Context is Key: Evaluating SWPBIS Tier 1 Implementation in High Schools

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE

UNIVERSITY OF MINNESOTA

BY

Greta E. Kos, M.A.

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Advisor: Dr. Faith G. Miller

June 2025

COPYRIGHT

Acknowledgements

This dissertation and the many years of education prior would not have been possible without some of the most supportive people in the world by my side. First, I would like to thank my advisor Dr. Faith Miller who balanced pushing me as an academic while caring about my wellbeing as a human with such tact over these last 5 years. She not only has modeled the type of writer and academic I want to be, but the leader and person I hope to become.

Second, I would like to thank my incredible friends and family. My parents for always accepting me as I am and raising me to be a problem solver. Their sacrifices over the years will never go unnoticed or unappreciated. My partner, Quinn Conlon, who has made numerous sacrifices to help me reach my professional goals. He has been my rock and helped me navigate all stressors and anxieties with kindness and patience in his heart. My Mankato "school friends" Dr. Abbey Riese and *almost* Drs. Courtney Sowle and Desmond Spann. From Dan Houlihan's ABA class to supporting doctoral studies from a distance, you three have been the best personal and professional friends I could ever ask for. Lastly, my "soul dog" Meira who has reminded me every day to slow down and smell the roses.

There is also no way I would have achieved this milestone without so many others, including multiple semesters writing with Dr. Lisa Aguilar and class, peer reviews from Mahasweta Bose, Elizabeth Shaver, Peter Reinke, Anna Li, and many others. I also owe a huge thank you to Dr. Mohammed Abulela who provided invaluable consultation and expertise regarding the statistical analyses in my study. Plus, a huge thank you to my therapist, Nicole, who pushes me to be the best version of myself, one session at a time.

Finally, I want to thank the other members of my committee for the time they have put in to review and provide feedback including Drs. Mollie Weeks and Kim Gibbons.

Abstract

The Schoolwide Positive Behavior and Supports (SWPBIS) framework was developed as a positive approach to school discipline, prioritizing the development and implementation of preventative systems that support school and student success. A majority of the SWPBIS research has prioritized implementation of Tier 1 (universal) structures in elementary schools, yet high schools have been implementing Tier 1 at increasing rates with varied success. This dissertation explored Tier 1 SWPBIS implementation in high schools through: (a) a narrative review of the research on Tier 1 SWPBIS implementation in high school settings and (b) a secondary analysis of a large, national, dataset of SWPBIS implementation. To this end, we analyzed the conceptual roots and history of SWPBIS, Tier 1 fidelity and core practice elements, along with an updated review of research examining the relations between high school Tier 1 implementation fidelity and office discipline referrals (ODRs). Narrative review findings highlighted the need to disaggregate data by school level, engage in more comprehensive reporting of fidelity, and establish greater experimental evidence to support deeper understanding of Tier 1 implementation and its effects at the high school level. Through the secondary analysis we found racially and ethnically marginalized (REM) students had significantly higher ODR rates than White students, and mixed results on the relationship between Tier 1 implementation fidelity and ODR rates. Overall, the narrative review demonstrated a positive relationship between Tier 1 implementation in high schools and outcomes. However, the secondary analysis revealed conflicting results regarding the association between Tier 1 fidelity components and ODRs.

Keywords: S*WPBIS, Schoolwide Positive Behavior Interventions and Supports, Tier 1, high school, implementation fidelity, Tiered Fidelity Inventory, student outcomes, ODRs, discipline*

Table of Contents

List of Tables

List of Figures

Positionality

As a White female who grew up middle-class in a rural Midwest town, my experiences have been primarily rooted in White supremacy culture. As such, it is critical that I discuss and analyze my personal connection to my research as a reflexive practice in line with my commitment to antiracism. During my years as a K-12 public education student I was consistently referred to as “difficult” due to my hyperactivity, stubbornness, and independent nature. Yet, many of my teachers also called me their favorite student and I was rarely disciplined, even when it was deserved. For years I internalized these experiences and believed something about me was special, when in reality, it was a product of my privilege. My privileges in the school context included my family name, academic and social skillsets, and most critically, my race. In my current practical experiences, I find myself building trusting relationships with the students who behave similarly to my young self. Yet, I have observed them be frequently subjected to exclusionary discipline, verbal belittlement, and social rejection because they do not hold the same privileges I did in school. Within my practice I support and advocate for every student with a whole-child approach prioritizing community and care, but I recognize my personal reach is limited. Therefore, I conduct this research with students in mind and in doing so spend substantial time in critical self-evaluation to diminish the impact my White worldview and lived experiences have on interpretations. Research on school systems requires that I unlearn my conceptualization of the way schools “should be” because of how they were for me, and instead critically evaluate the current status of education. Through conducting research and engaging in practice, I remain hopeful that someday every student will receive the same compassion and care that I was offered in school. Compassion that looks like a whole community rooting for them to reach their full potential, even when they make mistakes.

Chapter 1

Introduction

Schoolwide Positive Behavioral Interventions and Supports (SWPBIS) is an evidence-based, system-wide, preventative framework that historically emphasized four key elements to support student success: implementation *systems*, *data* based decision making, evidence-based *practices* (EBPs), and measurable *outcomes* (Center on PBIS, 2023). Through the framework, services are organized at multiple levels (e.g., school-wide, classroom, non-classroom, and individual student), and a continuum of behavior supports are provided that are universal (Tier 1), targeted (Tier 2), and intensive (Tier 3) in scope (Sugai & Horner, 2002). Of these service levels, it is argued that Tier 1 is the most influential in impacting school and student outcomes and is therefore the primary focus of initial implementation (Noltemeyer, Palmer, James, & Wiechman, 2019). That is, Tier 1 supports form the foundation for subsequent tiers. Further, early work in this area postulated that the impact of SWPBIS would likely be the greatest at the elementary level given both the student developmental level (i.e., early intervention) and systemic structures in place (i.e., the majority of time spent with one teacher; Walker et al., 1996), leading to a robust evidence base for elementary school implementation. However, high schools have been increasingly implementing SWPBIS in recent years with less of a robust evidence base (Flannery et al., 2018). As such, it is critical to evaluate Tier 1 implementation in the high school context specifically in order to expand on previous research. Through this dissertation, the high school level was thoroughly evaluated via a narrative literature review and secondary analysis of a national dataset. Questions related to implementation, student outcomes, fidelity, and equity are addressed.

SWPBIS

The development of SWPBIS originated from calls for less reactive (e.g., suspensions) and more proactive (e.g., clear, established expectations) approaches to discipline in schools given concerns related to school safety and violence (Colvin et al., 1993; Sugai, Sprague, et al., 2000; Sugai & Horner, 2020). Initially, the framework was conceptualized to prioritize adaptability as a way of meeting the ever-changing needs of students and schools. That is, adjusting priorities of the framework, which have included mental health supports, trauma-informed practices, and bullying prevention (Sugai & Horner, 2020). A recent example of these evolving needs relates to the emphasis on equity-centered SWPBIS in response to the concurrent and ongoing mass trauma associated with the COVID-19 pandemic and the national reckoning with racial injustice (Fallon et al., 2021; Sullivan et al., 2020, 2021). Subsequently, the Center on PBIS (2023) added equity as a fifth key element of the SWPBIS framework. Recent efforts to center equity have prioritized Tier 1 practices specifically, including calls to consistently collect and disaggregate data as a means to better address the needs of every student (Sullivan et al., 2021, 2024). This focus on universal practices is consistent with previous findings that Tier 1 is the most essential component of SWPBIS and should be prioritized during initial implementation to impact as many students as possible while maximizing school resources (Gage et al., 2018; Noltemeyer, Palmer, James, & Wiechman, 2019).

Tier 1

Implementation of Tier 1 requires establishing and applying a number of core practice elements at the school, classroom, and student levels (Center on PBIS, 2024; see column 5 of Table 1). These core practice elements are foundational to the implementation of the framework and are intended to set a school system up for success in managing both low-impact behavioral concerns and larger behavioral concerns. The specific actions utilized to implement these practices are adapted to fit the needs of the unique school building and students. Although the adaptability of Tier 1 is a commonly noted strength of SWPBIS as discussed, it can also make measuring fidelity of implementation more complicated than that of an individual-level intervention (Massar et al., 2019).

Fidelity and Effectiveness. Fidelity of implementation is the extent to which an intervention or framework is being implemented as designed. Numerous fidelity tools exist to measure SWPBIS fidelity of implementation. A majority of the tools evaluate Tier 1 implementation specifically to prioritize initial uptake (See Table 2 for a comparison of tools). In order to determine that Tier 1 is effective, research must provide evidence that Tier 1 was implemented as designed (i.e., with fidelity), otherwise there is a risk that inferences regarding effectiveness could be impacted by poor implementation. As such, studies have used fidelity tools to demonstrate that implementation of Tier 1 with fidelity had positive impacts on various student outcomes across grade levels. These impacts have included reductions in suspensions (Barrett et al., 2008; Grasley-Boy et al., 2019), reductions in office discipline referrals (ODRs; Algozzine et al., 2012; Bradshaw et al., 2010; Flannery et al., 2014), increased early literacy skills (Algozzine et al., 2012), and increased academic achievement (Gage et al., 2017). Of these student outcomes, ODRs have been the most widely used in research to evaluate Tier 1 effectiveness due to the feasibility of collecting ODR data and inferences that can be made regarding discipline patterns (Irvin et al., 2006; Sugai, Sprague, et al., 2000; Sugai & Horner, 2020).

A majority of SWPBIS research to date has evaluated Tier 1 effectiveness by analyzing implementation *across* a wide range of various contexts, for example: school levels, locales, school demographic characteristics, and student demographic characteristics. Although these findings have been positive, research has postulated that evaluating the effectiveness of an intervention or framework *within* specific contexts is critical (Fixsen et al., 2005). Therefore, recent efforts to evaluate effectiveness of Tier 1 in specific contexts is emerging, including evaluation of implementation differences in alternative education settings (Grasley-Boy et al., 2021), rural settings (Garbacz et al., 2022), and high schools (Estrapala et al., 2021; Flannery & McGrath Kato, 2017). Findings have been overall positive in these specific contexts, but multiple barriers and necessary adaptations have been discovered germane to specific contexts. For example, specific contextual barriers to implementation at the high school level have been identified, such as school size, organization, culture, and the student developmental level (Flannery & McGrath Kato, 2017).

Although the landscape of research evaluating SWPBIS Tier 1 implementation in high schools is relatively scarce, there has been an observed increase in the number of high schools implementing SWPBIS. As of July 2017, SWPBIS was being implemented in at least 3,367 high schools, compared to roughly 1,000 in 2010 (Flannery et al., 2018). This is both encouraging that more high schools are implementing SWPBIS and concerning due to the limited evidence-base tailored to high schools specifically. Given the uptake in high schools implementing SWPBIS and the lack of specific research evaluating the high school level, we propose research investigating high school implementation is critical. This research must evaluate outcomes, such as ODRs, Tier 1 fidelity of implementation, and the relationship between the two in order to better inform practices and allocate resources.

However, research in this realm has also been complex given the evolution of tools measuring Tier 1 fidelity over recent years. Of the various fidelity tools, the Tiered Fidelity Inventory (TFI; Algozzine et al., 2019) has been gaining momentum in research and practice since its development. The TFI allows separate assessments of all three tiers in one tool, where the tiered assessments can be used independently or together. The unique ability to assess each tier individually or comprehensively has resulted in the development of multiple measurement models for evaluation. For the full scale, these include a 3-factor model and a 10-factor model (Grasley-Boy et al., 2023; Massar et al., 2019; See Figure 1.1). The 3-factor model includes an independent score for each tier and the 10-factor model accounts for multiple subscales within each tier. For example, according to the 10-factor model Tier 1 includes three subscales: (a) teams ($n = 2$), (b) implementation ($n = 9$), and (c), evaluation ($n = 4$). Additionally, a weighted Tier 1 score was proposed by Grasley-Boy and colleagues (2023) and includes a simple weighted score of the three Tier 1 subscales. As Tier 1 is the focus of this dissertation, we will refer to these models as either a total Tier 1 score (average of all 15 Tier 1 questions), a weighted Tier 1 score (average of each Tier 1 subscale), or Tier 1 subscale scores (individual scores for each Tier 1 subscale) going forward (See Figure 1 to review the models).

TFI in High Schools

The application of the TFI Tier 1 scale in high schools is scarce. For example, in a systematic review of high school Tier 1 implementation by Estrapala and colleagues (2021), none of the 16 included studies used the TFI. However, research evaluating the psychometric properties of the full measure have included high schools in their sample. In the initial study conducted by McIntosh and colleagues (2017), 7.7% of the sample were high schools compared to 8.7% in the follow up study by Massar and colleagues (2019), and 11% in the most recent study by Grasley-Boy and colleagues (2023). These studies concluded the TFI had strong psychometric evidence *across* school buildings but with a small sample of high schools. Therefore, we aimed to explore the current use of the TFI in high schools.

Rationale

The goal of this dissertation was to review and analyze recent Tier 1 SWPBIS implementation practices in high schools, specifically related to the TFI and impacts on ODRs. This is important because high schools continue to implement SWPBIS without a robust evidence base related to adaptations, impacts on student outcomes, and best practices. Of the numerous student outcomes available, ODRs have been the most widely used in research to evaluate Tier 1 effectiveness (Irvin et al., 2006; Sugai, Sprague, et al., 2000; Sugai & Horner, 2020) and therefore will be analyzed in this dissertation to best align with previous research. Further, research thus far has strictly evaluated the TFI *across* school levels (e.g., Grasley-Boy et al., 2021; Massar et al., 2019; McIntosh et al., 2017) but not within specific levels, such as high schools. Prioritizing effectiveness at the high school level may provide further evidence for both the use of the TFI

Tier 1 scale and the impacts of Tier 1 on high school student outcomes more generally. Additionally, utilizing all three possible measurement model scores of the TFI Tier 1 scale in relationship to ODRs may help us better understand which score type(s) may be most valuable at this level and if any of the subscales may be stronger predictors of ODRs. This may advance our understanding of which specific elements of Tier 1 are the most influential for high school students. Lastly, disaggregation of data is critical to center equity within this practice. An overall goal of SWPBIS is to positively impact *every* student, but without disaggregating data and evaluating the potential for differential impacts for White students compared to racially and ethnically marginalized (REM)[1] students, we cannot determine if that goal is actually being met. In order to address these issues this dissertation (a) conducted a narrative review in order to better understand the current landscape of the SWPBIS Tier 1 literature in high schools and (b) analyzed School-Wide Information System (SWIS), from PBISApps, data to answer the following research questions related to the TFI and ODRs in high schools.

Research Questions

1. Which components of Tier 1 SWPBIS as measured by the TFI are implemented with the highest and lowest levels of fidelity in high schools using the SWIS platform?
    a. Which Tier 1 Subscale?
    b. Which Tier 1 items?
2. Overall, what is the relationship between the total Tier 1 TFI score and ODRs in high schools implementing using the SWIS platform?
    a. How are these relationships different for White students compared to REM[1] students?
    b. Does a weighted Tier 1 score impact these relationships?

[1]REM is used in this study to encompass all actively marginalized groups within Western Culture. However, we recognize this term and method unrightfully groups all of these students that face differing levels of injustice together. Although this is not best practice, this grouping allows more power in our analyses and may provide further insights on the differences between how marginalized and non-marginalized groups are impacted by SWPBIS.

3. Overall, what is the relationship between the three TFI Tier 1 subscales and ODRs in high schools using the SWIS platform?
    a. How are these relationships different for White students compared to REM[1] students?
4. Which model is a stronger predictor of ODRs in high schools implementing using the SWIS platform?

Chapter 2

Literature Review

Given the uptick in high schools implementing SWPBIS, we argue it is imperative for research to evaluate this specific level separate from other levels (e.g., elementary, middle). Although studies often include a small percentage of high schools within the larger sample, this unique context requires an individualized focus. In order to better understand the current status of implementation of Tier 1 SWPBIS in high school settings, efforts are needed to (a) evaluate the original theoretical development of SWPBIS as it applies to high schools; (b) review the current landscape of research examining implementation of SWPBIS and specific applications at the high school level; and (c) identify critical gaps within the literature. To this end, a narrative review of the research literature germane to these efforts was undertaken. First, a broad overview of the theoretical and conceptual development of SWPBIS is provided. Second, findings across all grade levels related to impacts and associations on student outcomes with Tier 1 and fidelity of implementation are discussed. Third, a review of research that examines implementation of SWPBIS in high school settings is provided. To conclude, findings were synthesized and directions for future research are discussed.

Theoretical Basis and Conceptual Development of SWPBIS

Through the 1970s to early 1990s, schools heavily relied on reactive approaches to school discipline, such as implementing zero tolerance policies, exclusionary discipline, and hiring security personnel (Sugai & Horner, 2002). These reactive approaches punished individual behaviors without focusing on sustainable change, resulting in null long-term benefits when used without a proactive support system (Gottfredson & Gottfredson, 2002). In fact, evidence has indicated that increased punishments for students with the most severe behaviors may actually result in increased intensity and frequency of the undesired behaviors (Sugai & Horner, 2006). Reliance on these reactive approaches in turn led to concerns from educators and the general public regarding school safety and violence (Colvin et al., 1993; Sugai, Sprague, et al., 2000; Sugai & Horner, 2020).

Consequently, researchers sought to identify alternative discipline practices to produce long-term change and increase school safety. Findings suggested that positive schoolwide systems centering prevention could be an effective way to approach discipline and decrease unexpected behavioral events in schools (Colvin et al., 1993). To organize these efforts, Walker and colleagues (1996) proposed a three-tiered system similar to the public health prevention model as an ideal framework to organize and deliver effective behavioral interventions in schools. The three-tiered model included primary (universal) prevention for all students, secondary prevention and intervention for students at risk of behavioral concerns, and tertiary intervention for students needing the most intensive and comprehensive support. The conceptual development of SWPBIS integrated multiple psychological theories and perspectives which prioritized manipulating the environment (e.g., adult behaviors) in order to promote prosocial behavior and decrease problem behavior (Dunlap et al., 2009). Although multiple theories were integrated, the primary theoretical grounding of SWPBIS is applied behavior analysis (ABA; Baer et al., 1968), which focuses on observable behaviors. Specifically, ABA emphasizes how observable behaviors were learned, in which contexts they occur, and what function these behaviors aim to serve (Dunlap et al., 2009; Sugai & Horner 2002, 2006). This prioritizes the *when and why* of a behavior as a key component in behavior change, promoting prosocial behaviors, and preventing future undesired behaviors.

The development of this three-tiered model led to an unprecedented decision by the United States government to reauthorize the Individuals with Disabilities Education Act (IDEA) in 1997 to include Positive Behavior Supports (PBS[2]). Turnbull and colleagues (2000) reviewed this reauthorization and argued that even though it occurred almost parallel to the development of SWPBIS and without a full understanding of the evidence for the framework, the move was warranted given the emerging research base, theoretical roots, and constitutional law (e.g., due process, least restrictive environment, equal protection). The 1997 reauthorization called for increased attention to universal prevention and intervention in schools to address the behavioral needs of most students in an efficient manner and therefore free up resources for students that need more intensive supports (Sugai, Sprague, et al., 2000). With continued development and empirical support of SWPBIS, the 2004 reauthorization of IDEA called directly for whole-school approaches and *positive behavioral interventions and supports* (e.g., § 1400[c][5][f]; Individuals with Disabilities Education Act, 2004). SWPBIS has been identified as one of the most important educational innovations for addressing behavioral problems (Bal, 2018).

Evolution of SWPBIS

In an effort to evolve with school systems, the SWPBIS framework has been adapted over time, including use of terminology and specificity of intervention types. For example, the original framework identified primary, secondary, and tertiary supports, which have since evolved in both label (i.e., “tiers”) and specificity (e.g., mode of intervention delivery). Tier 1 (primary) refers to universal or school-wide practices that all students receive, including a foundational level of behavioral expectations that are consistent across classroom and non-classroom settings. Tier 2 (secondary) provides interventions for students that need targeted support through behavior interventions, often in a group format. Finally, Tier 3 (tertiary) consists of targeted, individualized interventions for students with the most intensive needs. Although Tier 1 is universal for all students, Tiers 2 and 3 are conceptualized as a blended continuum of supports that can be added or retracted at any point in response to student behavior and progress (Sugai & Horner, 2020).

Similarly, priorities of the framework have changed over time to better meet the evolving needs of students and school systems. In 2014, various efforts were made to integrate mental health supports, trauma-informed practices, and bullying prevention into SWPBIS (Sugai & Horner, 2020). Even more recently, there has been a greater emphasis on equity-centered SWPBIS, which emerged in light of the concurrent and ongoing mass trauma associated with COVID-19 and the national reckoning with racial injustice (Fallon et al., 2021; Sullivan et al., 2020, 2021). Therefore, the Center on PBIS (2023) added equity as a fifth key element in SWPBIS expanding on the existing four (systems, data, practice, and outcomes; Horner et al., 2010). These changes and priority shifts have been primarily integrated into Tier 1 with a goal of impacting as many students as possible and decreasing the need for advanced tiered supports. These efforts have highlighted the importance of Tier 1 practices to better address the ever-changing range of student needs beyond the original goal of increasing school safety.

Tier 1 SWPBIS

Implementation of Tier 1 has been identified as the most essential component of SWPBIS to promote the behavioral success of students (Lewis et al., 1998; Noltemeyer, Palmer, James, & Wiechman, 2019). By effectively implementing Tier 1, a school system is set up for success in consistently managing typical, low-impact behavioral concerns (e.g., occasionally talking out of turn, not being seated when the bell rings) and larger behavioral concerns (e.g., violence and truancy). In the development of Tier 1, researchers established schoolwide practices critical to Tier 1 implementation targeting prevention and discipline (Colvin et al., 1993; Lewis et al., 1998; Sugai, Sprague, et al., 2000; Sugai & Horner, 2002). These practices have evolved over time but remain central to the present-day conceptualization of SWPBIS. The most recent PBIS

Implementation Blueprint (Center on PBIS, 2024) lists Tier 1 core practice elements as (a) co-designed positive Tier 1 behavior expectations are defined, taught, practiced, and reinforced schoolwide, including classroom routines consistent with schoolwide expectations; (b) continuum of procedures for encouraging expected behavior and discouraging unexpected or unwanted behavior; (c) procedures for connecting and encouraging school-family partnerships; (d) illustrations of low intensity supports threaded throughout the tiers (e.g., behavior support plans and pre-correction); (e) documentation of relation between social validity and implementation fidelity along with how screening predicts student outcomes (see Table 1 for a summary of Tier 1 core elements over time). Researchers have argued that Tier 1 is the most essential component of SWPBIS and should be prioritized during initial implementation in order to maximize school resources (Gage et al., 2018; Noltemeyer, Palmer, James, & Wiechman, 2019).

Effectiveness of Tier 1

Scientific findings from the last 20+ years indicate that SWPBIS is an EBP, with Tier 1 being the most essential for supporting the needs of every student (Filter et al., 2022). Specifically, experimental research has demonstrated significant impacts on various student outcomes when Tier 1 core practice elements are implemented as designed, including: reductions in suspensions (Barrett et al., 2008; Grasley-Boy et al., 2019), fewer office discipline referrals (ODRs; Algozzine et al., 2012; Bradshaw et al., 2010; Flannery et al., 2014), increased early literacy skills (Algozzine et al., 2012), and increased academic achievement (Gage et al., 2017). In an effort to synthesize current findings, Lee and Gage (2020) conducted a systematic review and meta-analysis of Tier 1 implementation and evaluated effectiveness across grade levels. This study focused on Tier 1 due to the small sample of studies including Tiers 2 and/or 3 in the search. The review included 32 experimental group design studies with a total of 6 randomized control trials (RCT) and 26 quasi-experimental designs (QED). Of these studies, 20 were peer-reviewed published articles and 12 were dissertations. The review included 16 elementary schools, 3 middle schools, 2 high schools and 11 mixed level schools. The percent of high schools within the 11 mixed level schools ranged from 9-18%. Inter-rater agreement was conducted at all phases of coding with 98-100% agreement across the 10% of abstracts screened, 50% of full texts screened, 50% of coded studies, and 30% of quality coding. In total, 19 studies met quality standards on What Works Clearinghouse Group Design Standards 4.0 (2019) and 24 met standards through the Council for Exceptional Children quality indicators (CEC, 2014; Cook et al., 2015). On average, schools required 3.4 years to fully implement Tier 1 practices. Results indicated that schools implementing Tier 1 had statistically significant improvements in behavioral outcomes (b = -0.26) such as exclusionary discipline, corporal punishment, and broad "problem behavior" with large heterogeneity of effects indicated with an I^2 of 77%. In addition, authors reported a statistically significant improvement in academics (b = 0.11) and organizational outcomes (b = 0.37), although the latter included only four studies. Conventional interpretive guidelines might suggest these are relatively small effects (e.g., Cohen et al., 1996).

However, we consider these effects *educationally meaningful* (see Kraft, 2020) due to the detrimental impact of exclusionary discipline practices on current and future student outcomes (e.g., Skiba et al., 2014). Kraft (2020) argued that school-based research can be more difficult to achieve traditionally large effect sizes due to the size of the school, variability of outcomes, and overall complexity. Moreover, SWPBIS is a complex framework with outcomes that do not directly align with the framework, meaning traditionally *small* effects may be deemed *educationally meaningful*. Further, the heterogeneity analysis indicated considerable heterogeneity of effects in relationship to behavioral outcomes, yet researchers did not find any significant effects of moderators tested (publication type, study design, outcome type). This unexplained heterogeneity supports the need to continue evaluating what moderators could be accounting for this level of heterogeneity. We hypothesize that school level could be an important moderator in this relationship and implementation.

Implementation Science: Fidelity, Core Practice Elements, and Context

Transferring EBPs to routine practice is a challenging endeavor across multiple service sectors, including schools. These challenges have led to a multi-disciplinary approach to study this phenomenon. Implementation science has been defined as "the scientific study of methods to promote the systematic uptake of research findings and other EBPs into routine practice, and, hence, to improve the quality and effectiveness of health services." (Eccles & Mittman, 2006, p. 1). This approach aims to identify barriers and facilitators to implementation success such as contextual variables including the target population and broad systems level components. The most frequently examined construct within implementation science is fidelity of implementation[2], which accounts for the extent to which core features of a program, intervention, or system are implemented as intended (Massar et al., 2019; Molloy et al., 2013).

SWPBIS Fidelity

Fidelity is a critical piece of SWPBIS implementation and has received a generous amount of attention as demonstrated by the development of numerous tools to measure fidelity and reliance on fidelity scores both in research and practice. Measuring fidelity of implementation is critical, but measuring fidelity in systems-level approaches like SWPBIS can be more complicated when compared to individual-level interventions (Massar et al., 2019). This is because SWPBIS is a framework of core practice elements situated within multiple systems rather than a manualized or scripted intervention targeted at a specific population. Yet, monitoring SWPBIS fidelity has helped schools implement data-based decision making, identify areas for improvement, establish next steps for continued success, and improve sustainability (McIntosh, Mercer, et al., 2018; Mercer et al., 2017). Indeed, without attending to the

[2] Fidelity of implementation has many names between and within fields including implementation fidelity, treatment integrity, treatment fidelity, and others (Breitenstein et al., 2010).

implementation process and fidelity in natural settings (e.g., schools) it is unlikely that outcomes will improve (Fixsen et al., 2010). Most tools assessing SWPBIS fidelity of implementation focus on Tier 1. In fact, it is commonly reported in the research literature that "implementing with fidelity" is achieved if Tier 1 scores are above a certain criterion set by the measure. Although there are tools that measure fidelity for Tiers 2 and 3, Tier 1 remains the most common focus of both research and practice.

Tier 1 Fidelity Measurement Tools. Numerous fidelity tools can be used to measure the extent to which a school is implementing Tier 1 core practice elements (See Table 2). This includes the School-Wide Evaluation Tool (SET; Sugai, Lewis-Palmer, et al., 2001), which was developed in 2001 and contains seven key practices for successful implementation (Horner et al., 2004). The SET has evidence for both validity and reliability of scores, in terms of construct validity, internal consistency, and test-retest reliability (Horner et al., 2004; Vincent & Tobin, 2011). A unique feature of the SET is the 80/80 criterion that requires an 80% score on both the "total score" and the "behavioral expectations taught" subscales to be considered at full implementation. Horner and colleagues (2004) suggested the 80/80 criterion based on preliminary observations of 44 schools in three states that indicated students were not likely to change their behaviors until they were taught the expected behaviors.

One of the more recent tools to measure fidelity of SWPBIS is the SWPBIS Tiered Fidelity Inventory (TFI; Algozzine et al., 2019). Unlike other fidelity measures, this tool has separate assessments of all three tiers in one instrument. These assessments can be used in isolation (e.g., individual tiers) or averaged together for a comprehensive SWPBIS fidelity score. Given this tool is unique in this way, multiple factor models and suggestions for scoring have been proposed. Evidence indicates that the TFI can be used to measure overall implementation, implementation by tier, and implementation by subscale. Research indicates that a school is implementing with fidelity if the Tier 1 questions average to over 70%. Validity and reliability evidence is present for the TFI in terms of content validity, factor structure, convergent validity, and internal consistency (Grasley-Boy et al., 2023; Massar et al., 2019). The TFI is the basis for the subsequent study within this dissertation and will be further analyzed in the next section.

Other tools include the Benchmarks of Quality (BoQ; Kincaid et al., 2010), the Team Implementation Checklist (TIC; Sugai, Horner, et al., 2001), and, the PBIS Self-Assessment Survey (SAS; Sugai, Horner, et al., 2000). The BoQ measures Tier 1 fidelity with a focus on action planning for future implementation. The BoQ has evidence to support reliability and validity of scores in terms of internal consistency, test-retest reliability, interrater reliability, and concurrent validity (Cohen et al., 2007). The TIC measures fidelity of Tier 1 but can be administered annually or quarterly for ongoing assessment. The TIC has been found to have strong internal consistency and concurrent validity evidence (Tobin et al., 2012). Lastly, the SAS is an annual assessment of fidelity across all tiers. Validity and reliability evidence have been reported in terms of internal consistency, construct validity, and criterion validity (Solomon et al., 2015).

Given the numerous tools to measure Tier 1 fidelity, Mercer and colleagues (2017) conducted a review to compare Tier 1 core practice element scores of all five tools (See Table 2 for more information on each tool). Results suggested moderate convergent validity among assessments (r = .59-.71, p < .001) excluding two comparisons of measures with the smallest sample sizes (SET and TFI, r = .92, p < .001, n = 36; TIC and TFI, r = .96, p < .001, n = 119). Additionally, there were no statistically significant mean differences between the TFI and the BoQ, SAS or TIC (d = 0.04-0.10). Authors noted that the SET had higher (i.e., inflated) scores on average compared to the other tools (d = 0.39-0.72), but also the most stringent *full* implementation criteria (80/80 criterion). The 80/80 criterion led to a similar proportion of schools that reached full implementation across measures, despite the SET yielding higher scores on average. These findings support relative consistency across the numerous tools measuring Tier 1 fidelity.

Although there are strengths and limitations to each fidelity tool, the SET, BoQ, and TFI appear to be the most heavily used in the research literature. Overall, Mercer and colleagues (2017) identified the SET to be the most common tool used and considered to be the most objective given the requirement of an external evaluator. In comparison, the TFI and BoQ do not require, simply recommend, an external evaluator. However, the TFI and BoQ have additional strengths in the areas of sustained implementation and action planning. Therefore, authors suggested the use of the TFI or BoQ with an external evaluator. Notably, in the summer of 2022, the School-Wide Information System (SWIS), a commonly used system for tracking fidelity and outcomes (see: https://www.pbisapps.org; PBISApps, 2022), retired the SET and TIC. They instead recommended schools replace these fidelity tools with the TFI. Given the ever-evolving status of Tier 1 SWPBIS and subsequent measurement methods, the TFI is evaluated in this dissertation. The TFI was chosen due to the recommendations for use (e.g., PBISApps), the psychometric properties, and the novelty of application in unique levels such as high schools.

Tiered Fidelity Inventory. The TFI is unique in that it is able assess each tier individually with the option to also compute a composite score. These scores can be analyzed through two different factor models, including a 3-factor and a 10-factor model to measure each individual tier and subscales within the tiers (Grasley-Boy et al., 2023; Massar et al., 2019; see Figure 1.1). The 3-factor model accounts for a score at each tier apart from the overall implementation score across all tiers (i.e., Tier 1 score, Tier 2 score, Tier 3 score). The 10-factor model accounts for subscales within each tier. This includes the same three subscales for Tier 1 and 2: (a) teams, (b) implementation, and (c) evaluation; and four subscales in Tier 3: (a) teams,

(b) resources, (c) support plans, and (d) evaluation. Evidence for model fit of both factor structures is present (Grasley-Boy et al., 2023; Massar et al., 2019). Yet, Grasley-Boy and colleagues found that the 10-factor model had a statistically significantly stronger model fit than the 3-factor model via the chi-square difference test ($\chi^2(9) = 125.373$, $p < .001$). Authors noted that although the difference was statistically significant, model fit was strong for both the 3-factor ($\chi2(942) = 1{,}301.616$, p = < .001) and 10-factor ($\chi2(933) = 1{,}223.668$, p = < .001) models. These factor models were assessed using the full scale, yet most research analyzes Tier 1 scores specifically using a single Tier 1 factor that averages all 15 questions (Grasley-Boy et al., 2019; James et al., 2019; McIntosh et al., 2021; Noltemeyer, Palmer, James & Petrasek, 2019). No identified studies have compared the total Tier 1 model to the Tier 1 three factor (subscale) model, nor a weighted Tier 1 score derived from the three subscales as outlined by Grasley-Boy and colleagues (2023).

Further psychometrics for the TFI have been evaluated. This included a large-scale validation study by McIntosh and colleagues (2017) with 789 schools (69.9% elementary, 19.7% middle schools, 7.7% high schools, 2.8% other) across seven states in the 2013-2014 school year. Results indicated overall internal consistency via coefficient alpha of .96 for the full measure and .87 for Tier 1 specifically. These findings suggest that the TFI is consistently measuring the same construct (i.e., fidelity of implementation for SWPBIS schools). Authors also found acceptable convergent validity evidence in comparison to other Tier 1 SWPBIS fidelity tools including statistically significant correlations between the TFI Tier 1 measure and the BoQ (.416-.643), SAS (.364-.551) and the TIC (.258-.544) (without external coach – with external coach). Further, Grasley-Boy et al. (2023) re-examined the reliability of the TFI with data from the 2018-19 school year in 693 California schools (67% elementary, 16% middle schools, 11% high schools, 6% other). Results indicated acceptable internal consistency evidence at Tier 1 with an ordinal coefficient alpha of 0.94 and for the total score across all three tiers 0.98. The psychometrics of the TFI are encouraging and provide support for the use of the measure in research and practice. It is critical to use tools with sound psychometric evidence and follow appropriate reporting practices.

Fidelity Reporting. Often, fidelity of implementation is collapsed into one score indicating if a school met or did not meet full implementation criteria (See Table 2 for criteria). However, the full implementation average score does not provide sufficient detail on the extent to which each core element is being implemented. Researchers have suggested that best practice is to assess and report fidelity for *all core practice elements* of the intervention or framework (Gottfredson & Gottfredson, 2002; Molloy et al., 2013). Reporting fidelity for all Tier 1 core practice elements can help researchers and practitioners better understand associations between practices and relevant outcomes in data review and can provide information on which Tier 1 elements are most impactful on outcomes or effectiveness measures.

Fidelity and Effectiveness. Efforts have been undertaken to identify which specific core practice elements may be more influential than others at improving student outcomes. Understanding which core practice elements are more influential may facilitate implementation by identifying which Tier 1 core practice elements should initially receive the bulk of resources (Molloy et al., 2013; Weick, 1984). Researchers have primarily explored these relationships via correlational analyses using primary or secondary data. Molloy and colleagues (2013) evaluated SET and ODR data from 166 K-6 schools across seven states. Authors found three Tier 1 elements to be the most influential, such that ODRs were the lowest when the expectations taught, reward system, and violation systems subscales were implemented with fidelity (i.e., met the 80% criterion for each subscale). Pas and colleagues (2019) evaluated the statistical relationship between SET scores and student outcomes for 180 schools, including 82 elementary schools, 40 middle schools, and 58 high schools. Researchers analyzed binary compiler cut points on SET fidelity scores (i.e., 70%, 80%, 90%) to predict same-year suspensions, truancy, and math and reading achievement. Across school levels, findings indicated a statistically significant relationship between the subscales of expectations defined, expectations taught, expectations reinforced, and overall fidelity with improvements in suspension and truancy. Further findings indicated that significant associations between SET cut points and student outcomes were the most common in elementary schools when compared to middle or high schools. Given both studies used the same outdated fidelity tool and did not address contextual factors (e.g., representation of high schools, addressing adaptations), the generalizability of the findings is limited.

Context of Implementation

In its development, SWPBIS was conceptualized as a universally applicable framework for educational environments. However, when considering the effectiveness of an intervention or framework, it is also important to consider the context of implementation such as the setting and target population characteristics (Fixsen et al., 2005). Contextual factors require adaptations when translating research to practice; by using an implementation science approach we can better understand the relationship between adaptations and effectiveness (Fixsen et al., 2010). Context is critical to evaluate within the SWPBIS literature because each school carries unique characteristics that can serve as barriers or facilitators to implementation. Without taking context into account, we are unable to determine for whom and under what conditions Tier 1 SWPBIS is the most effective.

One such critical context can be related to student demographics, which can vary widely by school buildings even within the same school district. Demographic variations require schools to adapt the framework to fit the culture of the school building, students, and families. The adaptability of the framework has resulted in SWPBIS to be cited as a means to promote equity and decrease school discipline disproportionality (e.g., Gage et al., 2018; Gregory et al., 2017; Sugai & Horner, 2020; Vincent & Tobin, 2011). However, critics have called out SWPBIS for being conceptualized as a "race-neutral" practice in schools for years (Bal, 2018; Calais & Green, 2022; Fallon et al., 2021) and the integration of equity as the fifth core element to the framework came decades after initial conceptualization (Center on PBIS, 2023). Therefore, most research has failed to disaggregate data by demographic characteristics, and researchers that have disaggregated have found that racial discipline disproportionality persists across the country, even in schools implementing SWPBIS (Fallon et al., 2021; McIntosh, Gion, et al., 2018; Sugai & Horner, 2020; Weeks & Sullivan, 2024). This discrepancy between one foundational purpose to promote equity versus real-world trends indicates that the framework may be working for some students, but we cannot say that it is working fo*r every* student. To address this issue, calls for culturally responsive SWPBIS that prioritizes equity within Tier 1 practices have been made (Bal, 2018; Fallon et al., 2021; Sullivan et al., 2020). In order to evaluate these equity-centered practices within Tier 1, researchers must disaggregate data and analyze trends for racially and ethnically marginalized (REM) students.

Further examples of research evaluating unique contexts include implementation differences in alternative education settings (Grasley-Boy et al., 2021), rural settings (Garbacz et al., 2022), and high schools (Estrapala et al., 2021; Flannery & McGrath Kato, 2017). Evaluating these settings as unique contexts can help better understand differences and necessary adaptations. For example, researchers have found that high schools require more time to implement Tier 1 with fidelity compared to elementary and middle level schools due to unique barriers (Flannery & McGrath Kato, 2017). Thus, it may take more time for high school settings to realize the benefits of SWPBIS implementation, which could impact buy-in, commitment, and lead to premature abandonment of the framework. This is one reason it is critical to better understand the implementation process and impact on outcomes in high school settings specifically.

Tier 1 SWPBIS in High Schools

Research applying an implementation science lens to SWPBIS is in its infancy in regard to the evaluation of specific contexts such as high schools. Yet, high schools continue to implement SWPBIS at increasing rates (Flannery et al., 2018). In a two-year cohort analysis across grade levels, Muscott and colleagues (2004, 2008) found high schools had the lowest Tier 1 fidelity scores compared to other grades, yet still demonstrated reductions in ODRs. Further, in reviewing the literature searching for barriers to implementation at the high school level, Flannery and McGrath Kato (2017) identified common barriers such as school size, organization and culture, and student developmental level. These identified barriers and the increase of implementation at the high school level highlights the need for further research evaluating the high school level specifically and may indicate a need for adapted fidelity tools.

Foundational High School Studies

Early studies evaluated Tier 1 implementation in high schools and demonstrated increased fidelity of implementation through case studies and cohort analyses (e.g., Bohanon et al., 2006; Muscott et al., 2004, 2008). One such case study from Bohanon and colleagues (2006) documented the three-year process of implementation in an urban high school with roughly 1,800 students. Researchers approached implementation in three phases; the first phase prioritized open communication amongst teachers, support staff, and administrators. The second phase included evaluating and presenting discipline data for the school. The final phase prioritized implementation, starting with training school staff on positive behavior supports and developing school-wide systems. At the end of the three years the school did not reach full implementation criteria via the SET, with a score of 78% overall and 40% on teaching behavioral expectations. Yet, a 20% reduction in ODRs was observed from year 2 to year 3 of the study. This foundational case study established promise for SWPBIS implementation in high schools by demonstrating gains in Tier 1 fidelity scores and subsequent decreases in ODRs.

In addition to increasing fidelity of implementation, foundational research sought to identify impacts of Tier 1 for high school students with specific behavior profiles. This was explored by Lane and colleagues (2007) that evaluated the effectiveness of SWPBIS across two high schools. To identify behavior profiles, English teachers completed a modified Systematic Screening for Behavior Disorders which identified one student from each class period with behavior profiles labeled “externalizing,” “internalizing,” “comorbid,” “typical,” or “high incidence disability.” The modification to the screening tool and inclusion of only one student identified by English teachers limited the validity of the behavior profile labels and generalizability. As such, results should be interpreted with caution. Findings suggested large effects on unexcused tardies for students with typical (d = -0.72), internalizing (d = -0.60), and high incidence (d = -0.46) profiles. Smaller, yet educationally meaningful, effects were found on both GPA for students with externalizing (d = 0.22) and internalizing (d = 0.39) profiles, and on suspensions for students with internalizing (d = -0.27), typical (d = -0.21), and high incidence (d = -0.16) profiles. The greatest and most impactful effects across groups were demonstrated on unexcused tardies and suspensions. Interestingly, moderate to large effects in unexpected directions were observed on unexcused tardies for students with comorbid concerns (d = 0.36) and disciplinary contacts for students with externalizing concerns (d = 0.52). Overall, these effects were mixed and should be examined with caution, especially given the small sample size (n = 178) and overrepresentation of White (88%), male students (66%).

Continued Evaluations of Effectiveness

Following these initial cohort and case studies, researchers continued to evaluate the effectiveness of Tier 1 SWPBIS implementation at the high school level in numerous ways. This included evaluating the impact of coaching and technical support on fidelity scores (e.g., (Flannery et al., 2013). In addition, researchers have evaluated the impacts of implementation on outcomes such as bullying (e.g., Bradshaw et al., 2015; Smolkowski et al., 2016), attendance (e.g., Freeman et al., 2015, 2016), classroom management (e.g., Bradshaw et al., 2021), dropouts (e.g., Freeman et al., 2015; Malloy et al., 2018), and suspensions (e.g., Childs et al., 2016; Malloy et al., 2018; Smolkowski et al., 2016; Vincent & Tobin, 2011). These findings have been overall positive and therefore have continued to establish evidence that SWPBIS can be implemented in high schools with positive associations on student and school outcomes.

Given positive impacts on student outcomes, efforts have been undertaken to examine high school implementation differences and fidelity scores along with student and school level outcomes. Estrapala and colleagues (2021) conducted a systematic review of Tier 1 implementation in high schools in order to synthesize the current status of research. Authors included studies published through December 2018 and identified 16 studies for inclusion with 100% agreement on all full texts screened and coding procedures. It is important to note that this systematic review did not assess study quality, discuss effect sizes, or use the full implementation criteria for the most common fidelity measure reported (i.e., SET 80/80 criteria). The review highlighted that most studies were descriptive in nature (*n* = 13) directly referencing the lack of large-scale experimental designs (*n* = 3). Some studies reported adaptations to address contextual factors at the high school level, including longer staff trainings (Childs et al., 2010), integrating staff feedback and priorities in the trainings (Bohanon & Wu, 2014; Childs et al., 2010; Flannery et al., 2014), and varying levels of student involvement in planning (Bohanon et al., 2006, 2012; Bradshaw et al., 2015; Flannery et al., 2014; Malloy et al., 2018; Muscott et al., 2008). Authors were not able to determine the extent to which fidelity scores for core practice elements influenced student outcomes given the variability in detail of reporting across different fidelity tools and student outcomes measured. Even so, steady gains in fidelity scores and subsequent reductions in ODRs were observed.

Fidelity and ODRs

ODRs have been the most widely used measure for monitoring Tier 1 SWPBIS (Irvin et al., 2006; Sugai, Sprague, et al., 2000; Sugai & Horner, 2020) with validity evidence in predicting behavioral outcomes (McIntosh et al., 2010). As such, ODRs have been the most prevalent student outcome reported in high school SWPBIS studies (e.g., Barrett et al., 2008; Bohanon et al., 2006, 2012; Bohanon & Wu, 2014; Bruhn et al., 2022; Childs et al., 2010, 2016; Elrod et al., 2021; Flannery et al., 2014; Freeman et al., 2016, 2019; Malloy et al., 2018; Muscott et al., 2004, 2008). Most commonly, these studies have evaluated basic descriptive analyses documenting changes in ODR rates rather than more advanced statistical analyses (e.g., regression). Although these results are promising they are merely correlational and do not permit any causal inferences, which would require experimental manipulation.

Descriptive analyses have demonstrated decreases in ODRs after Tier 1 implementation in high schools, both at levels of statistical significance (Flannery et al., 2014; Freeman et al., 2019), and non-significance (Bruhn et al., 2022; Childs et al., 2016). Statistical findings related to ODRs can be found in Table 3. It is evident that substantial variability of analyses exists across studies evaluating ODRs at the high school level. The most common analysis appears to be the change point test, or determination about whether a change has taken place. Yet, this analysis tells us very little information except that statistically significant changes in ODRs may be more likely to happen around March.

Few studies were identified that specifically evaluated the relationship between ODRs and fidelity at the high school level. Freeman and colleagues (2019) collected data from 15 high schools in the Midwest to analyze the relationship between ODRs and BoQ fidelity scores. Authors conducted multilevel linear modeling with a restricted maximum likelihood estimation that controlled for student demographic variables including grade, race, gender, special education status, and access to free or reduced-price lunch. Findings indicated that for each one unit increase in the overall fidelity score as measured by the BoQ, there were 0.06 fewer ODRs and 0.012 fewer suspensions per student. Similarly, Flannery and colleagues (2014) collected data from eight intervention schools and four comparison schools to evaluate SET and ODR data over three years of implementation. Authors implemented a cohort-sequential design with multilevel linear modeling for missing ODR data. When controlling for enrollment and free or reduced-price lunch, authors found a significant decrease in ODRs (b = -0.31) with increased SET scores. In sum, these findings suggest a relationship between increased quality of implementation and decreased ODRs at the high school level, but with a small sample of schools. Indeed, the research base for the relationship between fidelity and ODRs at the high school level remains in its infancy.

General Discussion and Conclusions

In an effort to better understand implementation of Tier 1 SWPBIS in high schools, a narrative review was undertaken that described the theoretical roots of SWPBIS, implementation science, and current research efforts in high schools. We argue the importance of analyzing context when evaluating SWPBIS and the relationship between fidelity and outcomes in these contexts. We have reviewed the literature and found support for implementation of Tier 1 SWPBIS in high school settings. This review included preliminary analyses evaluating the relationship between fidelity and ODRs specifically at the high school level. Based on this review, there is a theoretical base with emerging experimental evidence of implementation at the high school level. However, we support the need for potential adaptations and modifications to address the implementation barriers of size, organization, and developmental level.

Literature Gaps

Through this review, gaps in the literature related to Tier 1 implementation have emerged. These include the lack of studies disaggregating data by grade level to understand the unique contextual factors each level brings. For example, Lee & Gage (2020) demonstrated considerable unexplained heterogeneity; we argue that this could be partially explained by context (e.g., school setting, grade level) and should be evaluated further in future research. This could be done through experimental studies focused on specific contexts like the high school level or fully disaggregating data collected across grade levels to account for school level differences. Further, given studies at the high school level are in their infancy, there is a lack of research analyzing the role equity plays in high school SWPBIS. Equity is the most recent addition to the key elements of the framework and as such needs to be evaluated in all contexts to determine best practices. This could be further evaluated in studies by disaggregating high school data by racial and ethnic group to determine if the framework is working for every student, or a select group of students. Another gap remains between the analysis of fidelity tools within specific contexts and investigating whether tools should be adapted to address unique contextual variables. An established example of this is the inclusion of an extra component of 1.1: Team Composition related to student representation in the PBIS team for high schools using the TFI (Algozzine et al., 2019). Yet, the TFI includes this item without sufficient high school representation in psychometric studies. Further, studies evaluating the psychometrics of the TFI indicate that the 10-factor model has the best fit, yet research heavily relies on a single Tier 1 factor derived from the 3-factor models. Future research is needed to analyze this relationship between the different scoring methods of the TFI Tier 1 scale.

Conclusion

Evidence exists that Tier 1 SWPBIS may be implemented in high schools with positive impacts on ODRs, but the specific core practice elements that most impact ODRs and other student outcomes have yet to be identified. Calls for educational research to go beyond asking if an intervention or framework *worked* and instead ask under what conditions and for whom it *worked* have been made (Cook et al., 2018). Within the SWPBIS literature, this may be accomplished by evaluating the specific contexts in isolation such as high schools, key elements within those contexts, and through disaggregated data. Findings suggest that it may be critical for high schools to modify implementation to prioritize buy-in and student involvement during initial implementation, yet these modifications are not commonly assessed via fidelity tools. Research is needed to understand which core practice elements are implemented with the highest and lowest levels of fidelity in an effort to identify the most influential Tier 1 practices at the high school level. Efforts are also needed to better understand the application of the TFI in high schools specifically and determine if any of the factor model variations have significant impacts on student outcomes. Overall, further research on the Tier 1 SWPBIS implementation and the application of the TFI in high schools is critical in order to best understand this unique level and maximize associations with student outcomes and use of school resources.

Chapter 3

Methods

Data Source

This secondary analysis leveraged a national dataset, PBISApps, that was developed and is maintained by Educational and Community Supports (ECS), a research unit at the University of Oregon. Specifically, the data in this study were retrieved from the School-Wide Information System (SWIS). According to PBISApps (2019), SWIS is a web-based information system that was built to collect, summarize, and use behavior data for decision making. The system stores school documented discipline events and allows schools easy access to behavior trends. It should be noted that SWIS is primarily used by schools that are invested in SWPBIS implementation given the costs associated with the platform. Therefore, the sample reflects a specific subsample of schools implementing SWPBIS. The specific data set for this dissertation was from the 2022-2023 school year, and all data were reported at the school level. Variables in the dataset included school demographics, office discipline referral (ODR) data, including ODRs per racial and ethnic group, and Tiered Fidelity Inventory (TFI; Algozzine et al., 2019) Tier 1 scores. Since the data were non-identifiable and included only school-level information, this study was deemed "not human research" by the Institutional Review Board (IBR) at the University of Minnesota through a determination form, indicating formal IRB approval was not necessary.

Measure

Tiered Fidelity Inventory

The TFI measures fidelity of implementation for all tiers in schools implementing SWPBIS (Algozzine et al., 2019). The Tier 1 measure includes 15 questions that can be divided into three subscales: Subscale 1: Teams (n = 2), Subscale 2: Implementation (n = 9), and Subscale

3: Evaluation (*n* = 4). Scores from the TFI Tier 1 measure can be calculated several different ways: (a) a total Tier 1 score representing a simple mean of all 15 item responses, (b) a weighted total Tier 1 score representing the mean of each subscale, or (c) separate scores for each of the three subscales. Indeed, various measurement models have been tested using the TFI, which are discussed in further detail below. Higher scores on the total score, weighted score, items, and subscales indicate higher fidelity of implementation.

Psychometrics of the TFI were explored in a large-scale validation study by McIntosh and colleagues (2017) with 789 schools across seven states in the 2013-2014 school year. Results indicated overall internal consistency evidence via coefficient alpha of .96 for the full measure and .87 for Tier 1. These findings suggested that the TFI indeed measures the same construct (i.e., fidelity of implementation for SWPBIS schools). Authors also found acceptable convergent validity evidence in comparison to other Tier 1 SWPBIS fidelity tools including significant correlations between the TFI Tier 1 measure and the Benchmarks of Quality (BOQ; r = .416-.643), PBIS Self-Assessment Survey (SAS; r = .364-.551) and the Team Implementation Checklist (TIC; r = .258-.544) scores are reported with and without an external coach respectively. Further, Grasley-Boy et al., (2023) re-examined the reliability and construct validity evidence of the TFI with data from the 2018-2019 school year in 703 California schools. Reliability findings indicated strong internal consistency evidence both for Tier 1 with an ordinal coefficient alpha of .94 and for the total score across all three Tiers of .98. Two Confirmatory Factor Analyses were conducted comparing a simple and more complex hierarchical model of the full measure structure (see Figure 1). Results suggested adequate content validity evidence and model fit for both measurement models (Comparative Fit Index: Model 1 = 0.961, Model 2 = 0.967). Although these numbers are similar, the model fit indices demonstrated a statistically significant better fit ($\chi^2(9) = 125.373$, $p < .001$) for the more complex hierarchical model (Figure 1, measurement model 3) compared to the less complex model (Figure 1, measurement model 1).

Variables

Office Discipline Referrals (ODRs)

The primary outcome variable in this study included the total count of both major and minor ODRs per school. ODRs were disaggregated to include total ODRs, ODRs for White students, and ODRs for racially and ethnically marginalized (REM) students. Although a more nuanced examination of ODRs by race/ethnicity would be preferred, given the number of schools in our sample (n = 161), there were not enough students in each racial or ethnic group to make meaningful interpretations of fully disaggregated data. Although limited, evaluating differences in ODRs between White and REM students was worthwhile given the persistence of racial and ethnic discipline disproportionality across the country (U.S. Department of Education 2023). Further, prior research has documented disproportionality in ODRs between White students compared to Black, Latino/a, and multiracial students (e.g., Anyon et al., 2018; Losen & Whitaker, 2017). The inclusion disaggregated ODRs in this dissertation by White and REM students allowed us to explore these discipline differences in our sample of high schools and explore any differential impacts of fidelity on this relationship.

Covariates

Common covariates in SWPBIS fidelity and discipline literature include race and ethnicity (Kim et al., 2018), student enrollment (Childs et al., 2016; Flannery et al., 2014; Grasley-Boy et al., 2019), the number of students receiving free and reduced priced lunch (FRL; Flannery et al., 2014; Grasley-Boy et al., 2019; Kim et al., 2018), gender or sex assigned at birth (Freeman et al., 2019), and locale (Grasley-Boy et al., 2019). Covariates are important in SWPBIS literature because they may reduce the error in the model that is due to extraneous variables. Given previous findings and the research questions in this dissertation, FRL, locale, school days, and sex assigned at birth were explored as potential covariates. The FRL variable included a percentage of students in each school that qualified for free or reduced priced lunch according to the most recent National Center for Education Statistics (NCES; 2021-2022) data. Locale was a four-level categorical variable that included city, suburban, town, and rural classifications. These categories were based on the NCES locale classifications and criteria. The school days variable included a count of the total number of school days reported by each school. The sex assigned at birth variable was according to the most recent NCES data and included the percent of males in each school during the 2021-2022 school year. It is important to note the sex variable only included the male-female binary and did not account for any students whose sex assigned at birth and gender do not align. Although studies do not control for sex assigned at birth consistently in SWPBIS literature, it was included as a potential covariate given male students are more likely to receive ODRs when compared to female peers (US Department of Education, 2023). Other common covariates such as student enrollment and student race/ethnicity were not included as covariates as they were already accounted for by outcomes and predictors in later discussed models.

Missing Data and Dataset Cleaning

Prior to computing any analyses, data cleaning and the percentage of missing data were evaluated across the dataset. First, variables that could be accurately computed using other variables in the dataset were added. For example, when the number of school days were not reported for a particular school, the average number of school days was input given the small range across the sample of schools (147-189). Further, the SWIS dataset included school-input demographic variables along with NCES data. Researchers used the school-input data as a foundation, but leveraged the NCES data to fill in demographic gaps where information was missing from the provided school-input data. It is important to note the NCES data was from the 2021-2022 school year, while the school-input demographic data was from the 2022-2023 school year. However, NCES data from the prior year is likely a more accurate estimate than using other imputation methods. Next, ODR rates (i.e., ODRs per day per 100 students) were calculated using school days, enrollment numbers for total students, White students, and REM students along with total ODRs for all students, White students, and REM students. Finally, listwise deletion was executed for 18 schools that reported mismatched ODR numbers by race and ethnicity. These schools had fewer than 90% of their total ODRs accounted for in the different race and ethnicity columns. For example, one school reported 328 total ODRs but only 3 White ODRs and 0 REM ODRs, another school reported 443 total ODRs but only 266 White ODRs and 28 REM ODRs. The reason for these discrepancies was unknown but was speculated to be due to reporting errors or schools not entering data on race and ethnicity. Given the use of White and REM ODRs in the analyses for this study, these 18 schools were not included in the final analytic sample.

After data cleaning, the total percent of missingness was calculated. There is little guidance to support specific thresholds for missing data beyond estimates that data with more than 10% missingness may be biased and 5% may be a more conservative threshold (Hyuk Lee & Huber Jr., 2021). Therefore, it was determined a priori that, if more than 5% of data were missing, data would be analyzed to determine if the data appeared to be missing at completely random (MCAR), missing at random (MAR), or missing not at random (MNAR). In total, 5.79% of the data was missing across the dataset. The only missing data were potential covariates which included locale, sex assigned at birth, and FRL. Patterns of missingness were explored and found to *not* be MCAR via Little's missing completely at random test ($p < .01$). Researchers then determined the data to be MNAR given there were patterns for missingness related to FRL by states. For example, every Georgia, Oregon, Ohio, and Massachusetts school in the dataset did not report FRL. After exploring the mechanism of missingness, multiple imputation (MI) was implemented. MI is a powerful approach to handle missing data that helps reduce bias that would be present if deleting data under all missing data conditions (Enders et al., 2020; Woods et al., 2021). Woods and colleagues (2021) note that MI is most useful in cases of MAR and MCAR but still offers advantages under MNAR compared to other approaches (e.g., deletion).

In order to obtain the most accurate and strongest MI dataset, three MI methods were initiated and compared using RStudio Version 2023.09.1+494. The first round of imputation included five datasets, which is the minimum recommended number (Graham et al., 2007). Second, as recommended by the same authors, 20 datasets were developed. Then the "HowManyImputations" (Von Hippel, 2020) function in R was used which resulted in a recommendation of 61 total datasets for MI. When comparing the three options (i.e., 5, 40, or 61 datasets), significance values for all three datasets ranged from .487 to .486, but researchers found a statistically significant difference amongst the standard errors across the datasets. Notably, ANOVA results between the dataset with 20 datasets and with 61 datasets was statistically significant ($F = 10183$, $p < .01$). Therefore, the MI with 61 datasets was used in this study aligning with Von Hippel's (2020) recommendations related to standard error estimates and MI.

Data Analysis Plan

All statistical analyses were conducted in RStudio Version 2023.09.1+494 using the following packages: CRAN, dplyr, performance, MASS, and mice. First, descriptive statistics were analyzed in order to understand characteristics of the data set, including sample *n*'s, means, standard deviations, minimums, and maximums of school demographics, locales, ODR counts, and ODRs per day per 100 students in the sample. Next, regression models were built. The dependent variables of interest in this dissertation were ODR rates. Traditional linear regression models may not be appropriate for count or rate data as they may violate assumptions and result in inaccurate model coefficients or biased standard errors (Coxe et al., 2009). Further, although Poisson regression models are often used for count and rate data, there was overdispersion present in the Poisson models of this dissertation, meaning the variance of the dependent variable was greater than the mean. Therefore, a negative binomial regression (NBR) was more appropriate for this study (see also: Coxe et al., 2009; Edmondson et al., 2022; Gardner et al., 1995). In NBR, coefficients are derived from changes in the outcome variable by a multiplicative factor of the predictor, which can then be exponentiated and interpreted as an incident rate ratio. In total, nine NBR models were developed for this study to answer research questions 2-4 and incidence rate ratios were evaluated. Predictors of these models included total Tier 1 scores, weighted Tier 1 scores, and Tier 1 subscale scores. Only main effects were tested and interpreted in this study due to limited power. The specifics of the models, assumptions, and analyses are discussed further below.

Outliers

Prior to analyses, the presence of potential outliers that could impact the NBR models were assessed. First, the Mahalanobis distance test (Mahalanobis, 1936) for multivariate outliers was conducted using all potential model variables and five cases were flagged. Next, the initial nine NBR models were run excluding the five flagged cases and two more cases that were identified as potential outliers via visual analysis of residual plots. The models were computed again excluding the seven cases total. As a sensitivity analysis, model fit of all three possible datasets were compared using Nagelkerke's R^2 and Akaike Information Criterion (AIC) values. Researchers found the models using the full dataset (including potential outliers) had the strongest model fit and the significance values of the models did not differ with or without identified outliers. Findings from Abulela & Harwell (2020) suggest the dataset with the strongest model fit should be used in the presence of outliers as long as they do not impact the significance values of predictors. Additionally, after investigation it was clear that the flagged outliers were from schools with a high number of ODRs for REM students, which was critical to include in the analyses of this dissertation. Therefore, the full dataset of 161 schools was included for NBR models.

Further, as an exploratory analysis, the relationship between total, White, and REM ODRs were explored via t-test analysis. Although the same outliers were present in the t-test analysis, sensitivity tests determined that the outliers did not impact the results of the t-test (i.e., significant values did not change when outliers were removed). This approach was discussed in consultation with a staff member of the Research Methodology Consulting Center (M. Abulela, personal communication, July 29, 2024).

Testing of Covariates

Sensitivity analyses were conducted to test the addition of potential covariates in the models. Each potential covariate was added to the models one by one to determine if they improved the overall model fit via Nagelkerke's R^2 and AIC values. Locale was the only covariate found to significantly improve model fit. Therefore, all other potential covariates (FRL, sex assigned at birth, school days) were dropped from the final models as they were deemed to not improve model performance. This approach was discussed in consultation with a staff member of the Research Methodology Consulting Center (M. Abulela, personal communication, June 20, 2024).

ODRs as an Outcome

ODRs offset by enrollment were used as the outcome variable in all NBR models. An offset variable may be used in a NBR to account for event rates when modeling count data (Gagnon et al., 2008). Adding enrollment as an offset variable allowed researchers to adjust the model for enrollment, meaning we were able to include rates rather than counts in our models (Roback & Legler, 2021, Chapter 4). Therefore, all models were offset by either log(total enrollment/100), log(White enrollment/100), or log(REM enrollment/100). The offset variable adjusted the count data to be comparable across schools of different sizes and with varied enrollment numbers of White and REM students. Beyond the ability to compare ODR rates across schools of different sizes, ODRs were also offset by enrollment to align closer to previous research given many studies to date have reported ODRs per day per 100 students (e.g., Bohanon et al., 2006, 2012; Bohanon & Wu, 2014; Bradshaw et al., 2010; Flannery et al., 2014). Although we were not able to account for both enrollment and school days as an offset variable, the approach did align closer to previous research compared to the alternative of total counts. Therefore, results for NBR models were interpreted as ODRs per 100 students, White student ODRs per 100 White students, and REM student ODRs per 100 REM students.

Research Question 1

Which components of Tier 1 SWPBIS as measured by the TFI are implemented with the highest and lowest levels of fidelity in high schools using the SWIS platform? (a) Which Tier 1 subscale? and (b) Which Tier 1 item(s)? was answered via descriptive statistics which included averages, standard deviations, minimums, and maximums of the TFI Tier 1 item score, total score, and subscale score in our dataset.

Research Question 2

Overall, what is the relationship between the total Tier 1 TFI score and ODRs in high schools implementing using the SWIS platform? and 2a: *How are these relationships different for White students compared to REM students?* was answered via three separate NBR models with the total Tier 1 TFI scores as a predictor variable, locale as a covariate, and outcome variables of: (1) total ODRs offset by total enrollment, (2) White student ODRs offset by White enrollment, and (3) REM student ODRs offset by REM enrollment. The specific equations were as follows:

Model 1: Y_1(Total ODRs offset by total enrollment) = $b_0 + b_1X_1$(total Tier 1 score) + b_2X_2(Locale) + ε

Model 2: Y_2(White student ODRs offset by total White students) = $b_0 + b_1X_1$(total Tier 1 score) + b_2X_2(Locale) + ε

Model 3: Y_3(REM student ODRs offset by total REM students) = $b_0 + b_1X_1$ (total Tier 1 score) + b_2X_2(Locale) + ε

Research Question 2b

Does a weighted Tier 1 score impact these relationships? was answered with the same outcome variables and locale covariate but with the weighted Tier 1 score as a predictor. These weighted scores were calculated by averaging scores in each Tier 1 subscale and then averaging all three subscale scores for an overall score. The specific equations were as follows:

Model 4: Y_1(Total ODRs offset by total enrollment) = $b_0 + b_1X_1$(weighted Tier 1 score) + b_2X_2 (Locale) + ε

Model 5: Y_2(White student ODRs offset by total White students) = $b_0 + b_1X_1$ (weighted Tier 1 score) + b_2X_2 (Locale) + ε

Model 6: Y_3(REM student ODRs offset by total REM students) = $b_0 + b_1X_1$ (weighted Tier 1 score) + b_2X_2 (Locale) + ε

Research Question 3

Overall, what is the relationship between the three TFI Tier 1 subscales and ODRs in high schools using the SWIS platform? and 3a: *How are these relationships different for White students compared to REM students?* was answered via three NBR models that included each

Tier 1 subscales (i.e., teams, implementation, evaluation) as predictors plus the same outcomes and covariate as previous models. The specific model equations were as follows:

Model 7: Y_1(Total ODRs offset by total students) = $b_0 + b_1X_1$ (Subscale 1: Teams) + b_2X_2(Subscale 2: Implementation) + b_3X_3(Subscale 3: Evaluation) + b_4X_4 (Locale) + ε

Model 8: Y_2(White student ODRs offset by total White students) = $b_0 + b_1X_1$ (Subscale 1: Teams) + b_2X_2(Subscale 2: Implementation) + b_3X_3(Subscale 3: Evaluation) + b_4X_4 (Locale) + ε

Model 9: Y_3(REM student ODRs offset by total REM students) = $b_0 + b_1X_1$ (Subscale 1: Teams) + b_2X_2(Subscale 2: Implementation) + b_3X_3(Subscale 3: Evaluation) + b_4X_4 (Locale) + ε

Research Question 4

Which model is a stronger predictor of ODRs in high schools implementing using the SWIS platform? Model fit was analyzed to determine if the total Tier 1 score, weighted Tier 1 score, or subscale scores were the strongest predictor of (1) total ODRs offset by total enrollment, (2) White student ODRs offset by White enrollment, and/or (3) REM student ODRs offset by REM enrollment. These relationships were compared by evaluating the Nagelkerke's R^2 and the AIC values. In order to answer this research question, three tests to compare performance were analyzed as only models with the same outcome variable can be compared. For example, Models 1, 4, and 7 were compared to each other for model fit as they all have total ODRs offset by total enrollment as the outcome variable. In general, higher Nagelkerke's R^2 values and lower AIC values indicate a stronger model fit.

Chapter 4

Results

Descriptive Analyses

School Characteristics

In total, 161 high schools were included in this study. Of these schools, 34 were in urban settings, 69 in suburban, 17 in towns, and 41 in rural settings. Schools from 25 states in the sample included Arizona (*n* = 3), Arkansas (*n* = 1), California (*n* = 17), Georgia (*n* = 36), Idaho (*n* = 2), Illinois (*n* = 3), Iowa (*n* = 1), Kansas (*n* = 1), Kentucky (*n* = 2), Maine (*n* = 2), Massachusetts (*n* = 2), Michigan (*n* = 32), Minnesota (*n* = 7), Missouri (*n* = 3), Nebraska (*n* = 3), Nevada (*n* = 1), New Hampshire (*n* = 1), North Dakota (*n* = 2), Ohio (*n* = 2), Oregon (*n* = 6), Pennsylvania (*n* = 10), Texas (*n* = 2), Virginia (*n* = 5), Washington (*n* = 16), and Wisconsin (*n* = 1). Student enrollment in the 161 schools totaled 192,222 students, ranging from 70 to 3099 with an average of 1194 enrolled in each school. The number of White students totaled 86,558 (45% of the sample), ranging from 7 to 1780 with an average of 537 enrolled in each school. The number of racially and ethnically marginalized (REM) students totaled 105,664 (55% of the sample), ranging from 7 to 2948 with an average of 656 enrolled in each school. Refer to Table 4 for more information on student demographics.

Office Discipline Referrals (ODRs)

All schools (*n* = 161) included information on ODRs. The count of major and minor ODRs totaled 183,372 ranging from 24 to 5740 with an average of 1139 per school. These ODRs averaged 0.67 per day per 100 students in our sample during the 2022-2023 school year. Further information related to ODRs can be found in Table 5.

ODRs for White Students and REM Students. The count of White student ODRs totaled 79,706 ranging from 0 to 4571, with an average of 495 per school. The count of White student ODRs per day per 100 White students ranged from 0-5.17 with an average of 0.55 per day per 100 White students. The count of REM student ODRs totaled 102,338 ranging from 1 to 4526 with an average of 636 per school. The count of REM student ODRs per day per 100 REM students ranged from 0.021-7.373 with an average of 0.86 ODRs per day per 100 REM students. Further, we found a statistically significant difference between White student ODRs per day per 100 White students and REM student ODRs per day per 100 REM students (t = -3.136, df = 273.77, p < .001) via t-test analysis.

Research Question 1: *Which Components of Tier 1 SWPBIS as Measured by the TFI are Implemented with the Highest and Lowest Levels of Fidelity in High Schools Using the SWIS Platform? a. Which Tier 1 Subscale? b. Which Tier 1 Items?*

All schools (n = 161) reported information on Tiered Fidelity Inventory (TFI; Algozzine et al., 2019) Tier 1 scores. In total, the sample averaged 74.3% on the total Tier 1 score with ranges from 10% to 100% and a standard deviation of 20.75%. According to the 70% cut score set by authors (Algozzine et al., 2019), 108 out of the 161 or 67% of the schools in our sample achieved full implementation fidelity via to the total Tier 1 score. Similarly, 111 out of the 161 or 69% of our sample achieved scores above 70% on the weighted Tier 1 score, however, previous research has not validated the use the 70% criterion for the weighted Tier 1 score.

Examining the Tier 1 Subscale scores, Subscale 3: Evaluation was implemented with the highest levels of fidelity with an average of 78.34%, a range from 0% to 100% and a standard deviation of 23.54% across all schools. The next highest was Subscale 1: Teams, with an average of 76.9%, a range from 0% to 100% and a standard deviation of 24.23% across all schools. Lastly, Subscale 2: Implementation was implemented with the lowest level of fidelity with an average of 71.95%, a range from 5.56% to 100% and a standard deviation of 21.84% across all schools.

Examining the Tier 1 item responses, the highest fidelity scores for individual items included Item 1.12: Discipline Data with an average score of 1.776 out of 2, Item 1.14 Fidelity Data with an average score of 1.721 out of 2, and Item 1.3: Behavioral Expectations with an average score of 1.665 out of 2. Alternatively, the lowest individual items included Item 1.11: Student, Family, and Community Involvement with an average score of 1.229 out of 2 and Item 1.9: Feedback and Acknowledgement with an average score of 1.273 out of 2. See Table 6 for more detailed TFI reporting.

Negative Binomial Regression (NBR) Models

Results related to the full NBR models can be found in Table 7. Overall, we found all nine models were statistically significant via chi-squared likelihood ratio tests that compared the full model to the intercept only model. However, examination of main effects resulted in interesting and unexpected patterns, which are described in further detail below.

Research Question 2: *Overall, What is the Relationship Between the Total Tier 1 TFI score and ODRs in High Schools Implementing Using the SWIS Platform? Research Question 2a: How are These Relationships Different for White Students Compared to REM Students?*

Table 8 displays the negative binomial regression (NBR) results of the total Tier 1 score predicting ODRs per 100 total students (Model 1), White students (Model 2), and REM students (Model 3). The results demonstrated that the total Tier 1 score was not a significant predictor of ODRs in any of the three models, holding locale constant. However, in all three models, the categorical locale variables were found to be statistically significant with urban as the reference group, holding all other variables constant. For example, in Model 1, students in suburban schools had an 83% higher rate of ODRs per 100 students than students in urban schools ($b = 0.606$, $p < .001$), holding all other variables constant. In Model 2, White students in rural schools had a 3.6

times higher rate of ODRs per 100 White students compared to White students in urban schools ($b = 1.278$, $p < .001$), holding all other variables constant.

Research Question 2b: Does a Weighted Tier 1 Score Impact These Relationships?

Table 9 displays NBR results of the weighted Tier 1 score predicting ODRs per 100 total students (Model 4), White students (Model 5), and REM students (Model 6). The weighted Tier 1 score was not a significant predictor of ODRs in Models 4 or 5 but was found to be a significant predictor in Model 6, while holding locale constant. Model 6 demonstrated a 0.7% *increase* in rates of REM student ODRs per 100 REM students for each 1-unit increase in the weighted Tier 1 score ($b = 0.007$, $p = .026$), holding locale constant. Similar to the first three models, Models 4-6 found the categorical locale variable to be statistically significant with urban as the reference group, holding all other variables constant. For example, in Model 4 this included an 82.5% higher rate of total ODRs per 100 students for students in suburban settings compared to urban settings ($b = 0.602$, $p < .001$), holding all other variables constant. In Model 5, White students in rural schools had a 3.5 times higher rate of ODRs per 100 White students compared to White students in Urban schools ($b = 1.270$, $p < .001$), holding all other variables constant.

Research Question 3: *Overall, What is the Relationship Between the Three TFI Tier 1 Subscales and ODRs in High Schools Using the SWIS Platform? Research Question 3a: How are these Relationships Different for White students Compared to REM Students?*

Table 10 displays NBR results of Tier 1 subscale scores predicting ODRs per 100 total students (Model 7), White students (Model 8), and REM students (Model 9). Subscale 1: Teams was not found to be a statistically significant predictor in Models 7, 8, or 9, holding all other variables constant. Subscale 2: Implementation was found to be a statistically significant predictor in Model 7 such that a 1% decrease in total ODR rates per 100 students for each 1 unit increase in the Subscale 2: Implementation score ($b = -0.009$, $p = .03$) was observed, holding all other variables constant. Similarly, Subscale 2: Implementation was also found to be a statistically significant predictor in Model 9 with a 1.2% decrease in REM student ODR rates per 100 REM students for each 1 unit increase in the Subscale 2: Implementation score (b = -0.012, p = .02), holding all other variables constant. Subscale 3: Evaluation was found to be statistically significant in all three models, holding all other variables constant. Model 7 demonstrated a 0.9% increase in total ODR rates per 100 students for each one unit increase in the Subscale 3: Evaluation score (b = 0.009, p = .02), holding all other variables constant. Model 8 also demonstrated a 0.9% increase in White student ODR rates per 100 White students for each one unit increase in the Subscale 3: Evaluation score (b = 0.009, p = .03), holding all other variables constant. Model 9 demonstrated an increase of 1.3% REM student ODR rates per 100 REM students for each one unit increase in the Subscale 3: Evaluation score (b = 0.013, p = .004), holding all other variables constant. Similar to all other models in our study, Models 7-9 found the categorical locale variable statistically significant with urban as the reference group and while holding all other variables constant. For example, in Model 7, this included a 72% higher rate of total ODRs per 100 students for students in suburban settings compared urban settings (b = 0.542, $p < .001$), holding all other variables constant. In Model 8, White students in rural schools had a 3.4 times higher rate of ODRs per 100 White students compared to White students in urban schools (b = 1.237, $p < .001$), holding all other variables constant.

Research Question 4: ***Which Model is a Stronger Predictor of ODRs in High Schools Implementing Using the SWIS Platform?***

Across predictors, Models 7-9 with subscales had the strongest model fit (Nagelkerke's R^2 = .307, .337, .292; Akaike Information Criterion (AIC) = 2543, 2150, 2330) for total ODRs, White ODRs, and REM ODRs respectively. The second strongest were Models 4-6 with the weighted Tier 1 score (Nagelkerke's R^2 = .261, .307, .236; AIC = 2545, 2149, 2333) for each grouping respectively, followed by Models 1-3 with the total Tier 1 score (Nagelkerke's R^2 = .253, .304, .223; AIC = 2545, 2150, 2335) for each grouping respectively. Overall, these findings indicate that the models which included subscales as predictors had the strongest model fit (See Table 7).

Chapter 5

Discussion

In an effort to better understand implementation of Tier 1 SWPBIS in high schools, a secondary analysis of data from PBISApps that included high schools using the School-Wide Information System (SWIS) data suite was conducted. The purpose of this analysis was to explore the extent to which Tier 1 was implemented with fidelity as measured by the Tiered Fidelity Inventory (TFI; Algozzine et al., 2019) and explore any associations between fidelity and office discipline referrals (ODRs) for total (all), White, and racially and ethnically marginalized (REM) students. These relationships were explored via nine negative binomial regression (NBR) models. There were a number of findings within this study that demonstrated mixed results in the relationship between Tier 1 implementation and ODRs. The total Tier 1 score was not a statistically significant predictor of ODRs per 100 students, holding all other variables constant. The weighted Tier 1 score was a statistically significant predictor of *increased* REM ODR rates per 100 REM students, holding all other variables constant. Subscale 2: Implementation was a statistically significant predictor of *decreased* total and REM ODR rates per 100 students, holding all other variables constant. Subscale 3: Evaluation was a statistically significant predictor of *increased* ODR rates per 100 students, holding all other variables constant. A more thorough overview of these mixed results is presented along with more detailed discussions of item level scores and the overall findings of ODRs in the sample of high schools.

Office Discipline Referrals

Research has established that *major* ODRs have been decreasing in SWIS high schools (PBISApps, 2023), yet, our findings suggest that the total (i.e., minor and major) number of ODRs may be remaining constant when comparing our results to previous findings. For example, the average ODRs per day per 100 students in our sample was 0.67, which is strikingly similar to findings from Freeman and colleagues (2017) that also reported a median of 0.67 ODRs per day per 100 students in 328 high schools that used the SWIS suite during the 2013-2014 school year. This suggests that the average (i.e., major and minor) ODRs in SWIS utilizing high schools have potentially not changed over the last nine years. However, it is important to note that neither our sample nor the Freeman and colleagues (2017) sample encompassed *all* high schools using the SWIS suite in the respective school years. Although total ODR rates may be persistent, it is a positive sign that researchers have found *major* ODRs decreasing in high schools using the SWIS suite (PBISApps, 2023). This could reflect that SWPBIS schools may be better at preventing major behavioral incidents. It could also reflect staff having more time to intervene and prioritize minor behavioral concerns if the large ones are not occurring as frequently, thus increasing documentation of minor ODR events. However, the finding that total ODR rates in high schools have remained the same is contrary to what would be hypothesized given implementation is increasing in both frequency and duration. It is also important to consider that these rates reflect the averages of our total sample and may not reflect the unique experiences of REM students. As such, attending to the findings for REM students is important.

Our finding of discipline differences between White and REM students emphasizes the need to disaggregate data by race and ethnicity. This is especially true given that the decreasing trends in major ODRs reported by PBISApps (2023) were not disaggregated by race and ethnicity. Yet, we found REM students had a statistically significant higher average number of ODRs per day per 100 students than White students in SWPBIS implementing schools using SWIS. Furthermore, a report on SWIS schools across grade levels found that Black students were 3.67 times more likely to experience an out of school suspension, another commonly reported discipline incident (McIntosh, Gion, et al., 2018). Our findings align with recent calls to consistently collect and disaggregate data as a means to better evaluate disproportionality and address the needs of every student (Sullivan et al., 2021, 2024). Reporting ODR data without disaggregation does not allow for evaluation of adding equity as a fifth key element to the SWPBIS framework, nor does it reflect the efforts of researchers that have been calling for racial and ethnic disaggregation of data to be the standard for years (e.g., Bal, 2018; Fallon et al., 2021; Skiba et al., 2014; Sullivan et al., 2020). Our findings are consistent with previous research that has documented the persistence of racial discipline disproportionality across the country, even in schools implementing SWPBIS (Fallon et al., 2021; McIntosh, Gion, et al., 2018; Sugai & Horner, 2020; Weeks & Sullivan, 2024). Understanding the relationship between discipline events, like ODRs, and fidelity may help to bridge this gap moving forward by highlighting any Tier 1 practices that are more or less influential at decreasing REM ODRs. Therefore, it is critical to consider the extent to which schools in our sample were implementing Tier 1 with fidelity when interpreting these discipline differences.

Fidelity Scores

Compared to earlier developed fidelity measures, research reporting fidelity scores using the TFI for high schools is limited. Within our sample, we found high schools that used the SWIS suite and the TFI in the 2022-2023 school year averaged 74.3% on the total Tier 1 score. This is above the 70% Tier 1 cut score set by the authors (Algozzine et al., 2019), which indicates that on average, our sample of high schools implemented Tier 1 SWPBIS with fidelity. Similar data has been reported in three peer reviewed studies that evaluated reliability and validity evidence of the TFI, along with one Educational and Community Supports (ECS) report. However, it is important to note that these previous studies were across all grade levels and the percentage of high schools in each sample was small, ranging from 7.7%-11.4%. In a TFI pilot study, McIntosh and colleagues (2017) found an average of 83.9% on the Tier 1 total score in the 2013-2014 school year in 789 schools. Notably, this was a validation study and conducted *before* the TFI was made publicly available. In a follow up study from the first year the TFI was publicly available (2014-2015 school year), Massar and colleagues (2019) found an average of 67% on the Tier 1 total score in 1708 schools. Four years following initial public availability (2018-2019 school year), Grassley-Boy and colleagues (2023) found an average of 83.87% for the Tier 1 total score in 703 California schools. Lastly, the ECS report from Kittelman and colleagues (2018) found an average of 74% for the Tier 1 total score in 8467 ECS schools during the 2016-2017 school year. Notable variability amongst these scores is present. However, in the limitations, McIntosh and colleagues (2017) noted that participating schools were likely more enthusiastic about the new measure and therefore potentially more committed to SPWBIS in general, given it was an initial validation study. In contrast, Massar and colleagues (2019) included data from ECS (SWIS) schools in the first year the TFI was made publicly available, with no direct contact between the schools and researchers. Therefore, it is reasonable that the Massar and colleagues (2019) follow up study reported a lower average Tier 1 total score compared to the initial validation study. Further, it is promising that the ECS report from Kittelman and colleagues (2018) documented higher scores, on average, in a larger sample two years after the Massar and colleague (2019) study. Overall, it is encouraging that our analytic sample reported a similar TFI Tier 1 total score average (74.3%) to the two most recent studies across grade levels including Grasley-Boy and colleagues (2023; 83.87%) and Kittelman and colleagues (2018; 74%). This is true especially given prior research has documented that high schools have struggled to achieve Tier 1 fidelity cut scores (Flannery & McGrath Kato, 2017). Yet, given the six-year gap between our data and the most comparable data from the Kittelman and colleagues (2018) report, we could expect schools in our sample to have higher fidelity scores, as time has been found to be the strongest predictor of fidelity (Bradshaw et al., 2021; Flannery & McGrath Kato, 2017; McIntosh, Mercer, et al., 2018). However, it is also unknown how long the specific schools in our sample have been implementing SWPBIS or using the SWIS suite. Although the average fidelity score in our sample was above the 70% cut score, it is also critical to analyze the proportion of schools that were able to achieve full implementation fidelity in our sample.

Research has suggested that a TFI Tier 1 total score above 70% is considered the threshold for 'implementing SWPBIS with fidelity' (Algozzine et al., 2019). In total, 108 out of the 161 or 67% of the schools in our sample scored above the 70% total Tier 1 cut score, while 53 scored below. Most large-scale studies to date have discussed the average total Tier 1 score of the sample, but not the proportion of schools meeting or exceeding cut scores for fidelity tools (e.g., Bradshaw et al., 2010; Elrod et al., 2021; Grasley-Boy et al., 2023; McIntosh et al., 2017). However, one report by Freeman and colleagues (2017) explored fidelity of implementation in 1668 high schools that used SWIS and PBISApps during the 2013-2014 school year did report this data. Researchers found that according to respective cut scores (See Table 2) 69.3% of the 792 schools that reported Benchmarks of Quality (BoQ; Kincaid et al., 2010), 54.5% of the 222 schools that reported School-Wide Evaluation Tool (SET; Sugai, Lewis-Palmer, et al., 2001), 33.6% of the 601 schools that reported Team Implementation Checklist (TIC; Sugai, Horner, et al., 2001), and 23.1% of the 917 schools that reported PBIS Self-Assessment Survey (SAS; Sugai, Horner, et al., 2000) scores achieved full implementation fidelity. Given research has reported adequate convergent validity amongst these measures (e.g., Mercer et al., 2017), it is encouraging that the proportion of high schools that achieved full fidelity via the TFI total Tier 1 score in our sample was higher than those from the 2013-2014 school year. This may support that SWPBIS has been implemented with higher fidelity over the last nine years. However, a more nuanced comparison with a larger sample and the same fidelity tools (e.g., the TFI) would provide more information. It would also be important to include any other Tier 1 scoring methods, such as the weighted Tier 1 score in these analyses.

In this study, we compared the traditional total Tier 1 score to the weighted Tier 1 score as proposed by Grasley-Boy and colleagues (2023). In our sample, 111 out of the 161 schools, or 69%, scored above the 70% cut score while 50 scored below, when weighted. However, the use of the 70% cut score has not been validated with the weighted Tier 1 score option. Overall, these findings on proportions of schools above the 70% cut score were similar between the total Tier 1 score and the weighted Tier 1 score in our sample. Further, there was not a meaningful difference in the NBR models that evaluated the total Tier 1 score compared to those that evaluated the weighted Tier 1 score. This may add to the evidence base that either scoring method may be useful as an overall Tier 1 score. However, it is interesting to note that Grasley-Boy and colleagues (2023) found the strongest model fit with subscales loading onto tiers which aligns conceptually with the weighted Tier 1 score. Indeed, either scoring method may suffice, but evaluating subscale and individual item scores may better represent implementation practices.

Subscale and Individual Item Scores

Researchers have suggested that best practice is to assess and report fidelity for all core practice elements of a framework, like SWPBIS (Gottfredson & Gottfredson, 2002; Molloy et al., 2013). Therefore, all subscales and item level scores for the TFI Tier 1 fidelity tool were reported in this study (See Table 6). Within our sample, we found an average of 76.9% on Subscale 1: Teams, 71.95% on Subscale 2: Implementation, and 78.34% on Subscale 3: Evaluation. In comparison, the sample from the previously discussed ECS report across grade levels from Kittelman and colleagues (2018) averaged 78%, 74%, and 73% respectively. Although these averages are similar, it is of note that Subscale 2: Implementation is lower in our sample of high schools and Subscale 3: Evaluation is higher. This could be due to the noted challenges with implementation of SWPBIS in high schools (Flannery & McGrath Kato, 2017). However, we would expect all school levels to report equivalent scores on Subscale 3: Evaluation as all

SWPBIS schools preferably have a focus on data and evaluation. Discussing some key differences in item level scores may provide a more robust understanding of subscale score differences.

Three individual items that were notably lower in our sample compared to the Kittelman and colleagues (2018) ECS report were all from Subscale 2: Implementation. These included Item 1.4: Teaching Expectations which was 1.44 in our study and 1.61 in the prior report, Item 1.9: Feedback and Acknowledgement 1.27 compared to 1.49, and Item 1.8: Classroom Procedures, 1.39 compared to 1.55. These comparisons align with previous findings that high schools struggle to implement teaching and reinforcing behaviors with fidelity (Bohannon et al., 2006; Kos, 2022). Yet, prior studies that have explored the most critical core practices of Tier 1 implementation across grade levels have identified that expectations defined, taught, rewarded, and the violations systems (i.e., discipline practices) may be the most influential on student outcomes (Molloy et al., 2013; Pas et al., 2019). This is of concern given teaching and reinforcing behaviors were two of the lowest implemented items in our sample of high schools and lower than findings in the ECS report. However, the studies that documented these critical core practices used the SET, not the TFI. One notable difference between the SET and the TFI is the number of questions for each core practice, for example, five questions represent expectations taught on the SET compared to only one question on the TFI. Therefore, the comparison should be interpreted with caution. Further, the lower fidelity scores for classroom procedures in our study may be due to the common barriers of high school implementation like school organization and structure (Flannery & McGrath Kato, 2017). Indeed, the three items that were lower in this study compared to the Kittelman and colleagues (2018) report aligns with previous research that has documented the challenges with high school implementation compared to implementation in lower-level schools.

Alternatively, three items were notably higher in our sample than in the Kittelman and colleagues (2018) ECS report. The first two items were from Subscale 2: Implementation and third item was from Subscale 3: Evaluation. These included Item 1.10: Faculty Involvement which was 1.51 compared to 1.37, Item 1.11: Student, Family, and Community Involvement which was 1.23 compared to 1.08, and Item 1.12: Discipline Data which was 1.78 compared to 1.60. Studies have identified that student involvement is a crucial element in high schools given the developmental level and desired autonomy of students (Estrapala et al., 2021; Flannery et al., 2009). Therefore, it is encouraging that Item 1.11: Student, Family, and Community Involvement was higher in our sample compared to the across grade levels sample. Yet, it was one of the lowest implemented items in our sample. Further, previous findings have discussed that, on average, high schools implement discipline practices with the most fidelity (Kos, 2022). Although Item 1.6: Discipline Policies was equivalent at 1.47 amongst our sample and the Kittelman and colleagues (2018) report, the higher score on Item 1.12: Discipline Data still may reflect higher implemented discipline practices in high schools as a whole. Although it is positive to report a higher fidelity score for Item 1.12: Discipline Data, this may still suggest that some high schools could be overly reliant on discipline practices and data rather than preventative approaches like teaching and reinforcing behavioral expectations. Overall, these item level differences indeed reflect the differences found in subscale scores between this study and the Kittelman and colleagues (2018) ECS report. These findings also align with previous item level findings on fidelity in high schools compared to lower-level schools. Next, it is critical to consider the NBR models in this study to better understand the overall relationship between fidelity and ODRs.

Negative Binomial Regression (NBR) Models

All models explored in this study were found to be statistically significant via a chi-squared likelihood ratio test that compared the full model to the intercept only model (see Table

7). Overall, the likelihood ratio test indicated that Tier 1 fidelity, with a locale covariate, was a statistically significant predictor of ODR rates in our sample. This statistically significant relationship of all models allowed us to further analyze predictor variables and model fit across NBR models, with a specific focus on main effects.

Fidelity and ODRs in High Schools

The most consistent finding in all of the models was the statistical significance of the categorical locale variable when holding all other variables constant. Results indicated that suburban schools, schools in towns, and rural schools had higher ODR rates per 100 students compared to urban schools for total students, White students, and REM students in our sample, holding all other variables constant. The reason for these locales reporting higher ODRs than urban schools is unclear given that all models were offset by enrollment. One possibility is that this could be a product of decreased resources and time available to document and track ODRs in urban settings. Another possibility is that there are different behavioral norms and expectations in how behaviors are perceived and managed in urban locales compared to other settings, such that misbehavior may be more strictly disciplined in other locales. Further, there could be inconsistency on what behaviors meet ODR thresholds across school personnel in different locales. Additionally, this study combined both major and minor ODRs together, so it could be that urban schools spend more time responding to major ODRs and therefore are less likely to report minor ODRs whereas schools with smaller student populations may be able to track both more accurately. Although speculative, these are all potential hypotheses as to why the urban schools reported lower ODR rates than suburban, town, and rural schools when holding all other variables constant. Locale remained the most consistent predictor in terms of statistical significance, but fidelity scores demonstrated mixed results when holding locale constant.

Overall, the total and weighted Tier 1 scores predicted *increased* ODR rates for all groupings when holding locale constant, albeit at mostly non-statistically significant levels. Yet, the weighted Tier 1 score *was* a statistically significant predictor ($p = .026$) of *increased* ODR rates per 100 REM students, holding locale constant. Specifically, Model 6 resulted in a 0.7% increase in REM student ODR rates per 100 REM students for each one unit increase in the weighted Tier 1 score when holding locale constant. This was not substantially different from the non-significant findings in Model 3 that demonstrated a 0.5% increase in REM student ODR rates per 100 REM students ($p = .093$) for each one unit increase in the total Tier 1 score when holding locale constant. The potential reasons for increased ODR rates and the statistically significant increase in REM ODR rates is unclear. Overall, this suggests that the total and weighted Tier 1 fidelity scores did not predict improvement of ODRs in our sample. This could reflect the implementation practices of the schools in our sample, perhaps there was not enough done to meet the developmental age of students, or the schools were early on in their implementation journey and ODRs have not been impacted yet by the framework. However, further evaluation of the subscale scores as predictors may provide more insight.

Subscale scores in Models 7-9 predicted mixed findings on ODR rates when holding all other variables constant. Subscale 1: Teams was not found to be a statistically significant predictor for ODRs when holding all other variables constant. Subscale 2: Implementation was a statistically significant predictor of *decreased* ODR rates on total and REM ODRs per 100 students but a non-significant predictor of decreased White ODR rates, holding all other variables constant. For example, Model 9 demonstrated that for each one-unit increase on Subscale 2: Implementation there was a statistically significant ($p = .025$) decrease of 1.2% REM ODR rates per 100 REM students when holding all other variables constant. Given the items that comprise Subscale 2: Implementation align most closely with core practice elements outlined in the most recent PBIS Implementation Blueprint (Center on PBIS, 2024), these findings reflect previous research that has found when Tier 1 core practice elements were implemented as designed, reductions in ODRs have been found (Algozzine et al., 2012; Bradshaw et al., 2010; Flannery et al., 2014). Further, this may suggest that, with time, prioritizing the core practice elements documented in Susbcale 2: Implementation may decrease the persistent discipline gap that is evident in this study and previous studies (e.g., Fallon et al., 2021; McIntosh, Gion, et al., 2018; Sugai & Horner, 2020; Weeks & Sullivan, 2024). However, Subscale 3: Evaluation was a statistically significant predictor of *increased* ODR rates (ranging from 0.9% - 1.3%), holding all other variables constant. Although it is unclear why this increase occurred in our sample, multiple hypotheses can be made. One hypothesis is that higher evaluation efforts correspond with higher surveillance efforts, thus increasing the number of ODRs. Another is schools that spend time reviewing discipline data may have a better understanding of what qualifies as an ODR for their district. This could potentially increase the number of ODR reports if team members are all clear on the criteria. Schools that spend time on data analysis may also have a more organized system for documenting and tracking ODRs. These schools could also have more resources available and have team members that spend more of their time documenting and tracking ODRs and are therefore more accurate in their accounts. Indeed, the total Tier 1 score, weighted Tier 1 score, and subscale scores as predictors resulted in mixed findings on ODR rates when holding all other variables constant. Next, evaluating model fit is critical to interpret these models and findings as a whole.

Model Fit

Across all nine models in this study, the models that included subscale scores as predictors resulted in the strongest model fit for total ODRs, White ODRs, and REM ODRs. The second highest model fit came from the weighted Tier 1 score predictor and the lowest model fit from the total Tier 1 score predictor according to both Nagelkerke's R^2 and Akaike Information Criterion (AIC) values. No other studies to date have compared the weighted Tier 1 score to the total Tier 1 score in this manner. Further, the Nagelkerke's R^2 values for all models ranged from .223 to .337 regarding them as potentially moderate according to traditional R^2 interpretations (e.g., Cohen, 1988; Ozili, 2023; See Table 7). However, it is important to note that interpretation of Nagelkerke's R^2 is unclear in the NBR literature beyond its use of comparing goodness of fit across models with the same outcome variable and from the same dataset. This may suggest that traditional methods for interpretation of R^2 may not be appropriate for Nagelkerke's R^2 and therefore the relationship should be interpreted with caution. Beyond the Nagelkerke's R^2 and AIC values, we found the models with subscale scores (i.e., Models 7-9) were the most informative given the significance of both Subscale 2: Implementation and Subscale 3: Evaluation when holding all other variables constant. Overall, these findings suggest models with subscale scores may be the most reflective of implementation and have the strongest model fit when evaluating the relationship between fidelity and ODRs at the high school level.

A Note on Equity

Overall, we observed that REM students had statistically significant higher ODR rates compared to White students in our sample. Additionally, on average, our sample achieved full fidelity of implementation according to the TFI Tier 1 scale. This again aligns with previous findings that discipline disproportionality exists, even in schools implementing SWPBIS (Fallon et al., 2021; McIntosh, Gion, et al., 2018; Sugai & Horner, 2020). This is especially worrisome given the historical context referencing the adaptability of the SWPBIS framework as a means to promote equity and decrease school discipline disproportionality (e.g., Gage et al., 2018; Gregory et al., 2017; Sugai & Horner, 2020; Vincent & Tobin, 2011). These findings align with critiques that SWPBIS was built to be race-neutral and has not historically done enough to de-center the influence of White supremacy in school systems (Calais & Green, 2022; Fallon et al., 2021; Gregory et al., 2017). Further, it is important to note again that the Center on PBIS only recently added equity as a fifth key element to the framework in 2023. Therefore, the relationship between equity and discipline disproportionality should be a focus moving forward but may not have been evident in our sample from the 2022-2023 school year.

Limitations

Although this study contributes to the literature base, it is not without limitations. First, it is important to consider the analytic sample. All schools analyzed in this study used the SWIS platform, which may be reflective of schools highly invested in SWPBIS given the cost associated with the platform. These findings are likely not reflective of schools that may attempt to implement SWPBIS without the same level of funding and resources as those using the SWIS platform. Our sample included 43% suburban schools; therefore, it is critical to note some of these findings may be more representative of suburban schools rather than urban, town or rural schools. Additionally, our sample was limited even further to only include high schools that used both the TFI and reported ODRs during the 2022-23 school year. These data and results were not representative of high schools using other fidelity tools. Furthermore, it is important to note that the analytic sample consisted of only 161 high schools, which is a relatively small number, therefore some analyses may not have been adequately powered and should be interpreted with caution.

Overall, the percentages of White and REM students were similar in our sample, but we were not able to report and analyze more specific and comprehensive outcomes for individual racial/ethnic groups. Grouping all REM students into one category was necessary for these analyses but does not align with best practice. As such, these findings do not reflect the experiences of each racial and ethnic group, only REM students in our sample as a whole. This approach may diminish the experiences of unique racial/ethnic groups, should be avoided if possible, and interpreted with caution.

Further, all of the data were reported at the school level. This was appropriate for the analyses and research questions posed in the current study, but there is no information on the impacts for individual students. Student level data would provide a better understanding of the unique experiences of students within schools. This may be critical when interpreting data related to ODRs for White students compared to REM students, yet we were missing that deeper level of understanding by restricting our analysis to schools. It is also critical to note that we discuss student outcomes in the research literature, but only evaluated ODRs in this study, leaving out critical information about associations between academic performance, mental health, and other disciplinary incidents like suspensions and expulsions. Additionally, given the relatively small sample, major and minor ODRs were combined into one category representing total ODRs. This should be avoided in future research if possible because the severity of these two types of ODRs can be vastly different. For example, in this study we treated an ODR where a student is in a physical altercation the same as an ODR where a student talked out of turn in class. Therefore, findings should be interpreted with caution. Analyses and comparisons of fidelity scores should also be interpreted with caution given we do not know how long the schools in our sample were implementing SWPBIS, nor the amount of outside support schools received. This made comparing the proportion of schools achieving fidelity via the TFI Tier 1 scale to previous research difficult.

Related to the analyses specifically, interaction terms could not be explored in the current study given the small sample and limited power. Therefore, results should be interpreted with caution, as we were unable to adequately test whether the effect of Tier 1 fidelity on ODRs changes as a function of locale. Further, Subscale 1: Teams only includes two questions that contribute to the score. Typically, it is recommended that a factor has at least three items loading onto it for analyses. The factor was used with only two items in this study as in alignment with previous research (e.g., Grasley-Boy et al., 2022 & Massar et al., 2019) and because no other options to analyze the subscales were available. Ideally, fidelity measures would provide more robust measurement opportunities in order to more appropriately evaluate specific subscales.

Implications for Practice

Research has documented discipline disproportionality between White students and REM students for years (e.g., Girvan et al., 2017; McIntosh, Gion, et al., 2018, 2021; Skiba et al., 2014). Unfortunately, we still found that ODR rates for REM students were significantly higher than ODR rates for White students in our sample. Given our findings and previous research, high schools should evaluate their Tier 1 framework to determine if it is designed to work for *every* student, or specifically White students. This could be done through evaluating disaggregated building level discipline data, conversations with students and parents (e.g., student, family, and community involvement), along with consulting outside Tier 1 and equity experts. Our findings indicate that high schools should also prioritize Subscale 2: Implementation as it demonstrated statistically significant decreases in REM ODR rates in our sample. Further, high schools should prioritize students, family, and community involvement when implementing SWPBIS. Although this may be seen as *only one* question of many on the TFI, research to date has identified student involvement to be a key element of high school implementation and therefore high schools may benefit from it being prioritized.

Implications for Future Research

As evident in the narrative review portion of this study, there is the need for a more robust research base on SWPBIS implementation in high schools. One area this study did not cover, but future studies should, is expanding the type of student outcome variables analyzed.

This could include more behavioral outcome variables like suspensions, truancy, tardies, etc. Other underexplored areas could include mental health and academic performance. These future analyses would help gain a more comprehensive understanding and address the evolving needs of our schools and culture. Further, future research should continue to report detailed fidelity scores including item and subscale scores, along with proportions of schools achieving full implementation fidelity via cut scores.

This study left unanswered questions that should be explored further such as investigating why some subscales resulted in increased ODR rates and the differences amongst locales. These questions can be further investigated to determine what could be driving these results. Future research should also prioritize disaggregated data so we can better understand if SWPBIS is impacting every student or a certain subset of students. In this study we found a significant difference between White and REM ODR rates, future research should explore this relationship more thoroughly with specific racial or ethnic groups that fall under the REM category. Relatedly, the impact of adding equity as a key element in the SWPBIS framework should be further evaluated given its recent addition. This study does not reflect any potential impacts of adding equity as an element to the framework, but future studies with samples after the 2023 school year may be able to explore this.

Further, as fidelity tools keep evolving and expanding, our research needs to follow suit. Including further exploration comparing the total and weighted Tier 1 options to understand if either one is more advantageous in measuring fidelity. Additionally, research on Tiers 2 and 3 in high schools should continue to grow in order to best serve all high school students with differing levels of support needs.

Conclusion

The present study sought to better understand implementation of Tier 1 SWPBIS in high schools through a secondary analysis that evaluated the relationship between fidelity and ODRs. The sample included 161 schools across 25 states. We found that on average, the high schools in our sample were able to achieve full fidelity via the 70% TFI total Tier 1 cut point. This included 67% of the schools achieving full implementation fidelity, which is a larger proportion of schools than documented in previous research. Findings demonstrated that Item 1.9 Feedback and Acknowledgement and Item 1.11 Student/Family/Community Involvement were two of the lowest implemented items in our sample, though arguably being two of the most critical. Further, we found that Subscale 2: Implementation resulted in a statistically significant *decreases* in ODR rates for both total students and REM students when holding all other variables constant. Therefore, it may be critical for schools to continue prioritizing Subscale 2: Implementation practices which align with SWPBIS core practices, while also working to understand the role that Subscale 3: Evaluation may play. Overall, our findings were mixed on the extent to which TFI Tier 1 scores predicted ODR rates in our sample of high schools. As SWPBIS and the TFI continue to grow in popularity, it is critical to evaluate specific contexts, such as high schools, and disaggregate data to the fullest extent possible to determine the degree to which SWPBIS fidelity impacts outcomes for *every* student.

Table 1

Historical Lens of Tier 1 Core Elements

Checklist for Determining the Adequacy of an Existing School-wide Discipline Plan	SET – *The Seven Key Features of School-Wide Positive Behavior Support*	Center on PBIS, 2015	Center on PBIS, 2024	Center on PBIS, 2024
(Colvin et al., 1993)	(Horner et al., 2004)	PBIS Implementation Blueprint ***Practices***	PBIS Implementation Blueprint ***Systems***	PBIS Implementation Blueprint ***Practices***
1. The purpose of school-wide discipline is clearly stated. 2. School-wide behavioral expectations are clearly stated. 3. School-wide structures are in place to reinforce demonstrations of expected behaviors. 4. School-wide structures are in place to teach expected behaviors. 5. Staff are clear in which behaviors should be dealt with by staff and which should warrant office referrals. 6. Structures are in place for staff to work together to address persistent minor behavior. 7. A continuum of structures is available to address serious office referral behavior. 8. Structures are in place to use building resources to assist students with chronic serious behavior. 9. Record keeping procedures are in place to meet the needs for readily tracking student behavior.	1. Define 3 to 5 school-wide expectations for appropriate behavior. 2. Actively teach the school-wide behavioral expectations to all students. 3. Monitor and acknowledge students for engaging in behavioral expectations. 4. Correct problem behaviors using a consistently administered continuum of behavioral consequences. 5. Gather and use information about student behavior to evaluate and guide decision-making. 6. An administrator actively supports and is involved in the effort. 7. The school district provides support.	1. Set of school-wide positive expectations and behaviors are defined and taught. 2. Procedures for establishing classroom expectations and routines that are consistent with school-wide expectations. 3. Continuum of procedures for encouraging expected behavior. 4. Continuum of procedures for discouraging problem behavior. 5. Procedures for encouraging school-family partnerships.	1. Illustrations or posters with “rules” or guide points followed by concrete examples and non-examples. 2. Diverse leadership team representative of partners with active administrator, family, and community participation. 3. Efficient routine, schedule, and structure for team operating procedures. 4. Commitment or vision statement for establishing a positive social culture. 5. Procedures for: (a) on-going databased monitoring, evaluation, and dissemination; (b) selection, training and coaching of new personnel; and (c) evaluation of personnel related to PBIS implementation.	1. Co-designed positive Tier 1 behavior expectations are defined, taught, practiced, and reinforced schoolwide, including classroom routines consistent with schoolwide expectations. 2. Continuum of procedures for: (a) encouraging expected behavior; and (b) discouraging unexpected or unwanted behavior 3. Illustrations of low intensity supports threaded throughout the tiers (e.g., Behavior Support Plan, pre-correction). 4. Documentation of relation (a) between social validity (i.e., acceptability) and implementation fidelity; and (b) how screening data predicts student outcomes.

Table 2

Comparing Tier 1 Fidelity Tools

Measure	Number of Tier 1 Items	Respondent(s)	Tier 1 Subscales	*Full* Implementation Criteria
SET (Tier 1 specific)	28	External Evaluator	A. Expectations defined B. Behavioral expectations taught C. On-going system for rewarding behavioral expectations D. System for responding to behavioral violations E. Monitoring & decision making F. Management G. District-level support	'"80/80 criterion" – 80% on both subscale b and total (Horner et al., 2004)
TFI (Measures all 3 tiers)	15	SWPBIS team members with (recommended external) coach	1. Evaluation 2. Teams 3. Interventions	70% total on Tier 1 scale
BoQ (Tier 1 specific)	53	SWPBIS team members and internal or external coach	1. PBIS team 2. Faculty commitment 3. Effective procedures for dealing with discipline 4. Data entry & analysis plan established 5. Expectations & rules developed 6. Reward/recognition program established 7. Lesson plans for teaching expectations/rules 8. Implementation plan 9. Classroom systems 10. Evaluation	70% total
TIC (Tier 1 specific)	22	SWPBIS team members	1. Establish commitment 2. Establish & maintain team 3. Self-assessment 4. Establish school-wide expectations: Prevention system 5. Classroom behavior support system 6. Establish information systems 7. Build capacity for function-based support	80% total
SAS	18	All school staff	1. School-wide systems 2. Non-classroom setting systems	80% on school-wide systems

(Measures all 3 tiers)	3. Classroom systems 4. Individual student systems

Table 3

Statistical findings on the Impact of SWPBIS on ODRs in High Schools

Study	Outcome	Statistical Analysis	Statistical Finding	Fidelity
Bohanon et al. (2006)	ODRs (students with 0-1) from 2nd to 3rd year of implementation	Pearson's chi square	$\chi 2(3, N = 150) = 53.199, p < .001$	SET: did not meet 80/80
Bohanon et al. (2012)	ODRs	Change point test	February (Year 2) ($z = 3.6, p < .0001$) March (Year 3) ($z = 2.02, p < .0214$)	SET: met 80/80
Bohanon & Wu (2014)	ODRs	Change point test	March year 2 Treatment school 1 (z = 2.30, p D .022) Treatment school 2 (z = 3.07, p D .002 for T2	SET: Treatment school 1 = 99%, Treatment school 2 = 74%
Flannery et al. (2014)	ODRs (referenced as "problem behavior")	Multilevel latent growth model	Year 3: ($b = - 3.18, p < .01$)	SET (Not reported)
Freeman et al. (2016)	ODRs – SWIS	Structural equation modeling	Group 1: (-0.81, $p <.001$) Group 2: (-1.07, $p <.001$)	Group 1 = schools "approaching" fidelity Group 2 = schools meeting fidelity
Freeman et al. (2019)	ODRs	Multilevel linear modeling	$\gamma 01 = -0.060, p < .05$	BoQ: met, mean = 78.2
Malloy et al. (2018)	ODRs – SWIS	Change point test	December ($z = 3.67, p < .001$)	SET: Met 93 overall, 90 expectations taught

Note. ODRs can be inconsistent due to factors that impact the likelihood of a student receiving an ODR such as student behavior, adult response to that behavior, and organizational expectations (Sugai & Horner, 2020). Due to this variability, SWIS was developed in 2004 to monitor both the frequency of ODRs and the nature of the ODR (e.g., severity, type, location, administrator decision) along with fidelity scores. This had allowed for more consistency and detailed reporting for data analysis. As of the 2021-2022 school year, 4,233 schools used the SWIS to track ODRs (retrieved from www.pbisapps.org).

Table 4

Sample Demographic Data

Sample Across Schools	Count Or Average	Range	Standard Deviation
Total Number of High Schools	161		
Total Number of Students	192,222 ($\bar{x}$ = 1194)	70 - 3099	
% White Students	$\bar{x}$ = 50.42	0.37- 96.93	0.298
% REM Students	$\bar{x}$ = 49.47	3.84 – 99.62	0.297
Potential Covariates			
% FRL	$\bar{x}$ = 39.73	5.14 – 99.86	0.227
% Male	$\bar{x}$ = 50.49	37.72 – 56.23	0.027
Urban	34		
Suburban	69		
Town	17		
Rural	41		

Note. $\bar{x}$ = mean

Table 5

ODR Descriptive Data

ODR Type Across Schools	Average	Range	Standard Deviation
TOTAL ODRs	1138.95	24 - 5740	1104.86
Per day per 100 students	0.67	0.039 – 5.67	0.7327
TOTAL ODRs for White Students	495.068	0 - 4571	694.07
Per day per 100 White students	0.55	0 – 5.17	0.661
TOTAL ODRs for REM Students	635.95	1 – 4526	668
Per day per 100 REM students	0.857	0.021 – 7.373	1.022

Note. ODR = Office Discipline Referral. REM = Racially and Ethnically Marginalized, it is used in this study to encompass all actively marginalized groups within Western Culture. However, we recognize this term and method unrightfully groups all of these students that face differing levels of injustice together. Although this is not best practice, this grouping allows more power in our analyses and may provide further insights on the differences between how marginalized and non-marginalized groups are impacted by SWPBIS.

Table 6

TFI Tier 1 Item and Subscale Scores

Feature	Average Percentage (Score)	Range Across Schools (Score)	Standard Deviation Percentage (Points)
SUBSCALE 1: TEAMS	76.9% (3.07)	0% (0) – 100% (4)	24.23 (0.972)
1.1 Team Composition	1.466	0 – 2	0.524
1.2 Team Operating Procedures	1.608	0 – 2	0.604
SUBSCALE 2: IMPLEMENTATION	71.95% (12.95)	5.56% (1) – 100% (18)	21.84 (3.93)
1.3 Behavioral Expectations	1.665	0 – 2	0.524
1.4 Teaching Expectations	1.441	0 – 2	0.600
1.5 Problem Behavior Definition	1.553	0 – 2	0.558
1.6 Discipline Policies	1.466	0 – 2	0.623
1.7 Professional Development	1.416	0 – 2	0.648
1.8 Classroom Procedures	1.398	0 – 2	.0573
1.9 Feedback and Acknowledgement	1.273	0 – 2	0.707
1.10 Faculty Involvement	1.509	0 – 2	0.634
1.11 Student/Family/Community Involvement	1.229	0 – 2	0.776
SUBSCALE 3: EVALUATION	78.34% (6.27)	0% (0) – 100% (8)	23.542 (1.88)
1.12 Discipline Data	1.776	0 – 2	0.524
1.13 Data-Based Decision Making	1.403	0 – 2	0.702
1.14 Fidelity Data	1.721	0 – 2	0.561
1.15 Annual Evaluation	1.366	0 – 2	0.677
TOTAL TIER 1 SCORES	74.3% (22.29)	10% (3) – 100% (30)	20.75 (6.229)

Note. Full tool is available at https://www.pbis.org/resource/tfi

Table 7

Negative Binomial Regression Model Likelihood Ratio Test Results

Model	df (Resid. Df)	Likelihood Ratio Test Statistic	Pr(Chi)	Nagelkerke's R^2	AIC
Model 1	4 (156)	30.14	< .001	.253	2545.6
Model 2	4 (156)	37.72	< .001	.304	2150.35
Model 3	4 (156)	26.61	< .001	.223	2335.25
Model 4	4 (156)	31.27	< .001	.261	2545.02
Model 5	4 (156)	38.19	< .001	.307	2149.88
Model 6	4 (156)	28.27	< .001	.236	2333.59
Model 7	6 (154)	37.54	< .001	.307	2543.11
Model 8	6 (154)	42.42	< .001	.337	2150.06
Model 9	6 (154)	35.86	< .001	.292	2330.40

Note. Df = degrees of freedom. Resid. Df = residual degrees of freedom. Pr(Chi) = significance value of the chi squared test. AIC = Akaike Information Criterion.

Table 8

Negative Binomial Regression Models Comparing TFI Tier 1 Total Score and ODRs

Predictors	*b*	SE	*p*	IRR [95% CI]
Model 1				
Intercept	3.876	0.244	< .001	48.12 [28.99-82.37]
TFI total Tier 1 score	0.003	0.003	.28	1.003 [0.99-1.01]
Locale - Suburb	0.606	0.158	< .001	1.833 [1.34-2.48]
Locale – Town	0.869	0.224	< .001	2.385 1.55-3.75]
Locale – Rural	0.972	0.176	< .001	2.644 [1.86-3.73]
Model 2				
Intercept	3.537	0.276	< .001	34.37 [19.63-62.22]
TFI total Tier 1 score	0.002	0.003	.404	1.002 [0.99-1.01]
Locale - Suburb	0.723	0.179	< .001	2.061 [1.44-2.914]
Locale – Town	0.987	0.254	< .001	2.683 [1.65-4.48]
Locale – Rural	1.278	0.199	< .001	3.592 [2.42-5.31]
Model 3				
Intercept	3.849	0.279	< .001	46.95 [26.32-86.75]
TFI total Tier 1 score	0.005	0.003	.093	1.005 [0.99-1.01]
Locale - Suburb	0.771	0.18	< .001	2.161 [1.51-3.06]
Locale – Town	1.055	0.256	< .001	2.874 [1.76-4.82]
Locale – Rural	0.957	0.201	< .001	2.603 [1.75-3.86]

Note. N = 161 schools. IRR = incidence rate ratio, which is the exponentiated unstandardized regression coefficient. SE = Standard error, p = significance value. CI = confidence interval. Each regression contains an offset variable of enrollment with the coefficient restrained at 1.

Table 9

Negative Binomial Regression Models Comparing TFI Weighted Tier 1 Score and ODRs

Predictors	*b*	SE	*p*	IRR [95% CI]
Model 4				
Intercept	3.761	0.247	< .001	42.99 [25.77-73.79]
Weighted Tier 1 score	0.004	0.003	.109	1.004 [0.99-1.01]
Locale - Suburb	0.602	0.158	< .001	1.825 [1.33-2.47]
Locale – Town	0.858	0.224	< .001	2.359 [1.54-3.70]
Locale – Rural	0.961	0.175	< .001	2.614 [1.85-3.68]
Model 5				
Intercept	3.462	0.279	< .001	31.88 [18.05-58.27]
Weighted Tier 1 score	0.003	0.003	.256	1.003 [0.99-1.01]
Locale - Suburb	0.720	0.179	< .001	2.055 1.43-2.90]
Locale – Town	0.977	0.254	< .001	2.657 [1.63-4.43]
Locale – Rural	1.270	0.199	< .001	3.562 [2.40-5.26]
Model 6				
Intercept	3.707	0.281	< .001	40.75 [22.80-75.50]
Weighted Tier 1 score	0.007	0.003	.026	1.007 [1-1.01]
Locale - Suburb	0.767	0.179	< .001	2.154 [1.50-3.04]
Locale – Town	1.039	0.255	< .001	2.827 [1.74-4.73]
Locale – Rural	0.944	0.199	< .001	2.571 [1.73-3.80]

Note. N = 161 schools. IRR = incidence rate ratio, which is the exponentiated unstandardized regression coefficient. SE = Standard error, *p* = significance value. CI = confidence interval. Each regression contains an offset variable of enrollment with the coefficient restrained at 1.

Table 10

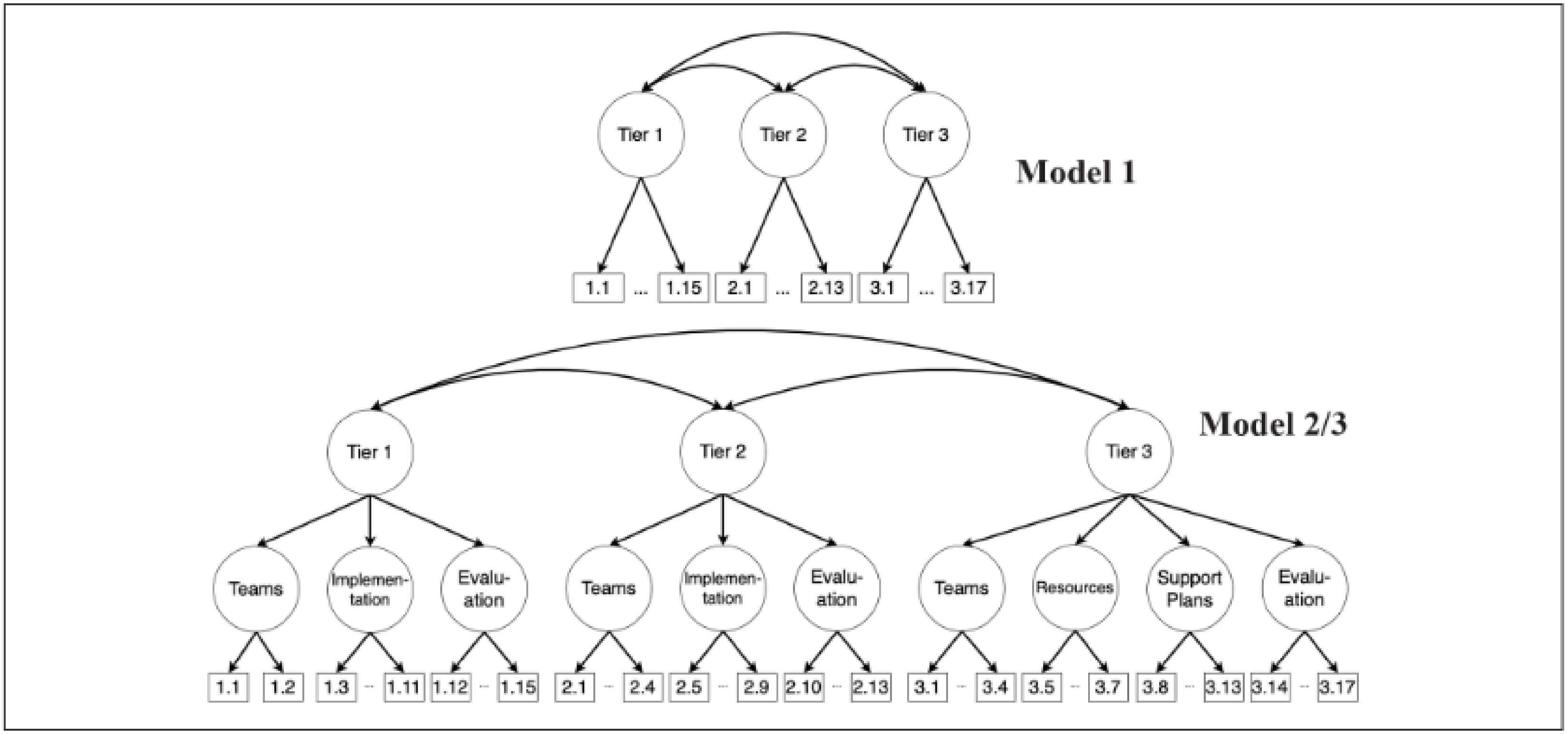
Negative binomial regression models comparing Subscale Tier 1 scores and ODRs

Predictors	*b*	SE	*p*	IRR [95% CI]
Model 7				
Intercept	3.765	0.243	< .001	43.16 [26.25-72.87]
Subscale 1: Teams	0.005	0.003	.157	1.004 [0.99-1.01]
Subscale 2: Implementation	-0.009	0.005	.032	0.99 [0.98-0.99]
Subscale 3: Evaluation	0.009	0.004	.021	1.009 [1-1.01]
Locale - Suburb	0.542	0.157	< .001	1.719 [1.26-2.32]
Locale – Town	0.770	0.221	< .001	2.16 [1.41-3.38]
Locale – Rural	0.928	0.173	< .001	2.529 1.79-3.55]
Model 8				
Intercept	3.455	0.276	< .001	31.68 [18.17-57.03]
Subscale 1: Teams	0.002	0.004	.552	1.002 [0.99-1.01]
Subscale 2: Implementation	-0.008	0.005	.117	0.991 [0.98-1]
Subscale 3: Evaluation	0.009	0.004	.036	1.009 [1-1.02]
Locale - Suburb	0.647	0.179	< .001	1.909 [1.33-2.71]
Locale – Town	0.895	0.252	< .001	2.447 [1.50-4.08]
Locale – Rural	1.237	0.197	< .001	3.445 [2.33-5.07]
Model 9				
Intercept	3.742	0.276	< .001	42.18 [24.07-76.31]
Subscale 1: Teams	0.005	0.004	.172	1.005 [0.99-1.01]
Subscale 2: Implementation	-0.012	0.005	.025	0.988 [0.978-0.998]
Subscale 3: Evaluation	0.013	0.004	.004	1.013 [1-1.02]
Locale - Suburb	0.699	0.178	< .001	2.012 [1.41-2.84]
Locale – Town	0.938	0.251	< .001	2.554 [1.57-4.25]
Locale – Rural	0.874	0.196	< .001	2.395 [1.62-3.53]

Note. N = 161 schools. IRR = incidence rate ratio, which is the exponentiated unstandardized regression coefficient. SE = Standard error, p = significance value. CI = confidence interval. Each regression contains an offset variable of enrollment with the coefficient restrained at 1

Figure 1

Confirmatory factor analysis models from Grasley-Boy et al., 2023



References

Abulela, M. A., & Harwell, M. (2020). Strengthening inferences in quantitative education studies conducted by novice researchers: Capitalizing on standards for sampling, research design, and instrumentation. *Research Highlights in Education and Science*, *20*(1), 59-78. https://doi.org/10.12738/jestp.2020.1.005

Algozzine, B., Barrett, S., Eber, L., George, H., Horner, R., Lewis, T., Putnam, B., Swain-Bradway, J., McIntosh, K., & Sugai, G. (2019). *SWPBIS Tiered Fidelity Inventory. Version 2.1.* Center on Positive Behavioral Interventions and Supports. https://www.pbis.org/resource/tfi

Algozzine, B., Wang, C., White, R., Cooke, N., Marr, M. B., Algozzine, K., Helf, S. S., & Duran, G. Z. (2012). Effects of multi-tier academic and behavior instruction on difficult-to-teach students. *Exceptional Children*, *79*(1), 45–64. https://doi.org/10.1177/001440291207900103

Anyon, Y., Lechuga, C., Ortega, D., Downing, B., Greer, E., & Simmons, J. (2018). An exploration of the relationships between student racial background and the school sub-contexts of office discipline referrals: A critical race theory analysis. *Race Ethnicity and Education*, *21*(3), 390–406. https://doi.org/10.1080/13613324.2017.1328594

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1*(1), 91–97. https://doi.org/10.1901/jaba.1968.1-91

Bal, A. (2018). Culturally responsive positive behavioral interventions and supports: A process–oriented framework for systemic transformation. *Review of Education, Pedagogy, and Cultural Studies*, *40*(2), 144–174. https://doi.org/10.1080/10714413.2017.1417579

Barrett, S. B., Bradshaw, C. P., & Lewis-Palmer, T. (2008). Maryland statewide PBIS initiative: Systems, evaluation, and next steps. *Journal of Positive Behavior Interventions*, *10*(2), 105–114. https://doi.org/10.1177/1098300707312541

Bohanon, H., Fenning, P., Carney, K. L., Minnis-Kim, M. J., Anderson-Harriss, S., Moroz, K. B., Hicks, K. J., Kasper, B. B., Culos, C., Sailor, W., & Pigott, T. D. (2006). Schoolwide application of positive behavior support in an urban high school: A case study. *Journal of Positive Behavior Interventions*, *8*(3), 131–145. https://doi.org/10.1177/10983007060080030201

Bohanon, H., Fenning, P., Hicks, K., Weber, S., Thier, K., Aikins, B., Morrissey, K., Briggs, A., Bartucci, G., McArdle, L., Hoeper, L., & Irvin, L. (2012). A case example of the implementation of schoolwide positive behavior support in a high school setting using change point test analysis. *Preventing School Failure*, *56*(2), 91–103. https://doi.org/10.1080/1045988X.2011.588973

Bohanon, H., & Wu, M. J. (2014). Developing buy-in for positive behavior support in secondary settings. *Preventing School Failure*, *58*(4), 223–229. https://doi.org/10.1080/1045988X.2013.798774

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. https://doi.org/10.1177/1098300709334798

Bradshaw, C. P., Pas, E. T., Debnam, K. J., & Johnson, S. L. (2015). A focus on implementation of positive behavioral interventions and supports (PBIS) in high schools: Associations with bullying and other indicators of school disorder. *School Psychology Review*, *44*(4), 480–498. https://doi.org/10.17105/spr-15-0105.1

Bradshaw, C. P., Pas, E. T., Debnam, K. J., & Johnson, S. L. (2021). A randomized controlled trial of MTSS-B in high schools: Improving classroom management to prevent EBDs. *Remedial and Special Education*, *42*(1), 44–59. https://doi.org/10.1177/0741932520966727

Breitenstein, S. M., Gross, D., Garvey, C. A., Hill, C., Fogg, L., & Resnick, B. (2010). Implementation fidelity in community-based interventions: Implementation fidelity. *Research in Nursing and Health*, *33*(2), 164–173. https://doi.org/10.1002/nur.20373

Bruhn, A. L., Estrapala, S., Rila, A., & Colbert, J. (2022). A case example of tier 1 PBIS implementation in a high school. *Preventing School Failure: Alternative Education for Children and Youth*, *66*(1), 66–76. https://doi.org/10.1080/1045988X.2021.1946463

Calais, J., & Green, M. (2022). The racial pandemic: Positive behavior intervention support as an asymptomatic carrier of racism. *Perspectives on Early Childhood Psychology and Education*, *6*(1) 107–130. https://doi.org/10.58948/2834-8257.1004

Center on Positive Behavioral Interventions and Supports. (2015). *SWPBS implementation blueprint*. Center on Positive Behavioral Interventions and Supports. http://www.pbis.org

Center on Positive Behavioral Interventions and Supports. (2023). *What is PBIS?* Center on Positive Behavioral Interventions and Supports. https://www.pbis.org/pbis/what-is-pbis

Center on Positive Behavioral Interventions and Supports. (2024). *SWPBS implementation blueprint (revised)*. Center on Positive Behavioral Interventions and Supports. https://www.pbis.org/resource/pbis-implementation-blueprint

Childs, K. E., Kincaid, D., George, H. P., & Gage, N. A. (2016). The relationship between school-wide implementation of positive behavior intervention and supports and student discipline outcomes. *Journal of Positive Behavior Interventions*, *18*(2), 89–99. https://doi.org/10.1177/1098300715590398

Childs, K. E., Kincaid, D., & Peshak George, H. (2010). A model for statewide evaluation of a universal positive behavior support initiative. *Journal of Positive Behavior Interventions*, *12*(4), 198–210. https://doi.org/10.1177/1098300709340699

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cohen, R. J., Swerdlik, M. E., & Phillips, S. M. (1996). *Psychological testing and assessment: An introduction to tests and measurement* (3rd ed.). Mayfield Publishing Co.

Cohen, R., Kincaid, D., & Childs, K. E. (2007). Measuring school-wide positive behavior support implementation: Development and validation of the benchmarks of quality. *Journal of Positive Behavior Interventions*, *9*(4), 203–213. https://doi.org/10.1177/10983007070090040301

Colvin, G., Kameenui, E. J., & Sugai, G. (1993). Reconceptualizing behavior management and school-wide discipline in general education. *Education and Treatment of Children*, *16*(4), 361–381. https://www.jstor.org/stable/42899326

Cook, B. G., Buysse, V., Klingner, J., Landrum, T. J., McWilliam, R. A., Tankersley, M., & Test, D. W. (2015). CEC's standards for classifying the evidence base of practices in special education. *Remedial and Special Education*, *36*(4), 220–234. https://doi.org/10.1177/0741932514557271

Cook, C. R., Kilgus, S. P., & Burns, M. K. (2018). Advancing the science and practice of precision education to enhance student outcomes. *Journal of School Psychology*, *66*, 4–10. https://doi.org/10.1016/j.jsp.2017.11.004

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to poisson regression and its alternatives. *Journal of Personality Assessment*, *91*(2), 121–136. https://doi.org/10.1080/00223890802634175

Council for Exceptional Children. (2014). *Council for Exceptional Children standards for evidence-based practices in special education*. https://exceptionalchildren.org/sites/default/files/2021-04/EBP_FINAL.pdf

Dunlap, G., Sailor, W., Horner, R. H., & Sugai, G. (2009). Overview and history of positive behavior support. In W. Sailor, G. Dunlap, G. Sugai, & R. H. Horner. (Eds.) *Handbook of Positive Behavior Support. Issues in Clinical Child Psychology* (pp. 3-16). Springer, Boston, MA. https://doi.org/10.1007/978-0-387-09632-2_1

Eccles, M. P., & Mittman, B. S. (2006). Welcome to implementation science. *Implementation Science*, *1*(1), 1-3. https://doi.org/10.1186/1748-5908-1-1

Edmondson, M. J., Luo, C., Nazmul Islam, Md., Sheils, N. E., Buresh, J., Chen, Z., Bian, J., & Chen, Y. (2022). Distributed quasi-poisson regression algorithm for modeling multi-site count outcomes in distributed data networks. *Journal of Biomedical Informatics*, *131*, 2-13. https://doi.org/10.1016/j.jbi.2022.104097

Elrod, B. G., Rice, K. G., & Meyers, J. (2021). PBIS fidelity, school climate, and student discipline: A longitudinal study of secondary schools. *Psychology in the Schools*, *59*(2), 376–397. https://doi.org/10.1002/pits.22614

Enders, C. K., Du, H., & Keller, B. T. (2020). A model-based imputation procedure for multilevel regression models with random coefficients, interaction effects, and nonlinear terms. *Psychological Methods*, *25*(1), 88-112. https://doi.org/10.1037/met0000228

Estrapala, S., Rila, A., & Bruhn, A. L. (2021). A systematic review of tier 1 PBIS implementation in high schools. *Journal of Positive Behavior Interventions*, *23*(4), 288–302. https://doi.org/10.1177/1098300720929684

Fallon, L. M., Veiga, M., & Sugai, G. (2021). Strengthening MTSS for behavior (MTSS-B) to promote racial equity. *School Psychology Review*, *52*(5), 518-533. https://doi.org/10.1080/2372966X.2021.1972333

Filter, K. J., Johnson, L. D., Ford, A. L. B., Sowle, C. A., Bullard, S. J., Cook, C. R., Kloos, E., & Dupuis, D. (2022). An expert consensus process to distill tier 1 PBIS into core practice elements essential to frontline implementation. *Education and Treatment of Children*, *45*(1), 51–67. https://doi.org/10.1007/s43494-021-00066-y

Fixsen, D. L., Blase, K. A., Duda, M. A., Naoom, S. F., & Van Dyke, M. (2010). Implementation of evidence-based treatments for children and adolescents: Research findings and their implications for the future. In J. R. Weisz & A. E. Kazdin (Eds.), *Evidence-based psychotherapies for children and adolescents* (2nd ed., pp. 435–450). The Guilford Press.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. The National Implementation Research Network.

Flannery, K. B., Fenning, P., McGrath Kato, M., & Bohanon, H. (2013). A descriptive study of office disciplinary referrals in high schools. *Journal of Emotional and Behavioral Disorders*, *21*(2), 138–149. https://doi.org/10.1177/1063426611419512

Flannery, K. B., Fenning, P., McGrath Kato, M., & McIntosh, K. (2014). Effects of school-wide positive behavioral interventions and supports and fidelity of implementation on problem behavior in high schools. *School Psychology Quarterly*, *29*(2), 111–124. https://doi.org/10.1037/spq0000039

Flannery, K. B., Hershfeldt, P., & Freeman, J. (2018). *Lessons learned on implementation of PBIS in high schools: Current trends and future directions.* Center for Positive Behavioral Interventions and Supports. https://www.pbis.org/resource/lessons-learned-on-implementation-of-pbis-in-high-schools-current-trends-and-future-directions

Flannery, K. B., & McGrath Kato, M. (2017). Implementation of SWPBIS in high school: Why is it different? *Preventing School Failure: Alternative Education for Children and Youth*, *61*(1), 69–79. https://doi.org/10.1080/1045988X.2016.1196644

Flannery, K. B., Sugai, G., & Anderson, C. M. (2009). School-wide positive behavior support in high school: Early lessons learned. *Journal of Positive Behavior Interventions*, *11*(3), 177–185. https://doi.org/10.1177/1098300708316257

Freeman, J., Kern, L., Gambino, A. J., Lombardi, A., & Kowitt, J. (2019). Assessing the relationship between the positive behavior interventions and supports framework and student outcomes in high schools. *Journal of At-Risk Issues*, *22*(2), 1–11. https://eric.ed.gov/?id=EJ1231342

Freeman, J., Simonsen, B., McCoach, D. B., Sugai, G., Lombardi, A., & Horner, R. (2015). An analysis of the relationship between implementation of school-wide positive behavior interventions and supports and high school dropout rates. *The High School Journal*, *98*(4), 290-315. https://www.jstor.org/stable/44077793

Freeman, J., Simonsen, B., McCoach, D. B., Sugai, G., Lombardi, A., & Horner, R. (2016). Relationship between school-wide positive behavior interventions and supports and academic, attendance, and behavior outcomes in high schools. *Journal of Positive Behavior Interventions*, *18*(1), 41-51. https://doi.org/10.1177/1098300715580992

Freeman, J., Wilkinson, S., & Vanlone, J., (2017). *Status of High School PBIS Implementation in the U.S.* Center on Positive Behavioral Interventions and Supports. https://www.pbis.org/resource/status-of-high-school-pbis-implementation-in-the-u-s

Gage, N. A., Leite, W., Childs, K., & Kincaid, D. (2017). Average treatment effect of school-wide positive behavioral interventions and supports on school-level academic achievement in Florida. *Journal of Positive Behavior Interventions*, *19*(3), 158–167. https://doi.org/10.1177/1098300717693556

Gage, N. A., Whitford, D. K., & Katsiyannis, A. (2018). A review of schoolwide positive behavior interventions and supports as a framework for reducing disciplinary exclusions. *The Journal of Special Education*, *52*(3), 142–151. https://doi.org/10.1177/0022466918767847

Gagnon, D. R., Doron-LaMarca, S., Bell, M., O'Farrell, T. J., & Taft, C. T. (2008). Poisson regression for modeling count and frequency outcomes in trauma research. *Journal of Traumatic Stress*, *21*(5), 448–454. https://doi.org/10.1002/jts.20359

Garbacz, S. A., Santiago, R. T., & Gulbrandson, K. (2022). Examining implementation of positive behavioral interventions and supports in rural schools. *Rural Special Education Quarterly*, *41*(2), 95–107. https://doi.org/10.1177/87568705221097117

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological Bulletin*, *118*(3), 392-404. https://doi.org/10.1037/0033-2909.118.3.392

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), 392–404. https://doi.org/10.1037/spq0000178

Gottfredson, D. C., & Gottfredson, G. D. (2002). Quality of school-based prevention programs: Results from a national survey. *Journal of Research in Crime and Delinquency*, *39*(1), 3–35. https://doi.org/10.1177/002242780203900101

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? some practical clarifications of multiple imputation theory. *Prevention Science*, *8*, 206–213. https://doi.org/10.1007/s11121-007-0070-9

Grasley-Boy, N. M., Gage, N. A., & Lombardo, M. (2019). Effect of SWPBIS on disciplinary exclusions for students with and without disabilities. *Exceptional Children*, *86*(1), 25–39. https://doi.org/10.1177/0014402919854196

Grasley-Boy, N. M., Gage, N. A., Lombardo, M., & Anderson, L. (2023). Re-examining the reliability and validity of the SWPBIS tiered fidelity inventory. *Journal of Positive Behavior Interventions, 25*(4), 215-226. https://doi.org/10.1177/10983007231167606

Grasley-Boy, N. M., Reichow, B., van Dijk, W., & Gage, N. (2021). A systematic review of tier 1 PBIS implementation in alternative education settings. *Behavioral Disorders*, *46*(4), 199–213. https://doi.org/10.1177/0198742920915648

Gregory, A., Skiba, R. J., & Mediratta, K. (2017). Eliminating disparities in school discipline: A framework for intervention. *Review of Research in Education*, *41*(1), 253–278. https://doi.org/10.3102/0091732X17690499

Horner, R. H., Sugai, G., & Anderson, C. M. (2010). Examining the evidence base for school-wide positive behavior support. *Focus on Exceptional Children*, *42*(8), 1–14. https://doi.org/10.17161/fec.v42i8.6906

Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A. W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions*, *11*(3), 133–144. https://doi.org/10.1177/1098300709332067

Horner, R. H., Todd, A. W., Lewis-Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. B. (2004). The school-wide evaluation tool (SET) a research instrument for assessing school-wide positive behavior support. *Journal of Positive Behavior Interventions*, *6*(1), 3–12. https://doi.org/10.1177/10983007040060010201

Hyuk Lee, J., & Huber Jr., J. C. (2021). Evaluation of multiple imputation with large proportions of missing data: How much is too much? *Iranian Journal of Public Health, 50*(7), 1372-1380. https://doi.org/10.18502/ijph.v50i7.6626

Individuals with Disabilities Education Act of 2004, Pub. L. No. 108-446, 118 Stat. 2647. (2004). https://www.congress.gov/bill/108th-congress/house-bill/1350

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. https://doi.org/10.1177/10983007060080010301

James, A. G., Noltemeyer, A., Ritchie, R., Palmer, K., & University, M. (2019). Longitudinal disciplinary and achievement outcomes associated with school-wide PBIS implementation level. *Psychology in the Schools*, *56*(9), 1512–1521. https://doi.org/10.1002/pits.22282

Kim, J., McIntosh, K., Mercer, S. H., & Nese, R. N. T. (2018). Longitudinal associations between SWPBIS fidelity of implementation and behavior and academic outcomes. *Behavioral Disorders*, *43*(3), 357–369. https://doi.org/10.1177/0198742917747589

Kincaid, D., Childs, K., & George, H. (2010). School-wide benchmarks of quality (BoQ). https://www.pbis.org/resource/boq

Kittelman, A., Eliason, B. M., Dickey, C. R., & McIntosh, K. (2018). How are schools using the SWPBIS Tiered Fidelity Inventory (TFI). *Center on Positive Behavioral Interventions and Supports*. https://www.pbis.org/resource/how-are-schools-using-the-swpbis-tiered-fidelity-inventory-tfi

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, *49*(4), 241–253. https://doi.org/10.3102/0013189X20912798

Lane, K. L., Wehby, J. H., Robertson, E. J., & Ann Rogers, L. (2007). How do different types of high school students respond to schoolwide positive behavior support programs? Characteristics and responsiveness of teacher-identified students. *Journal of Emotional and Behavioral Disorders*, *15*(1), 3–20. https://doi.org/10.1177/10634266070150010201

Lee, A., & Gage, N. A. (2020). Updating and expanding systematic reviews and meta-analyses on the effects of school-wide positive behavior interventions and supports. *Psychology in the Schools*, *57*(5), 783–804. https://doi.org/10.1002/pits.22336

Lewis, T. J., Sugai, G., & Colvin, G. (1998). Reducing problem behavior through a school-wide system of effective behavioral support: investigation of a school-wide social skills training program and contextual interventions. *School Psychology Review*, *27*(3), 446–459. https://doi.org/10.1080/02796015.1998.12085929

Losen, D. J., & Whitaker, A. (2017). *Lost instruction: The disparate impact of the school discipline gap in California.* Civil Rights Project-Proyecto Derechos Civiles. https://www.civilrightsproject.ucla.edu/resources/projects/center-for-civil-rightsremedies/school-to-prison-folder/summary-reports/lost-instruction-the-disparate-impact-of-the-school-discipline-gap-in-california

Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Sankhya A*, *80*(1), 1–7. https://doi.org/10.1007/s13171-019-00164-5

Malloy, J. M., Bohanon, H., & Francoeur, K. (2018). Positive behavioral interventions and supports in high schools: A case study from New Hampshire. *Journal of Educational and Psychological Consultation*, *28*(2), 219–247. https://doi.org/10.1080/10474412.2017.1385398

Massar, M. M., McIntosh, K., & Mercer, S. H. (2019). Factor validation of a fidelity of implementation measure for social behavior systems. *Remedial and Special Education*, *40*(1), 16–24. https://doi.org/10.1177/0741932517736515

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. https://doi.org/10.1080/02796015.2010.12087759

McIntosh, K., Gion, C., & Bastable, E. (2018). *Do schools implementing SWPBIS have decreased racial and ethnic disproportionality in school discipline?* Center on Positive Behavioral Interventions and Supports. https://www.pbis.org/resource/do-schools-implementing-swpbis-have-decreased-racial-and-ethnic-disproportionality-in-school-discipline

McIntosh, K., Girvan, E. J., Fairbanks Falcon, S., McDaniel, S. C., Smolkowski, K., Bastable, E., Santiago-Rosario, M. R., Izzard, S., Austin, S. C., Nese, R. N. T., & Baldy, T. S. (2021). Equity-focused PBIS approach reduces racial inequities in school discipline: A randomized controlled trial. *School Psychology*, *36*(6), 433–444. https://doi.org/10.1037/spq0000466

McIntosh, K., Massar, M. M., Algozzine, R. F., George, H. P., Horner, R. H., Lewis, T. J., & Swain-Bradway, J. (2017). Technical adequacy of the SWPBIS tiered fidelity inventory. *Journal of Positive Behavior Interventions*, *19*(1), 3–13. https://doi.org/10.1177/1098300716637193

McIntosh, K., Mercer, S. H., Nese, R. N. T., Strickland-Cohen, M. K., Kittelman, A., Hoselton, R., & Horner, R. H. (2018). Factors predicting sustained implementation of a universal behavior support framework. *Educational Researcher*, *47*(5), 307–316. https://doi.org/10.3102/0013189X18776975

Mercer, S. H., McIntosh, K., & Hoselton, R. (2017). Comparability of fidelity measures for assessing tier 1 school-wide positive behavioral interventions and supports. *Journal of Positive Behavior Interventions*, *19*(4), 195–204. https://doi.org/10.1177/1098300717693384

Molloy, L. E., Moore, J. E., Trail, J., Van Epps, J. J., & Hopfer, S. (2013). Understanding real-world implementation quality and "active ingredients" of PBIS. *Prevention Science*, *14*(6), 593–605. https://doi.org/10.1007/s11121-012-0343-9

Muscott, H. S., Mann, E., Benjamin, T. B., Gately, S., Bell, K. E., & Muscott, A. J. (2004). Positive behavioral interventions and supports in New Hampshire: Preliminary results of a statewide system for implementing schoolwide discipline practices. *Education and Treatment of Children*, *27*(4), 453–475. https://www.jstor.org/stable/42899817

Muscott, H. S., Mann, E. L., & LeBrun, M. R. (2008). Positive behavioral interventions and supports in New Hampshire: Effects of large-scale implementation of schoolwide positive behavior support on student discipline and academic achievement. *Journal of Positive Behavior Interventions*, *10*(3), 190–205. https://doi.org/10.1177/1098300708316258

Noltemeyer, A., Palmer, K., James, A. G., & Petrasek, M. (2019). Disciplinary and achievement outcomes associated with school-wide positive behavioral interventions and supports implementation level. *School Psychology Review*, *48*(1), 81–87. https://doi.org/10.17105/SPR-2017-0131.V48-1

Noltemeyer, A., Palmer, K., James, A. G., & Wiechman, S. (2019). School-wide positive behavioral interventions and supports (SWPBIS): A synthesis of existing research. *International Journal of School and Educational Psychology*, *7*(4), 253–262. https://doi.org/10.1080/21683603.2018.1425169

Ozili, P. K. (2023). The acceptable R-square in empirical modelling for social science research. In *Social research methodology and publishing results: A guide to non-native english speakers* (pp. 134–143). IGI global. https://doi.org/10.4018/978-1-6684-6859-3.ch009

Pas, E. T., Johnson, S. R., Debnam, K. J., Hulleman, C. S., & Bradshaw, C. P. (2019). Examining the relative utility of PBIS implementation fidelity scores in relation to student outcomes. *Remedial and Special Education*, *40*(1), 6–15. https://doi.org/10.1177/0741932518805192

PBISApps. (2023). *SWIS Data Summary 2022-23*. PBISApps. https://www.pbisapps.org/resource/2022-2023-swis-summary

PBISApps (2022, July 15). *PBISApps to retire the SET, ISSET, TIC, and SSS.* PBISApps. https://www.pbisapps.org/articles/pbisapps-to-retire-the-set-isset-tic-and-sss

PBISApps (2019). *SWIS user manual*. PBISApps. https://www.pbisapps.org/resource/swis-user-manual

RStudio Team (2023). RStudio: Integrated Development Environment for R (Version 2023.09.1+494) [Computer software]. RStudio, PBC. http://www.rstudio.com/

Roback, P., Legler, J. (2021). *Beyond multiple linear regression: Applied generalized linear models and multilevel models in R*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429066665

Skiba, R. J., Arredondo, M. I., & Williams, N. T. (2014). More than a metaphor: The contribution of exclusionary discipline to a school-to-prison pipeline. *Equity and Excellence in Education*, *47*(4), 546–564. https://doi.org/10.1080/10665684.2014.958965

Smolkowski, K., Strycker, L., & Ward, B. (2016). Scale-up of safe & civil schools' model for school-wide positive behavioral interventions and supports. *Psychology in the Schools*, *53*(4), 339–358. https://doi.org/10.1002/pits.21908

Solomon, B. G., Tobin, K. G., & Schutte, G. M. (2015). Examining the reliability and validity of the effective behavior support self-assessment survey. *Education and Treatment of Children*, *38*(2), 175–191. https://doi.org/10.1353/etc.2015.0007

Sugai, G., & Horner, R. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Child & Family Behavior Therapy*, *24*(1-2), 23–50. https://doi.org/10.1300/J019v24n01_03

Sugai, G., & Horner, R. (2006). A promising approach for expanding and sustaining school-wide positive behavior support. *School Psychology Review*, *35*(2), 245–259. https://doi.org/10.1080/02796015.2006.12087989

Sugai, G., & Horner, R. (2020). Sustaining and scaling positive behavioral interventions and supports: Implementation drivers, outcomes, and considerations. *Exceptional Children*, *86*(2), 120–136. https://doi.org/10.1177/0014402919855331

Sugai, G., Horner, R., & Lewis-Palmer, T. (2001). *Team Implementation Checklist (TIC).* Center on Positive Behavioral Interventions and Supports. https://www.pbis.org/resource/pbis-team-implementation-checklist-tic-3-1

Sugai, G., Horner, R., & Todd, A. (2000). *EBS self-assessment survey (SAS).* Center on Positive Behavioral Interventions and Supports. https://www.pbis.org/resource/sas

Sugai, G., Lewis-Palmer, T., Todd, A., & Horner, R. (2001). *School-wide evaluation tool (SET)*. Center on Positive Behavioral Interventions and Supports. https://www.pbis.org/resource/school-wide-evaluation-tool-set

Sugai, G., Sprague, J. R., Horner, R. H., & Walker, H. M. (2000). Preventing school violence: The use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of Emotional and Behavioral Disorders*, *8*(2), 94–101. https://doi.org/10.1177/106342660000800205

Sullivan, A. L., Harris, B., Miller, F. G., Fallon, L. M., Weeks, M. R., Malone, C. M., Kulkarni, T., Proctor, S. L., Johnson, A. H., Rossen, E., Nguyen, T., & Shaver, E. (2021). A call to action for school psychology to address COVID-19 health disparities and advance social justice. *School Psychology*, *36*(5), 410–421. https://doi.org/10.1037/spq0000463

Sullivan, A. L., Miller, F. G., McKevett, N. M., Muldrew, A., Hansen-Burke, A., & Weeks, M. (2020). Leveraging MTSS to advance, not suppress, COVID-related equity issues. *Communique*, *49*(1), 1–26. https://www.nasponline.org/publications/periodicals/communique/issues/volume-49-issue-1/leveraging-mtss-to-advance-not-suppress-covid-related-equity-issues

Sullivan, A. L., Weeks, M., Miller, F. G., Nguyen, T., Kulkarni, T., Williams, S., & Kim, J. (2024). No "top of the triangle kids": Toward conceptual clarity of students, behavior, and tiers in MTSS to advance social justice. *Journal of School Psychology*, *106*. https://doi.org/10.1016/j.jsp.2024.101325

Tobin, T., Vincent, C., Horner, R., Dickey, C. R., & May, S. (2012). Fidelity measures to improve implementation of positive behavioural support. *International Journal of Positive Behavioural Support*, *2*(2), 12–19.

Turnbull, H. R., Wilcox, B. L., Stowe, M., Raper, C., & Hedges, L. P. (2000). Public policy foundations for positive behavioral interventions, strategies, and supports. *Journal of Positive Behavior Interventions*, *2*(4), 218–230. https://doi.org/10.1177/109830070000200406

U.S. Department of Education (2023). Student discipline and school climate in U.S. public schools. https://www2.ed.gov/about/offices/list/ocr/docs/crdc-discipline-school-climate-report.pdf

Vincent, C. G., & Tobin, T. J. (2011). The relationship between implementation of school-wide positive behavior support (SWPBS) and disciplinary exclusion of students from various ethnic backgrounds with and without disabilities. *Journal of Emotional and Behavioral Disorders*, *19*(4), 217–232. https://doi.org/10.1177/1063426610377329

Von Hippel, P. T. (2020). How many imputations do you need? A two-stage calculation using a quadratic rule. *Sociological Methods and Research*, *49*(3), 699–718. https://doi.org/10.1177/0049124117747303

Walker, H. M., Horner, R. H., Sugai, G., Bullis, M., Sprague, J. R., Bricker, D., & Kaufman, M. J. (1996). Integrated approaches to preventing antisocial behavior patterns among school-age children and youth. *Journal of Emotional and Behavioral Disorders, 4*(4), 194–209. https://doi.org/10.1177/106342669600400401

Weeks, M. R., & Sullivan, A. L. (2024). Systematic review of the associations of SWPBS with exclusionary discipline and disproportionality in U.S. schools. *Journal of School Psychology*, *106*. https://doi.org/10.1016/j.jsp.2024.101327

Weick, K. E. (1984). Small wins: Redefining the scale of social problems. *American Psychologist*, *39*(1), 40–49. https://doi.org/10.1037/0003-066X.39.1.40

What Works Clearinghouse. (2019). *What works clearinghouse procedures and standards handbook, version 4.0.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). https://ies.ed.gov/ncee/wwc/Handbooks

Woods, A. D., Gerasimova, D., Van Dusen, B., Nissen, J., Bainter, S., Uzdavines, A., Davis-Kean, P. E., Halvorson, M., King, K. M., & Logan, J. A. (2021). Best practices for addressing missing data through multiple imputation. *Infant and Child Development*, *33*(1), 1-37. https://doi.org/10.1002/icd.2407