. Overall, the model fit indices indicate that the SAEBRS Social Behavior Subscale had acceptable model fit to White students in the sample.

Table 21

Confirmatory Factor Analysis: SAEBRS Social Behavior Subscale

	χ^2 (df)	CFI	RMSEA	SRMR
Total sample	469.345 (9) *p* =0.000	0.919	0.183	0.046
Black Students	212.173 (9) *p* =0.000	0.916	0.189	0.047
White Students	259.799 (9) *p* =0.000	0.897	0.177	0.055

Model Evaluation: SAEBRS Emotional Behavior Subscale

The original model fit to the total sample showed χ^2=305.672 (*df*=13), *p*=0.000, RMSEA=0.177, SRMR=0.072, and CFI=0.828, indicating the original model fit was poor. To improve model fit, an error correlation was included for Item 1 (“Sadness”) with item 4 (“Positive attitude”) because if a student receives a low rating for item 1, it is reasonable to expect the student will receive a high rating for item 4. Table 22 presents the CFA results of the new model fit to the total sample and shows χ^2=322.244 (*df*=13), *p*=0.000, RMSEA=0.125, SRMR=0.055, and CFI=0.920. The model fit indices indicate that the SAEBRS Emotional Behavior Subscale had acceptable model fit to the entire sample. When examined by group, Black students in the sample show χ^2 was 147.645 with a p-value of 0.000 (df=13), which indicates poor model fit in terms of exact fit. The CFI was acceptable at 0.914. The RMSEA was 0.128, indicating poor model fit, while the SRMR was 0.062, indicating good model fit. Overall, the model fit indices indicate that the SAEBRS Emotional Behavior subscale had acceptable model fit to Black students in the sample. For White students in the sample, χ^2 was 173.118 with a p-value of 0.000 (df=13), which indicates poor model fit in terms of exact fit. The CFI was acceptable at 0.922. The RMSEA was 0.118, which indicates poor model fit, while the SRMR was 0.048, which indicates good model fit. Taken together, the model fit indices indicate that the SAEBRS Emotional Behavior Subscale had acceptable model fit to White students in the sample.

Table 22

Confirmatory Factor Analysis: SAEBRS Emotional Behavior Subscale

	χ^2 (df)	CFI	RMSEA	SRMR
Total sample	322.244 (13) *p* =0.000	0.920	0.125	0.055
Black Students	147.645 (13) *p* =0.000	0.914	0.128	0.062
White Students	173.118 (13) *p* =0.000	0.922	0.118	0.048

Model Evaluation: mySAEBRS Academic Behavior Subscale

The original model fit to the total sample showed χ^2=42.723 (*df*=8), *p*=0.000, RMSEA=0.111, SRMR=0.084, and CFI=0.978. Consistent with the SAEBRS Academic Behavior Subscale, an error correlation was included for item 1 ("I like school") with item 2 ("I am ready for class") because if a student reports a high rating for item 1, it is reasonable to expect the student also to report a high rating for item 2. Table 23 displays the new model fit to the total sample and shows χ^2=48.631 (*df*=8), *p*=0.000, RMSEA=0.042, SRMR=0.016, and CFI=0.986. This suggests that the mySAEBRS Academic Behavior Subscale had good model fit to the entire sample. CFA by group was tested next. For Black students in the sample, χ^2 was 32.478 with a p-value of 0.000 (df=8), which implies poor model fit in terms of exact fit. The CFI for this group was 0.963, the RMSEA was 0.059, and the SRMR was 0.026, which each indicate good model fit. The model fit indices indicate that the mySAEBRS Academic Behavior Subscale had good model fit to Black students in the sample. For White students in the sample,

χ^2 was 25.033 with a p-value of 0.000 (df=8), which indicates poor model fit in terms of exact fit. The CFI for this group was 0.992, the RMSEA was 0.032, and the SRMR was 0.014, which each indicate good model fit. These data suggest the model fit indices indicate that the mySAEBRS Academic Behavior Subscale had acceptable model fit to White students in the sample.

Table 23
Confirmatory Factor Analysis: mySAEBRS Academic Behavior Subscale

	χ^2 (df)	CFI	RMSEA	SRMR
Total sample	48.631 (8) *p* =0.000	0.986	0.042	0.016
Black Students	32.478 (8) *p* =0.000	0.963	0.059	0.026
White Students	25.033 (8) *p* =0.002	0.992	0.032	0.014

Model Evaluation: mySAEBRS Social Behavior Subscale

Table 24 reveals the total sample CFA results and shows χ^2=688.111 (*df*=14), *p*=0.000, RMSEA=0.128, SRMR=0.070, and CFI=0.763. Various error correlations were attempted; however, the model fit did not improve. Therefore, the model fit indices imply that the mySAEBRS Social Behavior Subscale had poor model fit to the entire sample. CFA by group was examined next. For Black students in the sample, χ^2 was 183.440 with a p-value of 0.000 (df=14), indicating poor model fit in terms of exact fit. The CFI was 0.775 for Black students, indicating poor model fit, and the RMSEA was 0.117, which also indicates poor model fit. However, the SRMR was 0.067, which indicates good model fit. Overall, the model fit indices indicate the mySAEBRS Academic Behavior Subscale had poor model fit to Black students in the sample. For White students in the sample, χ^2 was 502.189 with a p-value of 0.000 (df=14), indicating poor model fit in terms of exact fit. The CFI was 0.747, indicating poor model fit, and the RMSEA was 0.130, which also indicates poor model fit. The SRMR was 0.070, which indicates good model fit. Overall, the model fit indices indicate the mySAEBRS Social Behavior Subscale had poor model fit to White students in the sample.

Table 24
Confirmatory Factor Analysis: mySAEBRS Social Behavior Subscale

	χ^2 (df)	CFI	RMSEA	SRMR
Total sample	688.111 (14) *p* =0.000	0.763	0.128	0.070
Black Students	183.440 (14) *p* =0.000	0.775	0.117	0.067
White Students	502.189 (14) *p* =0.000	0.747	0.130	0.070

Model Evaluation: mySAEBRS Emotional Behavior Subscale

The original model fit to the total sample showed χ^2=230.534 (*df*=13), *p*=0.000, RMSEA=0.125, SRMR=0.063, and CFI=0.844, indicating this original model had poor fit to the total sample. To improve model fit, an error correlation was included for Item 1 (“I feel sad”) with item 4 (“I am happy”). Consistent with the SAEBRS Emotional Behavior Subscale, if a student reports a low rating for item 1, it is reasonable to expect the student to report a high rating for item 4. The new model fit is displayed in Table 25 and shows χ^2=295.672 (*df*=13), *p*=0.000, RMSEA=0.086, SRMR=0.039, and CFI=0.931, suggesting that the mySAEBRS Emotional Behavior Subscale had acceptable model fit to the entire sample. CFA by group was examined next. For Black students in the sample, χ^2 was 79.605 with a p-value of 0.000 (df=13), which is less than .05 and indicates poor model fit in terms of exact fit. For this group, the CFI was 0.927, indicating acceptable model fit, the RMSEA was 0.076, indicating good model fit, and the SRMR was 0.042, also indicating good model fit. Overall, the model fit indices suggest that the mySAEBRS Emotional Behavior Subscale had good model fit to Black students in the sample. For White students in the sample, χ^2 was 249.124 with a p-value of 0.000 (df=13), which is less than .05 and indicates poor model fit in terms of exact fit. For this group, the CFI was 0.931, indicating acceptable model fit, the RMSEA was 0.094, which indicates poor model fit, and the SRMR was 0.039, which indicates good model fit. This implies that the model fit indices indicate that the mySAEBRS Emotional Behavior Subscale had acceptable model fit to White students in the sample.

Table 25
Confirmatory Factor Analysis: mySAEBRS Emotional Behavior Subscale

	χ^2 (df)	CFI	RMSEA	SRMR
Total sample	295.672 (13) *p* =0.000	0.931	0.086	0.039
Black Students	79.605 (13) *p* =0.000	0.927	0.076	0.042
White Students	249.124 (13) *p* =0.000	0.931	0.094	0.039

Interpretation of Major Parameters

Table 26 shows the total sample standardized factor loadings of each item of the SAEBRS with the associated factor. CFA factor loadings for each group (i.e., Black students and White students in the sample) showed patterns similar to the total sample, with most standardized factor loadings varying by less than 0.1 and a few varying by less than 0.2. High factor loadings support alignment to the latent factor. Specifically, an item is identified as a suitable indicator for the latent variable if the standardized factor loading is closer to one (Kline, 2015). The SAEBRS Social Behavior Subscale and Academic Behavior Subscale items showed the most alignment to the associated latent variables, with only one item from each of these subscales being below .7. The SAEBRS Emotional Subscale showed a range between .514 and .790. These lower factor loadings are still acceptable and indicate alignment to the latent variable. Table 27 shows the standardized factor loadings of each item of the mySAEBRS with the associated factor. The mySAEBRS Social Behavior Subscale ranged between .344 to .625, which are lower but are still acceptable for interpretation. Nearly all items on the mySAEBRS Academic Behavior Subscale fell between .335 and .740, meaning these are acceptable indicators. However, the item "I like school" showed a very low factor loading of .185. Nearly all items on the mySAEBRS Emotional Behavior subscale were between .362 and .746, which are also low but acceptable. However, the item “I feel sad” had a factor loading of .198, which falls below .3. Each item with a factor loading below .3 was kept, with the low factor loadings noted as a limitation.

Measurement Invariance

Measurement invariance tests if a scale measures a construct the same way across groups or time (Kim & Yoon, 2011; Meredith, 1993; Pendergast et al., 2017). The present study examined if the underlying structure of the SAEBRS and mySAEBRS subscale items is the same across racial groups (i.e., Black and White students). Both assessment tools were examined for measurement invariance before running a trifactor model to ensure measurement invariance across racial groups before breaking down the variance by common, item-specific, and unique perspective factors.

Table 26
Standardized Factor Loadings for the SAEBRS Subscales

Item	Social	Academic	Emotional
Arguing	0.882		
Cooperation with peers	0.665		
Temper outbursts	0.836		
Disruptive behavior	0.770		
Polite and socially appropriate responses toward others	0.761		
Impulsiveness	0.762		
Interest in academic topics		0.674	
Preparedness for instruction		0.708	
Production of acceptable work		0.891	
Difficulty working independently		0.866	
Distractedness		0.892	
Academic engagement		0.874	
Sadness			0.514
Fearfulness			0.574
Adaptable to change			0.664
Positive attitude			0.548
Worry			0.790
Difficulty rebounding from setbacks			0.706
Withdrawal			0.672

Table 27
Standardized Factor Loadings for the mySAEBRS Subscales

Item	Social	Academic	Emotional
I argue with others.	0.528		
I get along with my peers.	0.344		
I lose my temper.	0.540		
I disrupt class.	0.414		
I am respectful.	0.346		
Other people like me.	0.523		
I have trouble waiting my turn.	0.625		
I like school.		0.185*	
I am ready for class.		0.335	
I get good grades.		0.717	
I have trouble working alone.		0.500	
It's hard to pay attention in class.		0.522	
I participate in class.		0.740	
I feel sad.			0.198*
I feel nervous.			0.482
I like to try new things.			0.689
I am happy.			0.410
I am worried.			0.764
When something bad happens, it takes me a while to feel better.			0.362
I like being alone.			0.665

Note. * = factor loading < .30*

Measurement Invariance: SAEBRS Academic Behavior Subscale

Configural invariance. Configural invariance examines the equivalence of the model across groups and is assessed by examining how the items load onto the factors. A model holds configural invariance if the number of factors and the items that load onto those factors are the same across the groups (i.e., Black and White). The model constrained each group to have the same structure when testing for this type of invariance. The results in Table 28 show that χ^2=178.630 (*df*=16), *p*=0.000, RMSEA=0.116, SRMR=0.019, and CFI=0.975. These model fit indices indicate that there was good model fit.

Metric invariance. Metric invariance examines the equivalence of factor loadings across groups. Factor loadings are constrained to be the same across groups to test this type of invariance. The model fit indices of the metric model are then compared to the model fit indices of the configural model to determine which model has a better fit. The results in Table 28 show that χ^2=192.075 (*df*=21), *p*=0.000, RMSEA=0.103, SRMR=0.033, and CFI=0.974. This indicates that this model had good model fit. The comparisons between the two models show $\Delta\ \chi^2$ (Δdf)=13 (5), Δp=0.000, ΔRMSEA=-0.013, ΔSRMR=0.014, and ΔCFI=-0.001. The criteria employed for the differences in model fit indices are $\leq$.01 change in CFI, $\leq$.015 change in RMSEA, $<$.030 change in SRMR for metric invariance, and $<$.015 change in SRMR for scalar and strict invariance (Chen, 2007). Therefore, the differences between the two models were not significant. These data suggest that model fit was not significantly impacted by constraining the factor loadings across groups and invariance held at the metric level.

Scalar invariance. Scalar invariance examines the equivalence of the item intercepts in addition to the factor loadings. To test this type of invariance, the item intercepts are additionally constrained to be the same across groups. The model fit indices of the scalar model are then compared to the model fit indices of the metric model. The results from Table 28 show χ^2=208.575 (*df*=26), *p*=0.000, RMSEA=0.096, SRMR=0.037, and CFI=0.972, indicating there was good model fit with the more constrained model. The comparisons between the two models show $\Delta\chi^2$ (Δdf)=17 (4), Δp=0.000, ΔRMSEA=-0.007, ΔSRMR=0.004, and ΔCFI=-0.004. This suggests that the item intercepts were invariant across Black and White students at the scalar level.

Strict invariance. Strict invariance examines the residual variance of the items across groups. The residual variance is constrained to be the same across groups, and the model is compared to the scalar model. The results in Table 28 show χ^2=279.822 (*df*=32), *p*=0.000, RMSEA=0.101, SRMR=0.063, and CFI=0.962, indicating there was good model fit with this more constrained model. The comparisons between the two models show $\Delta\chi^2$ (Δdf)=71 (6), Δp=0.000, ΔRMSEA=0.005, ΔSRMR=0.026 and ΔCFI=-0.01, which indicates the item residual variance was invariant across groups, and the SAEBRS Academic Behavior Subscale was invariant across Black and White students at the strict level.

Table 28

Measurement Invariance: SAEBRS Academic Behavior Subscale

Model	χ^2 (df)	RMSEA	SRMR	CFI	$\Delta\chi^2$ (Δdf)	ΔRMSEA	ΔSRMR	ΔCFI
Model 1 Configural invariance	178.630 (16) *p* =0.000	0.116	0.019	0.975	-	-	-	-
Model 2 Metric Invariance	192.075 (21) *p* =0.000	0.103	0.033	0.974	13 (5) *p* =0.000	-0.013	0.014	-0.001
Model 3 Scalar Invariance	208.575 (26) *p* =0.000	0.096	0.037	0.972	17 (4) *p* =0.000	-0.007	0.004	-0.004
Model 4 Strict Invariance	279.822 (32) *p* =0.000	0.101	0.063	0.962	71 (6) *p* =0.000	0.005	0.026	-0.01

Measurement Invariance: SAEBRS Social Behavior Subscale

Configural invariance. The results in Table 29 show that χ^2=471.972 (*df*=18), *p*=0.000, RMSEA=0.182, SRMR=0.052, and CFI=0.906. These model fit indices indicate acceptable model fit.

Metric invariance. The results in Table 29 show that χ^2=564.874 (*df*=23), *p*=0.000, RMSEA=0.176, SRMR=0.120, and CFI=0.888, which implies that the more constrained model has acceptable model fit. The comparisons between the two models show $\Delta\chi^2$ (Δdf)=93 (5), Δp=0.000, ΔRMSEA=-0.006, ΔSRMR=0.068, and ΔCFI=-0.018. Therefore, the model fit was significantly impacted by constraining the factor loadings across groups, suggesting measurement invariance did not hold. Partial invariance was tested for next. However, model fit did not improve, and non-invariance could not be attributed to any items. Therefore, partial metric invariance did not hold, and the SAEBRS Social Behavior Subscale could not be included in the trifactor model analysis. Additionally, since measurement invariance is a hierarchical process and metric invariance did not hold, this indicates that measurement invariance also would not hold at the scalar or strict levels.

Table 29

Measurement Invariance: SAEBRS Social Behavior Subscale

Model	χ^2 (df)	RMSEA	SRMR	CFI	$\Delta\chi^2$ (Δdf)	ΔRMSEA	ΔSRMR	ΔCFI
Model 1 Configural invariance	471.972 (18) *p* =0.000	0.182	0.052	0.906	-	-	-	-
Model 2 Metric Invariance	564.874 (23) *p* =0.000	0.176	0.120	0.888	93 (5) *p* =0.000	-0.006	0.068	-0.018

Measurement Invariance: SAEBRS Emotional Behavior Subscale

Configural invariance. The results in Table 30 show that χ^2=320.763 (*df*=26), *p*=0.000, RMSEA=0.122, SRMR=0.054, and CFI=0.919, implying that there was acceptable model fit.

Metric invariance. The results in Table 30 show that χ^2=343.198 (*df*=32), *p*=0.000, RMSEA=0.113, SRMR=0.069, and CFI=0.914. This suggests that the more constrained model had good model fit. The comparisons between the two models show $\Delta\chi^2$ (Δdf)=22 (6), Δp=0.000, ΔRMSEA=-0.009, ΔSRMR=0.015, and ΔCFI=-0.005. Therefore, the differences between the two models were not significant. These data imply that model fit was not significantly impacted by constraining the factor loadings across groups, and the SAEBRS Emotional Behavior Subscale was invariant across Black and White students at the metric level.

Scalar invariance. The results in Table 30 show χ^2=497.607 (*df*=38), *p*=0.000, RMSEA=0.126, SRMR=0.086, and CFI=0.873, indicating there was poor model fit with the more constrained model. The comparisons between the two models show $\Delta\chi^2$ (Δdf)=17 (6), Δp=0.000, ΔRMSEA=0.013, ΔSRMR=0.017, and ΔCFI=-0.041. This implies that the item intercepts were not invariant across Black and White students, and measurement invariance did not hold at the scalar level. Partial invariance was tested for next. This involved relaxing item intercepts one at a time to observe the effect on the model fit and how it compared to the metric invariance model. The model fit did not improve by relaxing any single item intercept and was only significantly improved when items 1, 2, and 4 were relaxed (χ^2=350.933, *df*=35, *p*=0.000, RMSEA=0.109, SRMR=0.072, and CFI=0.913). The comparison between the partial scalar invariance model and the metric invariance model shows that the difference between the two models was not significant ($\Delta\chi^2$ (Δdf)=8 (3), Δp=0.000, ΔRMSEA=-0.004, ΔSRMR=0.003, and ΔCFI=-0.001). This suggests that the SAEBRS Emotional Behavior Subscale was measurement invariant except for items 1, 2, and 4. Because partial scalar invariance held, this supports the SAEBRS Emotional Behavior Subscale to be included in the trifactor model evaluation with the identified model adjustment.

Table 30
Measurement Invariance: SAEBRS Emotional Behavior

Model	χ^2 (df)	RMSEA	SRMR	CFI	$\Delta\chi^2$ (Δdf)	ΔRMSEA	ΔSRMR	ΔCFI
Model 1 Configural invariance	320.763 (26) *p* =0.000	0.122	0.054	0.919	-	-	-	-
Model 2 Metric Invariance	343.198 (32) *p* =0.000	0.113	0.069	0.914	22 (6) *p* =0.000	-0.009	0.015	-0.005
Model 3 Scalar Invariance	497.607 (38) *p* =0.000	0.126	0.086	0.873	17 (6) *p* =0.000	0.013	0.017	-0.041
Model 3a Partial Scalar Invariance (Items 1, 2, 4)	350.933 (35) p =0.000	0.109	0.072	0.913	8 (3) *p* =0.000	-0.004	0.003	-0.001

Measurement Invariance: mySAEBRS Academic Behavior Subscale

Configural invariance. The results in Table 31 show that χ^2=57.511 (*df*=16), *p*=0.000, RMSEA=0.042, SRMR=0.019, and CFI=0.985. These model fit indices indicate that there was a good model fit.

Metric invariance. The results in Table 31 show that χ^2=75.426 (*df*=21), *p*=0.000, RMSEA=0.042, SRMR=0.026, and CFI=0.981. This indicates that the more constrained model had good model fit. The comparisons between the two models show $\Delta\chi^2$ (Δdf)=18 (5), Δ*p*=0.000, ΔRMSEA=0, ΔSRMR=0.007, and ΔCFI=-0.004. Therefore, the differences between the two models were not significant, implying that model fit was not significantly impacted by constraining the factor loadings across groups and measurement invariance held at the metric level.

Scalar invariance. The results from Table 31 show χ^2=93.201 (*df*=26), *p*=0.000, RMSEA=0.042, SRMR=0.029, and CFI=0.976, indicating there was good model fit with the more constrained model. The comparisons between the two models show $\Delta\chi^2$ (Δdf)=18 (5), Δp=0.000, ΔRMSEA=0, ΔSRMR=0.003, and ΔCFI=-0.005. This implies that the item intercepts were invariant across groups, and the mySAEBRS Academic Behavior Subscale was invariant across Black and White students at the scalar level.

Strict invariance. The results in Table 31 show χ^2=227.528 (*df*=32), *p*=0.000, RMSEA=0.064, SRMR=0.066, and CFI=0.931, indicating there was good model fit with the more constrained model. The comparisons between the two models show $\Delta\chi^2$ (Δdf)=134 (6), Δp=0.000, ΔRMSEA=0.022, ΔSRMR=0.037 and ΔCFI=-0.045, which indicates measurement invariance did not hold at the strict level for Black and White students.

Table 31
Measurement Invariance: mySAEBRS Academic Behavior Subscale

Model	χ^2 (df)	RMSEA	SRMR	CFI	$\Delta\chi^2$ (Δdf)	ΔRMSEA	ΔSRMR	ΔCFI
Model 1 Configural invariance	57.511 (16) *p* =0.000	0.042	0.019	0.985	-	-	-	-
Model 2 Metric Invariance	75.426 (21) *p* =0.000	0.042	0.026	0.981	18 (5) *p* =0.000	0	0.007	-0.004
Model 3 Scalar Invariance	93.201 (26) *p* =0.000	0.042	0.029	0.976	18 (5) *p* =0.000	0	0.003	-0.005
Model 4 Strict Invariance	227.528 (32) *p* =0.000	0.064	0.066	0.931	134 (6) *p* =0.000	0.022	0.037	-0.045

Measurement Invariance: mySAEBRS Social Behavior Subscale

Configural invariance. The results in Table 32 show that χ^2=685.629 (*df*=28), *p*=0.000, RMSEA=0.126, SRMR=0.069, and CFI=0.755. These model fit indices indicate that there was a poor model fit. These findings align with the CFA, for which the mySAEBRS Social Behavior Subscale showed poor model fit.

Metric invariance. The results in Table 32 show that χ^2=739.624 (*df*=34), *p*=0.000, RMSEA=0.119, SRMR=0.076, and CFI=0.737, suggesting that the more constrained model had poor model fit. The comparisons between the two models show $\Delta\chi^2$ (Δdf)=54 (6), Δp=0.000, ΔRMSEA=-0.013, ΔSRMR=-0.014, and ΔCFI=-0.001. The differences between the two models were not significant, which implies that model fit was not significantly impacted by constraining the factor loadings across groups, and the mySAEBRS Social Behavior Subscale was invariant across Black and White students at the metric level.

Scalar invariance. The results from Table 32 show χ^2=854.528 (*df*=40), *p*=0.000, RMSEA=0.118, SRMR=0.089, and CFI=0.697, indicating poor fit with the more constrained model. The comparisons between the two models show $\Delta\chi^2$ (Δdf)=17 (6), Δp=0.000, ΔRMSEA=-0.001, ΔSRMR=0.013, and ΔCFI=-0.04, which implies the item intercepts were not invariant, and measurement invariance did not hold across Black and White students at the scalar level. Since invariance was not established for the SAEBRS Social Behavior Subscale, the mySAEBRS Social Behavior Subscale was also excluded from the conditional trifactor model evaluation.

Table 32
Measurement Invariance: mySAEBRS Social Behavior Subscale

Model	χ^2 (df)	RMSEA	SRMR	CFI	$\Delta\chi^2$ (Δdf)	ΔRMSEA	ΔSRMR	ΔCFI
Model 1 Configural invariance	685.629 (28) *p* =0.000	0.126	0.069	0.755	-	-	-	-
Model 2 Metric Invariance	739.624 (34) *p* =0.000	0.119	0.076	0.737	54 (6) *p* =0.000	-0.013	-0.014	-0.001
Model 3 Scalar Invariance	854.528 (40) *p* =0.000	0.118	0.089	0.697	17 (6) *p* =0.000	-0.001	0.013	-0.04

Measurement Invariance: mySAEBRS Emotional Behavior Subscale

Configural invariance. The results in Table 33 show that χ^2=328.728 (*df*=26), *p*=0.000, RMSEA=0.089, SRMR=0.040, and CFI=0.930. These model fit indices indicate that there was a good model fit.

Metric invariance. The results in Table 33 show that χ^2=372.359 (*df*=32), *p*=0.000, RMSEA=0.085, SRMR=0.053, and CFI=0.921. This indicates that the more constrained model had good model fit. The comparisons between the two models show $\Delta\chi^2$ (Δdf)=44 (6), Δp=0.000. ΔRMSEA=-0.004, ΔSRMR=0.013, and ΔCFI=-0.009. Since the differences between the two models were not significant, these data suggest that model fit was not significantly impacted by constraining the factor loadings across groups, and the mySAEBRS Emotional Behavior Subscale was invariant across Black and White students at the metric level.

Scalar invariance. The results in Table 33 show χ^2=502.620 (*df*=38), *p*=0.000, RMSEA=0.091, SRMR=0.057, and CFI=0.892, indicating there was poor model fit with the more constrained model. The comparisons between the two models show $\Delta\chi^2$ (Δdf)=130 (6), Δp=0.000, ΔRMSEA=0.006, ΔSRMR=0.004, and ΔCFI=-0.028. Therefore, the item intercepts were not invariant across Black and White students, and invariance did not hold at the scalar level. Partial scalar invariance was examined next by relaxing one item intercept at a time. The model fit was significantly improved when item 2 was relaxed (χ^2=426.86, *df*=37, *p*=0.000, RMSEA=0.085, SRMR=0.055, and CFI=0.910). The comparison between the partial scalar invariance model and the metric invariance model shows that the difference was not significant ($\Delta\chi^2$ (Δdf)=55 (1), Δp=0.000, ΔRMSEA=0, ΔSRMR=-0.002, and ΔCFI=-0.011), which shows support for partial scalar invariance. This indicates that the mySAEBRS Emotional Behavior Subscale held partial measurement invariance at the scalar level except for item 2. This supports the mySAEBRS Emotional Behavior Subscale to be included in the trifactor model evaluation with this model adjustment.

Table 33

Measurement Invariance: mySAEBRS Emotional Behavior Subscale

Model	χ^2 (df)	RMSEA	SRMR	CFI	$\Delta\chi^2$ (Δdf)	ΔRMSEA	ΔSRMR	ΔCFI
Model 1 Configural invariance	328.728 (26) p =0.000	0.089	0.040	0.930	-	-	-	-
Model 2 Metric Invariance	372.359 (32) p =0.000	0.085	0.053	0.921	44 (6) p =0.000	-0.004	0.013	-0.009
Model 3 Scalar Invariance	502.620 (38) p =0.000	0.091	0.057	0.892	130 (6) p =0.000	0.006	0.004	-0.028
Model 3a Partial Scalar Invariance (Item 2)	426.86 (37) p =0.000	0.085	0.055	0.910	55 (1) p =0.000	0	-0.002	-0.011

Research Question 2:

Does student race predict how students in grades seven through nine are scored on the SAEBRS and mySAEBRS subscales?

a. Do Black students receive systematically lower scores compared to White students in terms of the teacher perspective factor?

b. Do Black students receive systematically lower scores compared to White students in terms of the student perspective factor?

c. Is there any difference between Black and White students in terms of common factor?

Unconditional Trifactor Model

The unconditional trifactor model breaks down the multi-informant variance across the common factor, teacher unique perspective factor, student unique perspective factor, and item-specific factor. The present study investigated the relationship between these factors and each item on all SAEBRS and mySAEBRS subscales. However, the Social Behavior Subscales were not included since measurement invariance did not hold for these subscales. The model fit of the unconditional trifactor model to the multi-informant data sample was also investigated at this step.

Model Evaluation: Academic Behavior Subscales

Table 34 shows χ^2 was 512.061 with a p-value of 0.0000 (df=43), which is less than .05 and indicates poor model fit. Since the chi-square test assesses exact fit and is overly sensitive to large samples (Cheung & Rensvold, 2002), the CFI, RMSEA, and SRMR were examined to determine whether the model had good fit to the data. The CFI was 0.955, which is more than 0.95 and indicates good fit. The RMSEA was 0.061, which is less than .08 and indicates good model fit, and the SRMR was 0.057, which is under the cut-off score of .08 and indicates acceptable model fit. Overall, the model fit indices indicate that the Academic Behavior Subscales had acceptable model fit to the sample.

Table 35 shows the standardized factor loadings in terms of the common factor, student unique perspective factor, and teacher unique perspective factor. The teacher perspective factor shows a range of 0.538 to 0.737, while the common factor shows a range of 0.463 to 0.621, indicating the teacher perspective factor had a higher magnitude than the common factor. The student perspective factor, however, ranged from -0.471 to 0.375, indicating that the student perspective factor had a lower magnitude than the common factor. These data suggest that items on the SAEBRS Academic Behavior Subscale generally showed more of a relationship with the teacher perspective factor compared to the common factor, while items on the mySAEBRS Academic Behavior Subscale generally showed more of a relationship with the common factor compared to the student perspective factor.

Table 34

Unconditional Trifactor Model: Academic Behavior Subscales

χ^2 (df)	CFI	RMSEA	SRMR
512.061 (43) *p* =0.000	0.955	0.061	0.057

Table 35

Standardized Factor Loadings: Academic Behavior Subscales

	Common Factor	Student Perspective Factor	Teacher Perspective Factor
Item 1 (MS)	0.398	-0.405	
Item 2 (MS)	0.427	-0.220	
Item 3 (MS)	0.500	0.244	
Item 4 (MS)	0.449	0.267	
Item 5 (MS)	0.431	0.204	
Item 6 (MS)	0.517	0.317	
Item 1 (S)	0.398		0.466
Item 2 (S)	0.427		0.503
Item 3 (S)	0.500		0.618
Item 4 (S)	0.449		0.611
Item 5 (S)	0.431		0.629
Item 6 (S)	0.517		0.612

Model Evaluation: Emotional Behavior Subscales

Table 36 shows χ^2 was 785.999 with a p-value of 0.0000 (df=67), which is less than .05 and indicates poor model fit. The CFI was 0.912, which is more than 0.90 and indicates acceptable model fit. The RMSEA was 0.060, which is less than .08, and indicates good model fit, and the SRMR was 0.063, which is under the cut-off score of .08 and indicates good model fit. Overall, the model fit indices indicate that the Emotional Behavior Subscales had good model fit to the sample.

Table 37 shows the standardized factor loadings in terms of the common factor, student unique perspective factor, and teacher unique perspective factor. The teacher perspective factor shows a range of 0.133 to 0.249, while the common factor shows a range of -0.178 to 0.356, indicating the teacher perspective factor had a higher magnitude than the common factor. The student perspective factor shows a range of 0.249 to 0.650, indicating that the student perspective factor had a higher magnitude than the common factor. These data suggest that each item on the SAEBRS and mySAEBRS Emotional Behavior Subscales generally displayed more of a relationship with the teacher and student perspective factors, respectively, compared to the common factor.

Table 36

Unconditional Trifactor Model: Emotional Behavior Subscales

χ^2 (df)	CFI	RMSEA	SRMR
785.999 (67) *p* =0.002	0.912	0.060	0.063

Table 37

Standardized Factor Loadings: Emotional Behavior Subscales

	Common Factor	Student Perspective Factor	Teacher Perspective Factor
Item 1 (MS)	0.283	0.249	
Item 2 (MS)	0.046	0.424	
Item 3 (MS)	-0.128	0.526	
Item 4 (MS)	0.365	0.420	
Item 5 (MS)	-0.038	0.650	
Item 6 (MS)	0.060	0.340	
Item 7 (MS)	-0.178	0.537	
Item 1 (S)	0.283		0.249
Item 2 (S)	0.046		0.216
Item 3 (S)	-0.128		0.133
Item 4 (S)	0.356		0.232
Item 5 (S)	-0.038		0.181
Item 6 (S)	0.060		0.213
Item 7 (S)	-0.178		0.194

Conditional Trifactor Model

The conditional trifactor model tests a predictor variable in terms of the informant perspective, common, and item-specific factors. The present study utilized race as the predictor variable to examine the mean difference between Black and White students' scores in terms of the teacher perspective factor, student perspective factor, and the common factor. The model fit of the unconditional trifactor model to the multi-informant data sample was also investigated at this step. Race was dummy coded with 0 for White, meaning higher positive values were indicative of Black students receiving higher scores than White peers, and lower negative values were indicative of Black students receiving lower scores than White peers.

Model Evaluation: Academic Behavior Subscales

The results in Table 38 reveal that the unconstrained model had a common factor standardized regression coefficient of -0.240, a student unique perspective factor standardized regression coefficient of -0.133, and a teacher unique perspective factor standardized regression coefficient of -0.225. This means Black students overall had lower scores than White students in terms of the common factor, Black students had lower scores than White students in terms of the student perspective factor, and Black students had lower scores than White students in terms of the teacher perspective factor. To investigate whether Black students scored significantly lower than White students in terms of the teacher perspective factor compared to the student perspective factor (i.e., the significance of the difference between -0.133 and -0.225), a constrained version of this model was run. The results in Table 39 show χ^2= 558.604 (*df*=52), *p*=0.000, RMSEA=0.057, SRMR=0.062, and CFI=0.953 for the unconstrained model, while the constrained model results show χ^2= 568.871 (*df*=54), *p*=0.000, RMSEA=0.057, SRMR=0.059 and CFI=0.952. There is no significant difference between these two models, which suggests that while Black students scored lower than White students across the common, teacher perspective, and student perspective factors, these ratings were similar across all three factors.

Table 38

Standardized Regression Coefficients: Academic Behavior Subscales

	Common Factor			Student Perspective Factor			Teacher Perspective Factor		
	Estimate	S.E.	*p*	Estimate	S.E.	*p*	Estimate	S.E.	*p*
Unconstrained Model	-0.240	0.022	0.000	-0.133	0.027	0.000	-0.225	0.027	0.000
Constrained Model	-0.208	0.013	0.000	-0.208	0.013	0.000	-0.208	0.013	0.000

Note. S.E. = Standard Error

Table 39

Conditional Trifactor Model: Academic Behavior Subscales

	χ^2 (df)	CFI	RMSEA	SRMR
Unconstrained Model	558.604 (52) *p* =0.000	0.953	0.057	0.062
Constrained Model	568.871 (54) *p* =0.000	0.952	0.057	0.059

Model Evaluation: Emotional Behavior Subscales

The results in Table 40 reveal that the unconstrained model had a common factor standardized regression coefficient of -0.251, a student unique perspective factor standardized regression coefficient of 0.079, and a teacher unique perspective factor standardized regression coefficient of -0.266. This suggests that Black students overall had lower scores than White students in terms of the common factor, Black students scored higher than White students in terms of the student perspective factor, and Black students scored lower than White students in terms of the teacher perspective factor. A constrained version of this model was run to investigate whether Black students scored significantly lower than White students in terms of the teacher perspective factor compared to the student perspective factor (i.e., the significance of the difference between 0.079 and -0.266). The results in Table 41 show χ^2= 907.289 (*df*=78), *p*=0.000, RMSEA=0.060, SRMR=0.065, and CFI=0.902 for the unconstrained model, while the constrained model results show χ^2= 8580.164 (*df*=105), *p*=0.000, RMSEA=0.066, SRMR=0.076

and CFI=0.880. The constrained model had worse fit than the unconstrained model. The unconstrained model is supported, which implies that Black students received higher ratings than White students in terms of the student perspective factor, Black students overall received lower ratings than White students in terms of the common factor, and Black students received significantly lower ratings than White students in terms of the teacher perspective factor.

Table 40

Standardized Regression Coefficients: Emotional Behavior Subscales

	Common Factor			Student Perspective Factor			Teacher Perspective Factor		
	Estimate	S.E.	*p*	Estimate	S.E.	*p*	Estimate	S.E.	*p*
Unconstrained Model	-0.251	0.026	0.000	0.079	0.022	0.000	-0.266	0.026	0.000
Constrained Model	-0.088	0.017	0.000	-0.088	0.017	0.000	-0.044	0.010	0.000

Note. S.E. = Standard Error

Table 41

Conditional Trifactor Model: Emotional Behavior Subscales

	χ^2 (df)	CFI	RMSEA	SRMR
Unconstrained Model	907.289 (78) *p* =0.000	0.902	0.060	0.065
Constrained Model	8580.164 (105) *p* =0.000	0.880	0.066	0.076

CHAPTER 5: DISCUSSION

Universal SEB screening intends to be an equitable and proactive approach to the early identification of students for behavioral support or intervention. Equitable assessment and decision-making underly the effectiveness of an MTSS approach. Thus, universal screening tools should exhibit equitable performance across demographic groups, known as measurement invariance (Kim & Yoon, 2011; Meredith, 1993). If an assessment does not hold measurement invariance, scores may be systematically influenced by this test bias. This type of bias may be a function of the tool or reflective of explicit or implicit rater biases. Test bias poses a threat to the validity of conclusions drawn across these students and may lead to biased decision-making (Borsboom, 2006). Screening tools should be tested for measurement invariance across various characteristics, such as student race. Measurement invariance has been assessed for and established for the SAEBRS across grade clusters (Kilgus et al., 2021) and for the mySAEBRS across male and female students (von der Embse et al., 2017a). The current study intended to establish measurement invariance across Black and White students in grades seven through nine. Students from these grades were selected because these grades have been identified as more likely to belong to the high teacher perspective class than the congruent class (Kim & von der Embse, 2021), meaning the influence of informant perspectives was more likely to be observed for students in these grades. This would support the utility and validity of the SAEBRS to identify students for social, emotional, or academic behavioral support across seventh- through ninth-grade students within these racial groups (i.e., Black and White students).

Further, the present study examined a dataset for the influence of informant perspectives on student SEB screening scores. Rater bias is observed when systematic score differences reflect characteristics of the rater in addition to the student being rated (Hoyt, 2000). In comparison to test bias as a function of the assessment tool, rater bias is a function of the rater and their environment. The influence of the environment might include the school context or the classroom experience that day. For example, school characteristics, such as the percentage of Black students enrolled at the school, may impact which students are identified as needing behavioral support (Skiba et al., 2014). Regarding teacher characteristics, a teacher may have experienced challenging behavior from one group of students on the same day they are completing behavioral ratings for the students. This may influence the teachers' attitude towards the student and, in turn, the scores provided by the teacher.

Implicit racial bias refers to the unconscious, automatic associations related to race that are formed from personal experiences and social media and influence our understanding and behavior, although they may not align with our explicit beliefs (Staats, 2014; Russell-Brown, 2018). Implicit racial bias is one form of rater bias that may systematically influence how students are rated. This can be detrimental to students being universally screened for behavioral health support as rater bias may contribute to students being overlooked or identified without a true need for support. When multi-informant data are included in SEB screening, reports are likely to vary between informants, which may be due to unique views and observations of the behavior across settings (De Los Reyes, 2011). According to Bauer et al. (2013), the trifactor model can be used to decompose the informant ratings in terms of the common factor, or shared perspective among informants, and the perspective factors, or the unique perspectives across informants, which may include rater perspectives, biases, and unique access or observations of behaviors in various environmental contexts. The second set of research questions employed a trifactor model to observe how ratings on each subscale in terms of the common and perspective factors with race as a predictor variable. While the environmental context was unaccounted for, including a predictor may provide unique insight into processes that influences systematic differences in informant ratings. Identifying unique perspective factors with race as a predictor may provide insight into the extent to which teachers and students perceive Black students to have worse social, emotional, or behavioral functioning than White students.

Summary and Explanation of Findings

Confirmatory Factor Analysis

The first analysis conducted was a CFA to examine the proposed model fit of each subscale to the current data. The model fit indices of the SAEBRS Academic Behavior Subscale, Social Behavior Subscale, and Emotional Behavior Subscale each showed that the proposed model had good model fit. This indicates that the proposed model for each SAEBRS subscale demonstrated good fit to the data sample. Good model fit indicates further analyses can be conducted since the CFAs confirmed model fit to the data. The model fit indices of the mySAEBRS Academic Behavior Subscale and the model fit indices of the mySAEBRS Emotional Behavior Subscale each indicated the proposed model had good model fit. These data suggest that the proposed model for each of these subscales fits the current data sample, and further analyses can be conducted. However, the model fit indices of the mySAEBRS Social Behavior Subscale ($\chi 2$=688.111, *p*-value=0.000, RMSEA=0.128, SRMR=0.070, and CFI=0.763) indicated the proposed model had poor model fit. This indicates that the proposed model for this subscale did not fit the current data sample. Poor model fit to the data will be a significant limitation when conducting further analyses and may limit the interpretation of further analyses for this subscale.

Research Question 1

The first research question addressed the measurement invariance of the SAEBRS and mySAEBRS subscales across Black and White students. Establishing measurement invariance is necessary to ensure the constructs are being measured the same way across groups. There is a need for SEB screeners to be investigated for cultural responsiveness, as this is an essential component of the tool's usability. However, few SEB screeners have been assessed for this (Brann et al., 2022). An SEB assessment may be deemed culturally responsive once examined for use with marginalized populations. If the SAEBRS holds measurement invariance across Black and White students in the sample, it may be considered culturally responsive to these subpopulations. This would have to be tested with other student populations to establish measurement invariance across students from those cultural backgrounds. The SAEBRS Academic Behavior Subscale and the mySAEBRS Academic Behavior Subscale were measurement invariant across groups (i.e., Black and White students), with the SAEBRS Academic Behavior Subscale showing support at the configural, metric, scalar, and strict levels, and the mySAEBRS Subscale showing support at the configural, metric and scalar levels. This suggests that the items on the SAEBRS and the mySAEBRS Academic Behavior Subscales tested the same constructs across Black and White students in the sample. Additionally, since measurement invariance held for the Academic Behavior Subscales on both the SAEBRS and mySAEBRS, this indicates that the data from these subscales could be analyzed further with the trifactor model. The SAEBRS Social Behavior Subscale held measurement invariance across groups at the configural level, and the mySAEBRS Social Behavior Subscale held measurement invariance across groups at the configural and metric levels. Partial metric invariance was tested for the SAEBRS Social Behavior Subscale but was not supported. This suggests this subscale was not measurement invariance across groups, and the Social Behavior Subscales could not be included in the trifactor model analysis. The SAEBRS and mySAEBRS Emotional Behavior Subscales held measurement invariance across groups at the configural and metric levels. Partial scalar invariance testing supported this level of invariance among each of these subscales. This implies that the SAEBRS and mySAEBRS Emotional Behavior Subscales were measurement invariant across Black and White students in the sample and could be included in the trifactor model analysis. It was important that the current study identified each subscale as holding measurement invariance or partial measurement invariance because this indicates that any systematic ratings of students are not due to test bias. This supports moving forward with further analyses, as well as shows support for the use of the SAEBRS and mySAEBRS across Black and White students.

Research Question 2

The last research question addressed race as a predictor in terms of the informant perspective factors and common factor through employing the trifactor model. Examination of this research question involved decomposing each item to observe the standardized regression coefficient associated with the common factor, teacher unique perspective factor, and student unique perspective factor. Research on ODRs, a traditional approach to identifying students for SEB services, shows that Black students may receive more ODRs than peers, even when controlling for school characteristics such as the percentage of Black students enrolled and the percentage of students who qualified for free or reduced-price lunch (Girvan et al.,2021). Further, racial disproportionality in ODRs was significantly predicted by community-level explicit and implicit biases, but this relationship did not exist for out-of-school suspensions (Girvan et al.,2021). This indicates that decision-making at the classroom level may be influenced by teacher implicit or explicit racial biases and highlights the importance of exploring this topic with contemporary approaches to identifying students with behavioral needs (i.e., SEB screening). Therefore, the trifactor model was utilized in the current study to view the mean difference in Black and White students' scores in terms of the common factor, teacher perspective factor, and student perspective factor with race as a predictor variable.

The conditional trifactor model for the Academic Behavior Subscales showed that Black students scored lower than White students similarly across each factor. This indicates that the informant perspectives were consistent with the common factor, or shared perspectives, since the regression coefficient for the perspective factors was statistically similar to that of the common factor. The Emotional Behavior Subscales showed that Black students received higher ratings than White peers in terms of the student perspective factor, overall lower ratings than White students in terms of the common factor, and even lower ratings than White peers in terms of the teacher perspective factor. This indicates that this subscale somewhat reflected the common factor, or shared perspective, but was significantly influenced by the unique perspective factors. The current study finding is similar to that of von der Embse et al. (2019), in which the previous SAEBRS trifactor model results identified a student demographic variable as significantly predicting the unique perspective factors of the Emotional Behavior Subscale only. These findings together support the value in investigating the influence of perspective factors for each SAEBRS subscale and indicate that prioritizing one informant's rating may lead to biased decision-making.

While teacher informant perspectives of students' emotional behaviors may provide useful information that reflects the observed behavioral expression of students' internal states (De Los Reyes et al., 2013), the informant perspective is said to also include rater bias (Bauer et al., 2013). Rater bias is observed when the informant systematically inflates or lowers student ratings, which reflects something about the informant in addition to the student's behavior. Rater bias within the current informant perspectives is assumed to be implicit racial bias due to race being the predictor variable. Previous work that has suggested informant perspectives may be equivalent to rater bias involved mother and father raters who likely had similar access to the child’s behaviors (Bauer et al., 2013). In contrast, informants in the current study (i.e., student and teacher) had different access to the student’s behaviors. Thus, it is especially important to consider how perspectives may be influenced by the school context. For example, school setting such as the percentage of Black students at the school may influence the teacher’s perspective on whether the student has worse behavioral functioning than White peers. This has been supported by research in which a higher percentage of Black student enrollment was related to an increase in the likelihood that a Black student would receive an out-of-school suspension, even when controlling for student gender, the severity of the misbehavior and school-level achievement (Skiba et al., 2014). Other teacher characteristics, such as high levels of emotional exhaustion, are related to an increased use of ODRs (Eddy et al., 2020). School and teacher characteristics, such as these, would likely have a similar effect in influencing whether students are identified on an SEB screener to receive behavioral support. However, these were not able to be controlled for in the current study due to this being a secondary data analysis.

In addition to the specific school context, it is also necessary to address the systemic issues that may have a role in which students are identified for behavioral support. Critical Race

Theory posits that society's systems and institutions are fundamentally racist and have been created to meet the needs of White individuals (Sleeter, 2017). This can be observed in the community, such as segregated housing and poor healthcare for many who are racially minoritized (Bowman, 2019). The schools mirror the community, and racism can be seen in the disproportionate representation of racially minoritized students in special and gifted education and disciplinary data (Sleeter, 2017). Although teacher demographic variables were not available in the current dataset, teacher education programs continue to produce cohorts of teachers. of which about 80% are White (U.S. Department of Education, 2016). Since racism is rooted in the system, without active effort to work against this, these White teachers are positioned to uphold an oppressive system. This may have played a role in the teacher perspective significantly influencing lower scores for Black students in the sample, and therefore, the influence of this perception would have less to do with their individual bias and more to do with racism within society at large. In practice, many schools have opted to provide training to increase objectivity in SEB screening. Screening training prior to a SAEBRS screening has been related to improved acceptability and utility of the SAEBRS (von der Embse et al., 2018). Acknowledgment of and suggestions to mitigate response bias are embedded into this training, and a school may opt to include a discussion of implicit racial bias. To determine the school's need for this, schools can investigate disciplinary data for racial disproportionality since this has been linked with high levels of implicit racial bias in the community (Girvan et al., 2021).

Regarding student perception of SEB risk, it is essential to consider how one's cultural lens influences this. In the current study, Black students received higher emotional behavior scores than White students in terms of the student perspective factor. This means informant perspective influenced student scores in the opposite direction that the teacher informant perspective influenced student scores, with Black students reporting better emotional behavioral functioning than White peers. This may be related to Black students' cultural perception that mental health concerns are a sign of weakness, and they may believe it is shameful to report these types of concerns (National Alliance on Mental Illness, 2020). Schools can facilitate positive conversations about mental health to generate a shared understanding of internal behaviors that may be indicative of emotional behavioral risk. Schools can approach these conversations with an empathic understanding of historical oppression experienced by the Black community (Khalifa, 2018) and a desire to collaboratively support the students' needs. While beyond the scope of the current study, it is important that researchers investigate students' cultural perceptions of emotional behavioral risk when utilizing student self-ratings.

Further, student race is a demographic variable that has been investigated as a predictor for membership to a profile of rater agreement (von der Embse et al., 2021). Non-white students in the sample were more likely to receive congruent low teacher and student ratings. This implies that rater agreement was observed, for which the students and teachers rated non-white students lower, or as having worse behavioral functioning. The current study extends the findings of von der Embse et al. (2021) by moving beyond the shared perspectives and observing the influence of the informant perspective on student ratings. In the current study, teacher informant perspectives were observed to influence Black students' emotional behavior scores to be significantly lower than White peers, while student perspectives influenced Black students' scores to be higher than White peers'. Informants were observed to provide different information, and these findings add to the literature regarding perspective being inherently tied to how students are scored, particularly on emotional behavior. Within this context, it remains challenging to distinguish the nature of informant perspectives. The current findings call for research to continue investigating the influence of informant perspectives with other samples, as well as other SEB screeners, to better understand how various factors (e.g., school system, context, and individual characteristics) contribute to student SEB ratings.

Limitations

The current study has several limitations that must be addressed. The first limitation regards data dependency. Students in the current study sample were nested within teachers. To correct for data dependency, it is recommended to include adjusted standard errors in a single-level analysis (Stapleton et al., 2016). However, this could not be incorporated into the analysis due to the unavailability of the teacher identification variable. The second limitation concerns the missing data. The sample consisted of 2,948 mySAEBRS student self-ratings and 1,521 SAEBRS teacher ratings. Therefore, 48% of teacher ratings were missing. After investigating the mechanism of the missing data, MAR was able to be assumed. FIML was used to treat this data, and all available data could be used for the trifactor model estimation.

Third, the psychometric quality of the scale must be addressed as a limitation. Specifically, the CFA for the mySAEBRS Social Behavior Subscale showed that the subscale had an overall poor model fit to the data. Seven different error correlations were attempted; however, the model fit did not improve from these adjustments. This suggests that further analyses (i.e., measurement invariance) with the mySAEBRS Social Behavior Subscale hold little value for this sample. Also, subscales on the mySAEBRS demonstrated some very low factor loadings. In particular, the mySAEBRS Social and Emotional subscales each had an item with a factor loading below 0.2. This indicates those items may not be suitable indicators for the latent variable (i.e., social behavior or emotional behavior), and those items should be interpreted with caution.

The final limitation is that implicit racial bias may have influenced the unique perspective; however, this was not distinguished from school or other teacher characteristics that may also contribute to the teacher informant perspective. For example, the percentage of Black students enrolled at the school (Skiba et al., 2014) or the teacher’s level of emotional exhaustion (Eddy et al., 2020) may have influenced which students were selected for behavioral support. Controlling for alternative variables may indicate the potential of implicit bias to be influencing discipline disparities when there is a higher percentage of Black student enrollment. These findings emphasize the importance of ruling out school and other teacher characteristics' influence on the perception of students' behavioral functioning by collecting data on various school characteristics and examining the extent to which they predict student outcomes. Within the context of the current study, Black students received systematically lower scores than White peers on the SAEBRS Emotional Behavior Subscale. Since school and other teacher characteristics were not controlled for, this perceived need may have been related to the percentage of Black students at the school or the teacher’s level of emotional exhaustion, for example.

Future Directions

The current study focused on a target-specific characteristic (i.e., student race), and future research can focus on informant-specific characteristics, such as the influence of teacher race on student SEB scores in terms of the informant perspective factors and common factors. This will expand the current work by investigating how other variables influence student SEB screening scores. Additionally, accounting for the influence of students’ cultural perceptions of SEB risk on their own ratings was beyond the scope of the current study. Future research can include this aspect in data collection to investigate how this may impact shared perspectives with teacher ratings. Also, it will be valuable for future multi-informant data research to include a teacher identification variable to account for students nested within teachers' classrooms. With the inclusion of this information, data dependency will be able to be corrected for.

Future research needs to be conducted with other samples of students, including students from other racial backgrounds. The current study chose to focus on Black students with White students as the reference group for two reasons: 1) Black students are at significantly higher risk of experiencing exclusionary school discipline than peers (Young & Butler, 2018), and thus, there is a significant need to proactively identify these students for SEB support early on 2) at least 200 students must be in the sample to conduct a CFA (MacCallum et al., 1999). While previous research has combined non-white groups for data analysis, there is also value in examining the data of individual subgroups of students. Additionally, this work can be extended to students in other grade levels, as the current study only focused on students in grades seven through nine. Students in these grades were focused on due to the increased likelihood that student and teacher scores would be congruent, as well as to limit confounding variables. However, this is important to examine for students across all grades.

Future research may address the items with poor factor loadings to make them more representative of the latent variable. On the SAEBRS Social Behavior Subscale, the item "interest in academic topics" had an acceptable factor loading of 0.674. In contrast, its matching item on the mySAEBRS Social Behavior Subscale, "I like school" had an extremely low factor loading of 0.185. Future research is called to explore how various word selections on this mySAEBRS item might improve the factor loading.

Implications for Practice

The current research findings can be applied at the input and output of universal SEB screening. Universal SEB screening intends to be more equitable through its proactive approach to identifying students for support with early behavioral concerns indicative of future behavioral challenges. However, the current study findings speak to the potential of disproportionality and biased decision-making that could be involved within this process if steps are not taken to address this. Schools preparing to implement universal SEB screening can feel confident selecting the SAEBRS assessment tools across Black and White students as each SAEBRS and mySAEBRS subscale held some level of measurement invariance. Regarding screening more broadly, this research adds to the literature on SEB screeners that have been examined for cultural responsiveness. Developers of other SEB screeners can follow suit by conducting measurement invariance testing to examine usability among students from various backgrounds. The results also further the research on the potential impact of implicit racial bias on SEB screening scores and call for training to address this prior to implementing screening. Providing screening training to teachers before implementing the SAEBRS has improved the acceptability and utility of this tool (von der Embse et al., 2018). Response bias is often discussed when preparing teachers for universal screening, and a school may opt to embed a discussion of implicit racial bias within this. School-based practitioners can proactively investigate whether there is a need to train teachers on this particular implicit bias by examining their school's data, such as ODRs and suspensions, for racial disproportionality since this is associated with high levels of racial biases within the community (Girvan et al., 2021). If implicit racial bias training matches the school's needs, it should be provided at the classroom level compared to the administrative level because the relationship between community-level biases and disciplinary decisions no longer was observed at the administrative level (Girvan et al., 2021). Researchers can continue to explore the impacts of implicit bias training, such as the empathy intervention developed by Devine et al. (2012), which teaches strategies to reduce prejudice and has been associated with a decrease in implicit bias towards Black individuals among White pre-service teachers (Whiteford & Emerson, 2019). Additionally, school practitioners may attempt to control for other teacher characteristics that may influence their perception, such as emotional exhaustion. Schools can request informants to complete a rating of their emotional exhaustion at the time they complete the SEB screening.

At the output of SEB screening data, schools may decide to use the teacher as the informant to identify students for social and academic behavioral support or if they are interested in the behavioral expression of emotion. On the other hand, if schools are more interested in identifying students for emotional behavioral support based on the student's internal states, then using the student self-report appears to be more appropriate since the teacher informant perspective significantly influenced ratings on the Emotional Behavior Subscale in this study. When identifying students for academic behavioral support, incorporating both raters is likely to be best. A multi-informant data method has been established as best practice (Achenbach et al., 1987). Using one informant's ratings to inform decision-making can lead to students being overlooked or identified without a need for support. However, teachers and student self-raters in this study were both able to predict the common factor for the Academic Behavior Subscales, indicating that either informant may be called upon to provide student ratings for this SAEBRS and mySAEBRS subscales. Regarding universal SEB screening broadly, researchers can investigate the influence of informant perspectives on student scores to learn about the ability of each informant to predict the shared perspective.

Conclusion

The main goals of this study were to examine the SAEBRS and mySAEBRS subscales for measurement invariance across Black and White students and to employ a trifactor model with race as a predictor variable to examine the variance across the informant perspective factors and common factor. The SAEBRS and mySAEBRS held measurement invariance across Black and White Students, which supports the usability and validity of these assessment tools with these groups. This study also found that both students and teachers could predict the common factor for the Academic Behavior Subscales. However, items on the Emotional Behavior Subscales and Social Behavior Subscales primarily loaded on the informant perspective factors rather than the common factor. Overall, this study added to the usability and validity of the SAEBRS, demonstrated the potential impact of implicit racial bias on student screening scores, and proposed possible solutions to minimize an aversive impact on Black students who may experience this school phenomenon.

REFERENCES

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.

Anyon, Y., Jenson, J. M., Altschul, I., Farrar, J., McQueen, J., Greer, E., ... & Simmons, J. (2014). The persistent effect of race and the promise of alternatives to suspension in school discipline outcomes. *Children and Youth Services Review*, *44*, 379-386.

Antaramian, S. P., Huebner, E. S., Hills, K. J., & Valois, R. F. (2010). A dual-factor model of mental health: Toward a more comprehensive understanding of youth functioning. *American Journal of Orthopsychiatry*, *80*(4), 462.

Balfanz, R., & Fox, J. (2014). Sent home and put off-track: The antecedents, disproportionalities, and consequences of being suspended in the ninth grade. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, *5*(2), 13.

Basting, E. J., Naser, S., & Goncy, E. A. (2022). Assessing the factor structure and measurement invariance of the BASC-3 behavioral and emotional screening system student form across race/ethnicity and gender. *Assessment for Effective Intervention*, *48*(1), 43-51.

Bauer, D. J., Howard, A. L., Baldasaro, R. E., Curran, P. J., Hussong, A. M., Chassin, L., & Zucker, R. A. (2013). A trifactor model for integrating ratings across multiple informants. *Psychological methods*, *18*(4), 475.

Beachum, F. D., & Gullo, G. L. (2020). School leadership: Implicit bias and social justice. *Handbook on promoting social justice in education*, 429-454.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, *107*(2), 238.

Borsboom, D. (2006). When does measurement invariance matter?. *Medical care*, *44*(11), S176-S181.

Bowman, B. (2019). Programs, Practices, and Policies Affecting Diverse Children and Families. In B.H. Wasik & S. L. Odom (Eds.), Celebrating 50 years of child development research past, present, and future perspectives. 93–114.

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.

Brann, K. L., Daniels, B., Chafouleas, S. M., & DiOrio, C. A. (2022). Usability of social, emotional, and behavioral assessments in schools: A systematic review from 2009 to 2019. *School psychology review*, *51*(1), 6-24.

Briesch, A. M., Chafouleas, S. M., Nissen, K., & Long, S. (2020). A review of state-level procedural guidance for implementing multitiered systems of support for behavior (MTSS-B). *Journal of Positive Behavior Interventions*, *22*(3), 131-144.

Brown-Chidsey, R., & Steege, M. W. (2010). *Response to Intervention, Second Edition: Principles and Strategies for Effective Practice*. United States: Guilford Publications.

Burns, J. R., & Rapee, R. M. (2021). Barriers to universal mental health screening in schools: The perspective of school psychologists. *Journal of Applied School Psychology*, 1–18.

Chafouleas, S.M., Miller, F. G., Briesch, A. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2012). Usage Rating Profile – Assessment. Storrs, CT: University of Connecticut.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural equation modeling: a multidisciplinary journal*, *14*(3), 464-504.

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

Cook, C. R., Volpe, R. J., & Livanis, A. (2010). Constructing a roadmap for future universal screening research beyond academics. *Assessment for Effective Intervention*, *35*(4), 197-205.

Cullinan, D., & Epstein, M. H. (2013). Development, reliability, and construct validity of the Emotional and Behavioral Screener. *Preventing School Failure: Alternative Education for Children and Youth*, *57*(4), 223-230.

Dancey, C. P., & Reidy, J. (2007). *Statistics without maths for psychology*. Pearson education.

De Los Reyes, A., Talbott, E., Power, T. J., Michel, J. J., Cook, C. R., Racz, S. J., & Fitzpatrick, O. (2022). The Needs-to-Goals Gap: How informant discrepancies in youth mental health assessments impact service delivery. *Clinical Psychology Review*, *92*, 102114.

De Los Reyes, A. (2011). Introduction to the Special Section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, *40*(1), 1-9.

De Los Reyes, A., Thomas, S. A., Goodman, K. L., & Kundey, S. M. A. (2013). Principles Underlying the Use of multiple informants' reports. Annual Review of Clinical Psychology, 9, 123–149. https://doi.org/10.1146/annurev-clinpsy-050212-185617.

Devine, P. G., Forscher, P. S., Austin, A. J., & Cox, W. T. (2012). Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of experimental social psychology*, *48*(6), 1267-1278.

Deros, D. E., Racz, S. J., Lipton, M. F., Augenstein, T. M., Karp, J. N., Keeley, L. M., ... & De Los Reyes, A. (2018). Multi-informant assessments of adolescent social anxiety: Adding clarity by leveraging reports from unfamiliar peer confederates. *Behavior Therapy*, *49*(1), 84-98.

Dineen, J. N., Chafouleas, S. M., Briesch, A. M., McCoach, D. B., Newton, S. D., & Cintron, D. W. (2022). Exploring social, emotional, and behavioral screening approaches in US public school districts. *American Educational Research Journal*, *59*(1), 146-179.

Dixon, T. L. (2008). Network news and racial beliefs: Exploring the connection between national television news exposure and stereotypical perceptions of African Americans. *Journal of Communication, 58*(2), 321-337.

Donner, A. (1982). The relative effectiveness of procedures commonly used in multiple regression analysis for dealing with missing values. *The American Statistician*, *36*(4), 378-381.

Dowdy, E., Dever, B. V., DiStefano, C., & Chin, J. K. (2011). Screening for emotional and behavioral risk among students with limited English proficiency. *School Psychology Quarterly*, *26*(1), 14.

Eagle, J. W., Dowd-Eagle, S. E., Snyder, A., & Holtzman, E. G. (2015). Implementing a multi-tiered system of support (MTSS): Collaboration between school psychologists and administrators to promote systems-level change. *Journal of Educational and Psychological Consultation*, *25*(2-3), 160-177.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt brace Jovanovich college publishers.

Eddy, C. L., Huang, F. L., Cohen, D. R., Baker, K. M., Edwards, K. D., Herman, K. C., & Reinke, W. M. (2020). Does teacher emotional exhaustion and efficacy predict student discipline sanctions?. *School Psychology Review*, *49*(3), 239-255.

Eklund, K., & Dowdy, E. (2014). Screening for behavioral and emotional risk versus traditional school identification methods. *School Mental Health*, *6*, 40-49.

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural equation modeling*, *8*(3), 430-457.

Fabelo, T., Thompson, M. D., Plotkin, M., Carmichael, D., Marchbanks, M. P., & Booth, E. A. (2011). Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement. *New York: Council of State Governments Justice Center*.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272–299. https://doi-org.ezproxy.lib.usf.edu/10.1037/1082-989X.4.3.272.

Farkas, G., & Morgan, P. L. (2018). Risk and race in measuring special education need. *Contexts*, *17*(4), 72-74.

Fenning, P., & Jenkins, K. (2018). Racial and ethnic disparities in exclusionary school discipline: Implications for administrators leading discipline reform efforts. *Nassp Bulletin*, *102*(4), 291-302.

Gabel, S. L., Curcic, S., Powell, J. J., Khader, K., & Albee, L. (2009). Migration and ethnic group disproportionality in special education: An exploratory study. *Disability & Society, 24*(5), 625-639.

Gamm, S., Elliott, J., Halbert, J. W., Price-Baugh, R., Hall, R., Walstron, D., Uro, G., & Casserly, M. (2012). Common core state standards and diverse urban students: Using multi-tiered systems of support. Washington, DC: Council of the Great City Schools.

Gilliam, W. S., Maupin, A. N., Reyes, C. R., Accavitti, M., & Shic, F. (2016). *Do early educators' implicit biases regarding sex and race relate to behavior expectations and recommendations of preschool expulsions and suspensions?*. Research Study Brief. Yale University, Yale Child Study Center.

Girvan, E. J., Gion, C., McIntosh, K., & Smolkowski, K. (2017). The relative contribution of subjective office referrals to racial disproportionality in school discipline. *School psychology quarterly*, *32*(3), 392.

Girvan, E. J., McIntosh, K., & Santiago-Rosario, M. R. (2021). Associations between community-level racial biases, office discipline referrals, and out-of-school suspensions. *School Psychology Review*, *50*(2-3), 288-302.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of school psychology*, *45*(2), 117-135.

Gullo, G. L., & Beachum, F. D. (2020). Does implicit bias matter at the administrative level? A study of principal implicit bias and the racial discipline severity gap. *Teachers College Record*, *122*(3), 1-28.

Hartley, A. G., Zakriski, A. L., & Wright, J. C. (2011). Probing the depths of informant discrepancies: Contextual influences on divergence and convergence. *Journal of Clinical Child & Adolescent Psychology*, *40*(1), 54-66.

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.

Harro, B. (2000). The cycle of socialization. *Readings for diversity and social justice, 2*, 45-51.

Hawkins, C. (2018). Effects of Implementing the SWIS System in a PBIS School.

Houri, A. K., & Miller, F. G. (2020). A systematic review of universal screeners used to evaluate social-emotional and behavioral aspects of kindergarten readiness. *Early Education and Development*, *31*(5), 653-675.

Hoyt, W. T. (2000). Rater bias in psychological research: When is it a problem and what can we do about it?. *Psychological methods, 5*(1), 64.

Hu, L.T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Huang, F. L. (2018). Do Black students misbehave more? Investigating the differential involvement hypothesis and out-of-school suspensions. *The Journal of Educational Research, 111*(3), 284–294.

Humphrey, N., & Wigelsworth, M. (2016). Making the case for universal school-based mental health screening. *Emotional and Behavioural Difficulties, 21*(1), 22-42.

Irvin, L. K., Horner, R. H., Ingram, K., Todd, A. W., Sugai, G., Sampson, N. K., & Boland, J. B. (2006). Using office discipline referral data for decision making about student behavior in elementary and middle schools: An empirical evaluation of validity. *Journal of Positive Behavior Interventions, 8*(1), 10–23.

Jeong, S., & Lee, Y. (2019). Consequences of not conducting measurement invariance tests in cross-cultural studies: A review of current research practices and recommendations. *Advances in Developing Human Resources*, *21*(4), 466-483.

Kahneman, D. (2011). *Thinking, fast and slow*. macmillan.

Kamphaus, R. W., & Reynolds, C. R. (2007). *BASC-2 behavioral and emotional screening system manual.* Circle Pines, MN: Pearson.

Khalifa, M. A. (2018). *Culturally responsive school leadership*. Harvard Education Press.

Kilgus, S. P., & Eklund, K. R. (2016). Consideration of base rates within universal screening for behavioral and emotional risk: A novel procedural framework. *School Psychology Forum 10*(1), 120-130.

Kilgus, S. P., Eklund, K., von der Embse, N. P., Taylor, C. N., & Sims, W. A. (2016b). Psychometric defensibility of the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS) Teacher Rating Scale and multiple gating procedure within elementary and middle school samples. *Journal of School Psychology*, *58*, 21-39. https://doi.org/10.1016/j.jsp.2016.07.00.

Kilgus, S. P., Eklund, K., von der Embse, N. P., Weist, M., Barber, A. J., Kaul, M., & Dodge, S. (2021). Structural validity and reliability of social, academic, and emotional behavior risk screener–student rating scale scores: A replication study. *Assessment for Effective Intervention*, *46*(4), 259-269.

Kilgus, S. P., Sims, W. A., von der Embse, N. P., & Taylor, C. N. (2016a). Technical Adequacy of the Social, Academic, and Emotional Behavior Risk Screener in an Elementary Sample. *Assessment for Effective Intervention*, *42*(1), 46-59.

Kilgus, S. P., Sims, W. A., von der Embse, N. P., & Riley-Tillman, T. C. (2015). Confirmation of models for interpretation and use of the Social and Academic Behavior Risk Screener (SABRS). *School Psychology Quarterly*, *30*(3), 335.

Kilgus, S. P., & von der Embse, N. P. (2014). The Social, Academic, and Emotional Behavior Risk Screener—Teacher Rating Scale. Minneapolis, MN: FastBridge Learning.

Kim, E. S., Cao, C., Wang, Y., & Nguyen, D. T. (2017). Measurement invariance testing with many groups: A comparison of five approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(4), 524-544.

Kim, E. K., & Choe, D. (2022). Universal social, emotional, and behavioral strength and risk screening: Relative predictive validity for students' subjective well-being in schools. *School psychology review*, *51*(1), 40-54.

Kim, E., & von der Embse, N. (2021). Combined approach to multi-informant data using latent factors and latent classes: Trifactor mixture model. *Educational and Psychological Measurement*, *81*(4), 728-755.

Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling*, *18*(2), 212-228.

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Koblin, J. (2016). How much do we love TV? Let us count the ways. *The New York Times*, *30*.

Lambert, M. C., January, S. A. A., Cress, C. J., Epstein, M. H., & Cullinan, D. (2018). Differential item functioning across race and ethnicity for the Emotional and Behavioral Screener. *School psychology quarterly*, *33*(3), 399.

Lee, K., Quinn, P. C., & Pascalis, O. (2017). Face Race Processing and Racial Bias in Early Development: A Perceptual-Social Linkage. Current Directions in Psychological Science, 26(3), 256–262. https://doi.org/10.1177/0963721417690276.

Makol, B. A., & Polo, A. J. (2018). Parent-child endorsement discrepancies among youth at chronic-risk for depression. *Journal of abnormal child psychology*, *46*, 1077-1088.

Marsh, R. J., & Mathur, S. R. (2020). Mental Health in Schools: An Overview of Multitiered Systems of Support. *Intervention in School and Clinic*, *56*(2), 67–7.

Mash, E. J., & Hunsley, J. (2005). Evidence-based assessment of child and adolescent disorders: Issues and challenges. *Journal of clinical child and adolescent psychology*, *34*(3), 362–379.

May, S., Ard, W., Todd, A., Horner, R., Glasgow, A., Sugai, G., et al. (2000). School-wide Information System (SWIS ©). University of Oregon: Educational and Community Supports.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods*, *1*(2), 130.

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543.

McIntosh, K., Campbell, A. L., Carter, D. R., & Zumbo, B. D. (2009). Concurrent validity of office discipline referrals and cut points used in school-wide positive behavior support. *Behavioral Disorders*, *34*(2), 100–113.

McIntosh, K., Ellwood, K., McCall, L., & Girvan, E. J. (2018). Using discipline data to enhance equity in school discipline. *Intervention in school and clinic*, *53*(3), 146-152.

McIntosh, K., Frank, J. L., & Spaulding, S. A. (2010). Establishing research-based trajectories of office discipline referrals for individual students. *School Psychology Review*, *39*(3), 380-394.

McIntosh, K., & Goodman, S. (2016). Integrated Multi-Tiered Systems of Support: Blending RTI and PBIS. Guilford Publications.

McIntosh, K., Smolkowski, K., Gion, C. M., Witherspoon, L., Bastable, E., & Girvan, E. J. (2020). Awareness is not enough: A double-blind randomized controlled trial of the effects of providing discipline disproportionality data reports to school administrators. *Educational Researcher*, *49*(7), 533–537.

Mendez, L. M. R., & Knoff, H. M. (2003). Who gets suspended from school and why: A demographic analysis of schools and disciplinary infractions in a large school district. *Education and Treatment of Children*, 30-51.

Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving schools*, *12*(2), 101-113.

Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied psychological measurement*, *17*(4), 297-334.

Millsap, R. E., & Kwok, O. M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological methods*, *9*(1), 93.

Mizel, M. L., Miles, J. N., Pedersen, E. R., Tucker, J. S., Ewing, B. A., & D'Amico, E. J. (2016). To educate or to incarcerate: Factors in disproportionality in school discipline. *Children and youth services review*, *70*, 102-111.

Morris, E. W., & Perry, B. L. (2016). The punishment gap: School suspension and racial disparities in achievement. *Social Problems*, *63*(1), 68-86.

Muthén, L. K., & Muthén, B. O. (1998 –2013). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén and Muthén.

National Alliance on Mental Illness. (2020). *Mental Health in Black Communities: challenges, resources, community voices. Retrieved from https://namica.org/mental-health-challenges-in-african-american-communities/.*

National Center for Education Statistics. (2022). Public High School Graduation Rates. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved June 2, 2022, from https://nces.ed.gov/programs/coe/indicator/coi.

Naser, S., Brown, J., & Verlenden, J. (2018). The utility of universal screening to guide school-based prevention initiatives: Comparison of office discipline referrals to standardized emotional and behavioral risk screening. *Contemporary School Psychology*, *22*(4), 424-434.

Nowicki, J. M. (2018). K-12 Education: Discipline Disparities for Black Students, Boys, and Students with Disabilities. Report to Congressional Requesters. GAO-18-258. *US Government Accountability Office*.

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological science*, *26*(5), 617-624.

Pendergast, L. L., von der Embse, N., Kilgus, S. P., & Eklund, K. R. (2017). Measurement equivalence: A non-technical primer on categorical multi-group confirmatory factor analysis in school psychology. *Journal of School Psychology*, *60*, 65-82.

Predy, L., McIntosh, K., Frank, J. L., & Flitchcock, J. (2014). Utility of Number and Type of Office Discipline Referrals in Predicting Chronic Problem Behavior in Middle Schools. *School Psychology Review, 43*(4), 472–489. https://doi-org.ezproxy.lib.usf.edu/10.17105/SPR-13-0043.1.

Rabiner, D. L., Godwin, J., & Dodge, K. A. (2016). Predicting academic achievement and attainment: The contribution of early academic skills, attention difficulties, and social competence. *School Psychology Review*, *45*(2), 250–267.

Renk, K., & Phares, V. (2004). Cross-informant ratings of social competence in children and adolescents. *Clinical psychology review*, *24*(2), 239-254.

Reynolds, C. R., & Ramsay, M. C. (2003). Bias in psychological assessment: An empirical review and recommendations. In *Handbook of psychology* 1(4), 67–93). https://doi.org/10.1002/0471264385.wei1004.

Reyes, J. A., Elias, M. J., Parker, S. J., & Rosenblatt, J. L. (2013). Promoting educational equity in disadvantaged youth: The role of resilience and social-emotional learning. *Handbook of resilience in children*, 349-370.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological methods*, *17*(3), 354.

Rocque, M. (2010). Office discipline and student behavior: Does race matter?. *American Journal of Education*, *116*(4), 557–581.

Russell-Brown, K. (2018). The academic swoon over implicit racial bias: costs, benefits, and other considerations. *Du Bois Review: Social Science Research on Race*, *15*(1), 185–193.

Santiago-Rosario, M. R., Whitcomb, S. A., Pearlman, J., & McIntosh, K. (2021). Associations between teacher expectations and racial disproportionality in discipline referrals. *Journal of School Psychology, 85*, 80-93.

Scott, T. M., Gage, N., Hirn, R., & Han, H. (2019). Teacher and student race as a predictor for negative feedback during instruction. *School Psychology Quarterly: The Official Journal of the Division of School Psychology, American Psychological Association*, *34*(1), 22–31.

Shapiro, V. B., Kim, B. K., Accomazzo, S., & Roscoe, J. N. (2016). Predictors of Rater Bias in the Assessment of Social-Emotional Competence. *International Journal of Emotional Education*, *8*(2), 25-44.

Skiba, R. J., Chung, C. G., Trachok, M., Baker, T. L., Sheya, A., & Hughes, R. L. (2014). Parsing disciplinary disproportionality: Contributions of infraction, student, and school characteristics to out-of-school suspension and expulsion. *American Educational Research Journal*, *51*(4), 640-670.

Skiba, R. J., Horner, R. H., Chung, C. G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. *School psychology review*, *40*(1), 85-107.

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. doi:10.1023/A:1021320817372.

Skiba, R. J., & Rausch, M. K. (2006). Zero tolerance, suspension, and expulsion: Questions of equity and effectiveness.

Sleeter, C. E. (2017). Critical race theory and the whiteness of teacher education. *Urban Education*, *52*(2), 155-169.

Siceloff, E. R., Bradley, W. J., & Flory, K. (2017). Universal behavioral/emotional health screening in schools: overview and feasibility. *Report on emotional & behavioral disorders in youth, 17*(2), 32.

Smith-Millman, M. K., Flaspohler, P. D., Maras, M. A., Splett, J. W., Warmbold, K., Dinnen, H., & Luebbe, A. (2017). Differences between teacher reports on universal risk assessments. *Advances in School Mental Health Promotion*, *10*(4), 235-249.

Smolkowski, K., Girvan, E. J., McIntosh, K., Nese, R. N. T., & Horner, R. H. (2016). Vulnerable Decision Points for Disproportionate Office Discipline Referrals: Comparisons of Discipline for African American and White Elementary School Students. *Behavioral Disorders, 41(*4), 178–195.

Splett, J. W., Trainor, K. M., Raborn, A., Halliday-Boykins, C. A., Garzona, M. E., Dongo, M. D., & Weist, M. D. (2018). Comparison of Universal Mental Health Screening to Students Already Receiving Intervention in a Multitiered System of Support. *Behavioral Disorders, 43*(3), 344–356. https://doi.org/10.1177/0198742918761339

Staats, C. (2014). Implicit racial bias and school discipline disparities. *Exploring the connection*.

Stapleton, L. M., Yang, J. S., & Hancock, G. R. (2016). Construct meaning in multilevel settings. *Journal of Educational and Behavioral Statistics, 41*(5), 481–520. https://doi.org/10.3102/1076998616646200.

Starck, J. G., Riddle, T., Sinclair, S., & Warikoo, N. (2020). Teachers are people too: Examining the racial bias of teachers compared to other American adults. *Educational Researcher*, *49*(4), 273-284.

Sugai, G., & Horner, R. H. (2009). Responsiveness-to-intervention and school-wide positive behavior supports: Integration of multi-tiered system approaches. *Exceptionality*, *17*(4), 223–237.

Sugai, G., O'Keeffe, B. V., & Fallon, L. M. (2012). A contextual consideration of culture and school-wide positive behavior support. *Journal of Positive Behavior Interventions*, *14*(4), 197-208.

Suldo, S & Shaffer, E. (2008). Looking Beyond Psychopathology: The Dual-Factor Model of Mental Health in Youth, *School Psychology Review*, *37*:1, 52-68, DOI: 10.1080/02796015.2008.12087908

Taylor, C. N., Lovelace, R. W., Weaver, C. M., Wright Harry, S., Cato, T. A., & Ackley, M. M. (2023). Addressing Barriers to Universal Screening for Social, Emotional, and Behavioral Risk in Elementary Schools. *Perspectives on Early Childhood Psychology and Education*, *7*(1), 8.

Tilly, W. (2008). The Evolution of School Psychology to Science-Based Practice: Problem Solving and the Three-Tiered Model. Best Practices in School Psychology. *1*. 17-36.

U.S. Department of Education. (2016). The state of racial diversity in the educator workforce. Washington, DC: U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs. (2022). *43rd Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2021*. Washington, DC.

van de Vijver, F. & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *European Review of Applied Psychology*, *54*(2), 119–135. https://doi.org/10.1016/j.erap.2003.12.004.

von der Embse, N. P., Iaccarino, S., Mankin, A., Kilgus, S. P., & Magen, E. (2017a). Development and Validation of the Social, Academic, and Emotional Behavior Risk Screener–Student Rating Scale. Assessment for Effective Intervention, 42(3), 186-192. https://doi.org/10.1177/1534508416679410.

von der Embse, N. P., Kilgus, S. P., Eklund, K., Ake, E., & Levi-Neilsen, S. (2018). Training teachers to facilitate early identification of mental and behavioral health risks. *School Psychology Review, 47*(4), 372–384. https://doi.org/10.17105/SPR-2017-0094.V47-4.

von der Embse, N. P., Kilgus, S. P., Iaccarino, S., & Levi-Nielsen, S. (2017b). Screening for Student Mental Health Risk: Diagnostic Accuracy, Measurement Invariance, and Predictive Validity of the Social, Academic, and Emotional Behavior Risk Screener-Student Rating Scale (SAEBRS-SRS). *School Mental Health, 9(*3), 273–283.

von der Embse, N., Kim, E., Jenkins, A., Sanchez, A., Kilgus, S. P., & Eklund, K. (2021). Profiles of rater dis/agreement within universal screening in predicting distal outcomes. *Journal of Psychopathology and Behavioral Assessment*, 1-14.

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.

Wang, S., Rubie-Davies, C. M., & Meissel, K. (2018). A systematic review of the teacher expectation literature over the past 30 years. *Educational Research and Evaluation*, *24*(3-5), 124-179.

Whitford, D. K., & Emerson, A. M. (2019). Empathy intervention to reduce implicit bias in pre-service teachers. *Psychological reports*, *122*(2), 670-688.

Whitley, S. F., & Cuenca-Carlino, Y. (2020). Examining the technical adequacy of the social, academic, and emotional behavior risk screener. *Assessment for Effective Intervention*, *46*(1), 67-75.

Wood, B. J., & McDaniel, T. (2020). A preliminary investigation of universal mental health screening practices in schools. *Children & Youth Services Review,* 112, N.PAG. https://doi-org.ezproxy.lib.usf.edu/10.1016/j.childyouth.2020.104943.

Woolf, B. (1957). The log-likelihood ratio test (the G-test). *Annals of Human Genetics*, *21*(4), 397-409. https://doi.org/10.1111/j.1469-1809.1972.tb00293.x.

Young, J. L., & Butler, B. R. (2018). A student saved is not a dollar earned: A meta-analysis of school disparities in discipline practice toward Black children. *Taboo: The Journal of Culture and Education*, *17*(4), 6.

Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, *4*(2), 223–233.