Dissertations   Theses and Dissertations

2021

Disrupting the School-To-Prison Pipeline: A Mixed-Methods Systematic Review of Alternative Discipline Practices to Reduce Exclusion and Promote Equity

Cara Diclemente

Recommended Citation

LOYOLA UNIVERSITY CHICAGO

DISRUPTING THE SCHOOL-TO-PRISON PIPELINE:

A MIXED-METHODS SYSTEMATIC REVIEW OF ALTERNATIVE DISCIPLINE

PRACTICES TO REDUCE EXCLUSION AND PROMOTE EQUITY

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE GRADUATE SCHOOL

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

PROGRAM IN CLINICAL PSYCHOLOGY

BY

CARA M DICLEMENTE

CHICAGO, IL

AUGUST 2021

ACKNOWLEDGMENTS

To my partner, Matt, my sister, Anna, my mother, Lyn, and my father, Carlo: I would never have gotten to this point in my graduate career without you. You have been a calming presence amidst a very busy several years. I cannot express how much I have appreciated all the love and support you have provided me along the road and the belief you have had in my abilities from the moment I was accepted at Loyola University Chicago.

Dedicated to all of the educators, students, and families who have persevered and supported each other amidst an incredibly difficult school year.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Zero tolerance policies were designed to create safety by implementing automatic exclusion (e.g., suspensions, expulsions) for misbehavior in response to rising school violence in the United States. However, evidence over the past four decades shows that these policies fail to increase safety, and instead foster poor school climate and disproportionate rates of minority groups in the school-to-prison pipeline. Previous research and literature reviews suggest there are a host of developing Multi-Tiered System of Supports (MTSS) practices that have promising potential to reduce exclusionary outcomes and foster equitable treatment of vulnerable student populations, such as Positive Behavioral Intervention Supports and Restorative Practices.

This mixed methods systematic review collates available data regarding the effectiveness of various MTSS interventions as alternative discipline practices that intend to replace or reduce exclusionary outcomes in schools. This review identifies both the quantitative outcomes for each universal, secondary, and tertiary non-exclusionary disciplinary intervention, as well as the qualitative processes and perspectives associated with each MTSS strategy, such as subjective impact, acceptability, feasibility, and implementation barriers and facilitators across multiple types of stakeholders. Studies included in this systematic review were limited to peer-reviewed research in United States general education K-12 settings published in 1997 and later in order to capture the most relevant, high-quality work on this topic.

This systematic review incorporates school-level, staff-level, and student-level effects and perspectives across multiple types of study design, therefore highlighting prevalent themes across various practices so that these strategies can be more easily replicated in under-resourced settings with access to less support. In order to ensure more unbiased assessment of the evidence, the quality of each study's reporting, methodology, and evidence are discussed to frame how strongly these findings should be taken into consideration when drawing conclusions. Each of the 59 studies is discussed in terms of its design, setting, population, intervention focus and MTSS tier, methods, and outcomes, followed by an overview of the shared characteristics and key findings across studies.

While conclusions are limited by the available study designs and comparisons between interventions, recommendations are provided for administrators, educations, and practitioners in schools moving forward, as well as directions for future research on what we still need to know about alternative discipline practices to reduce exclusion and promote equity. Key themes for future implementers include: promoting school-wide culture change through relationship-building; using behavioral strategies to incentivize positive behaviors; implementing brief teacher-focused trainings to enhance empathy, cultural responsiveness, and awareness of implicit bias; offering trauma-informed trainings to enhance contextual awareness; providing ongoing professional development to prevent "initiative fatigue"; selecting core common practices across interventions to simplify implementation and reduce burden; tailoring interventions to fit students' developmental level; and involving stakeholders in development, implementation, and evaluation of these interventions. Implications for implementation science and disability critical race theory are also discussed.

CHAPTER ONE

INTRODUCTION

Adolescents spend the majority of their time in school, making this context incredibly influential for adolescent development and well-being (Brookmeyer, Fanti, and Henrich, 2006). Especially in low-income communities with economic and social constraints, the school serves as the primary source of consistent intervention for physical, psychological, and academic needs, and is a major determinant of Black youth trajectories (American Psychological Association (APA), 2008a). Unfortunately, schools can also be extremely harmful when the environment is not safe or supportive. Zero tolerance policies were designed to create safety by implementing automatic exclusion for misbehavior in response to rising school violence in the United States. However, evidence over the past four decades shows that these policies fail to increase safety, and instead foster poor school climate (i.e., the quality and character of school life) and detriments to other psychological and academic variables that predict future success. In essence, students are removed from classroom instruction without support for future behavior change and socioemotional learning. Even more detrimental, the criminalization of adolescent behaviors has exacerbated the school-to-prison pipeline, especially for Black adolescents who receive the most disproportionate rates of exclusionary discipline compared to all other ethnicities (Okonofua et al., 2016). In segregated communities plagued by decades of systemic racism, discipline practices are more likely to be biased and harsh (Riddle & Sinclair, 2019; Skiba et al., 2011).

School districts across the US have made concerted efforts to decrease exclusionary discipline practices in order to reduce the school-to-prison pipeline, but it is still unclear as to which alternative practices have successfully replaced exclusion, when exclusion or police notification may still be warranted, or when these methods might be paired together. *The current systematic review will fill these gaps by summarizing the current research evaluating the implementation of alternative, non-exclusionary disciplinary strategies, such as Restorative Justice, School-Wide Positive Behavioral Interventions and Supports, and other interventions that aim to supplant punitive practices.* This review will serve to identify key disciplinary tools and supports that serve as best practices for ensuring school safety and equitable treatment of student misbehavior, while illuminating areas of school discipline research that still need to be developed. This research comes at a crucial time for reform, given the rescinding of national guidance towards reducing discipline disparities by former Secretary of Education Betsy DeVos, which can take years to undo (Vara-Orta, 2018).

Federal Zero Tolerance Policies in Schools

Created during the "war on drugs" efforts by state and federal governments in the mid-1980s, zero tolerance policies in the United States, such as the Drug Free Schools Act (DFSA) in 1986, were initially intended to enforce strict penalties for drug use (Skiba & Rausch, 2006). Over time, these penalties were applied to a wider range of offenses in accordance with the "Broken Windows" theory of crime, which argues for tough action on minor offenses (e.g., broken windows) in order to prevent more serious crimes (e.g., fires set by squatters) (Kelling & Coles, 1997). In schools, the rise in juvenile arrests for violent crimes in the 1980s (Hockenberry & Puzzanchera, 2014) and increased shootings and violence in schools in the 1990s reinforced

the use of zero tolerance policies in an effort to make schools safer amidst disproportionate concerns relative to the actual safety risks (Heitzeg, 2014). Specifically, the Gun-Free Schools Act (GFSA) in 1994 required schools receiving federal funding to expel those students who brought a firearm to school for at least one year and refer them to the justice system (Martinez, 2009). After the highly publicized Columbine shooting in 1999, fear of youth violence was at an all-time high and schools became even more focused on maintaining security, with amendments added for other weapon types (Fuentes, 2014).

By definition, zero tolerance policies designate automatic and often punitive consequences for drug or weapon offenses, regardless of the severity of or circumstances surrounding the offense, in order to maintain a safe educational climate (Skiba & Rausch, 2006). Over the past three decades, school administrations have increasingly resorted to automatic punishment for rule-breaking and disruptive behaviors not identified in the DFSA or GFSA, even for minor or nonviolent offenses such as fighting, obscene language, damage to school property, tardiness, truancy, disobedience, or disrespect (Gregory et al., 2010; Skiba, Chung, et al., 2014; Marsh, 2014). This meant that more students received office referrals, suspensions, expulsions, and arrests than ever before, in turn removing these children from classroom instruction, even for first-time offenses (Skiba, 2000; Kang-Brown et al., 2013), and subsequently criminalizing normative adolescent behaviors (Jones, 2013). In the 2015-2016 school year, nearly 2.7 million public school students (5.3%) received a suspension, with 291,000 referred to law enforcement or arrested in school (Office for Civil Rights, 2018). This high rate of suspensions and arrests runs counter to the overall decline in juvenile crime and delinquency rates over the past two decades (U.S. Department of Education, 2014); especially since violent crime is more likely to

occur off rather than on school grounds (Justice Policy Institute, 2011). The accumulation of various zero tolerance policies over time has led to more hardship than success for students.

Negative Effects of Exclusionary Practices

While it was originally assumed that applying zero tolerance policies would promote greater school safety for undisrupted learning, the resulting increase of exclusionary practices (e.g., detention, suspension, expulsion, school-related arrest) has been ineffective in deterring disruptive behavior or violence (Gregory et al., 2010; Curran, 2016). The APA Zero Tolerance Task Force (2008b) found that these policies do not enhance consistency of school discipline practices and actually increase the amount of time spent on discipline matters (Sharkey & Fenning, 2012). In examining student reports, many researchers have shown that students regard suspensions and expulsions as unfair and do not use this exclusionary time to work on changing their behavior (APA, 2008b). Instead, the removed student loses formal learning time, which is correlated with their academic achievement (Gregory et al., 2010). Developmentally, it is normative for adolescents to take risks and disregard consequences, but automatic punishment does not take this into account nor uncover the root causes of disruptive behaviors and provide the support that adolescents need to learn from their mistakes (APA, 2008b; Morrison & Vaandering, 2012).

Even worse, exclusion can actually increase misbehavior and make schools less safe (APA, 2008b; Carter et al., 2014), especially when there are no interventions to promote behavior change and growth (Morrison & Vaandering, 2012). Several studies have shown notable pathways between exclusionary punishments and reduced school cohesion, sense of belonging, perceived sense of safety, satisfaction with school administration, and academic

achievement (e.g., repeating a grade) (Gregory et al., 2010; Skiba, Arrendonda, et al., 2014; APA, 2008b; McNeal & Dunbar, 2010), with graduation rates dropping significantly after just one suspension (Losen et al., 2016; Rumberger & Losen, 2016). In a study of Florida high school students, those who were suspended only once in 9th grade were subsequently twice as likely to drop out by the end of high school (Balfanz et al., 2012). In a qualitative study, both educator-youth and family-school relationships were damaged by suspensions, underscoring the need for sensitive and caring educators and fewer suspensions (Haight et al., 2014).

Additionally, School Resource Officers (SROs), city law enforcement officers responsible for safety and crime prevention in schools, have become an additional tool to implement and increase exclusion. The Safe Schools Act of 1994 and 1998 provided explicit funding for SROs in response to school shootings (Mallett, 2016). Past and recent mass shootings, such as Columbine in 1999 and Stoneman Douglas High School in 2018, have maintained the demand for police in schools, with $300 million spent on school policing since 1995 (American Civil Liberties Union, ACLU, 2017), $1 billion added to state budgets within six months of Stoneman Douglas (ACLU, 2019), and House lawmakers calling for even more funding through the School Resource Officer Act in November 2018 (Ujifusa, 2018). Nationwide statistics have illuminated increases of up to 400% in arrests on school property since the start of zero tolerance policies (Torres & Stefkovich, 2009), often for non-serious offenses due to unclear standards for when SROs should intervene in discipline (Na & Gottfredson, 2013; Kurtz et al., 2018; Curran, 2019). Additionally, the hostility these figures promote are associated with greater student alienation, poorer school climate, and lower academic achievement (Deakin & Kupchik, 2018; Ripski & Gregory, 2009).

In stark contrast, a report from the ACLU (2019) found that 14 million students are in schools with police yet no mental health professionals, and 90% of schools with professionals failed to meet the minimum staff to student ratio (i.e., at least one counselor and one social worker for every 250 students) despite rising suicide rates. While calls for more police have heightened over time, state education spending per pupil has decreased (Leachman & Mai, 2014) and fiscal disparities across schools fail to be reconciled (Advancement Project, 2010). Other educational policy shifts, such as the No Child Left Behind law in 2001, aimed to help under-performing students achieve academic success, but unintentionally led to the exclusion of these students through expulsions or referrals to alternative programs, so that schools could produce higher average test scores (Heitzeg, 2014). From this body of research, it appears that while the intentions of zero tolerance policy creators were positive, the outcomes of these policies are overwhelmingly negative.

School to Prison Pipeline: Disproportionate Effects for Black and Brown Students

As a result of the criminalization of many behaviors ranging in severity, exclusionary discipline practices have been fueling the school to prison pipeline by increasing the likelihood that youth will enter the juvenile justice system (Martinez, 2009; Stinchcomb et al., 2006). In this way, infractions that were typically handled in school or would not be considered to be dangerous are instead referred to police (Wald & Losen, 2003). Once suspended or expelled, students have difficulty avoiding recidivism upon re-entry given the social isolation and academic struggles they experience (Losen et al., 2014). For example, in a study of six million Texas students, receiving a non-weapon school offense increased the chances of juvenile court involvement three-fold (Fabelo et al., 2011). Exclusion from school also promotes increased

access to unstructured and unmonitored activities, therefore increasing the probability of future arrests and criminal offending outside of school (Cuellar & Markowitz, 2015). Concerningly, once a youth is involved in the court system in any way, the probability of detention and incarceration surges (Petrosino et al., 2010).

This pipeline has been especially detrimental for racial and ethnic minorities and students from lower socioeconomic status (SES) environments through systematically unfair applications of zero tolerance policies. Extensive evidence shows heavily biased exclusionary punishments directed towards these students across the United States despite efforts to make discipline objective (Advancement Project, 2010; Petras et al., 2011), and these biases exist after accounting for differences in student misbehavior (Huang & Cornell, 2017; Skiba, Chung, et al., 2014). Trends in the literature suggest that schools most likely to be punitive and to have police officers are those that are large, urban, have larger Black student populations, and have more students receiving free or reduced price lunches (Marchbanks et al., 2018; Curran, 2019; Welsh & Little, 2018). Schools with more Black and low-income students are more likely to have tough security measures in place, harsher discipline practices, and more than one disciplinary action per situation; even when controlling for neighborhood violence and misconduct in schools (Nance, 2015; Skiba, Chung, et al., 2014; Welch & Payne, 2010). Family characteristics have also been found to predict suspensions, such as whether the student has access to books, quiet space, time for homework, and more than one caregiver (Welsh & Little, 2018). These findings together suggest that educators may struggle to adequately address the underlying inequities in development and resource depletion in these communities.

Statistically, while Black students account for 15% of the national student body, they received 31% of arrests in 2015-2016, which represents an increase from two years prior and mirrors disproportionate rates for students with disabilities as well (i.e., 12% of student body and 28% of referrals to law enforcement) (Office for Civil Rights, 2018). This overrepresentation occurs across all poverty levels (Government Accountability Office, 2018) and for both Black males and females (Center for Intersectionality and Social Policy Studies, 2015; Morris, 2012) despite lack of evidence that Black students are actually more violent or disruptive (Gastic, 2017; Losen, Martinez, & Gillespie, 2012). There is also evidence of longer punishments for Black students for the same offenses as White peers, even when controlling for poverty (Anderson, Ritter, & Zamarro, 2017; Balfanz, Byrnes & Fox, 2015). Within the juvenile justice system, Black and Latinx youth represent two thirds of the youth in residential incarceration facilities, while they only make up one third of the adolescent population (Hockenberry & Puzzanchera, 2014), and Black youth are six times more likely to be incarcerated than White youth (Piquero, 2008). While Black boys represent the highest overall rates of exclusion (Government Accountability Office, 2018), Black girls account for 31% of girls referred to law enforcement by school officials and 43% of those arrested on school grounds, but only make up 17% of the student population (National Women's Law Center & NAACP Legal Defense and Education Fund, 2014). Black girls have also been shown to be punished more harshly and for more subjective offenses (Annamma et al., 2019) and are four times more likely to be arrested than White girls (Office for Civil Rights, 2018).

In an effort to explain these disproportionate statistics, APA (2008b) reviewed evidence suggesting that administrators' and teachers' poor classroom management, reduced cultural

competence, and racial stereotypes are some of the factors making punishment increasingly biased. Black student offenses have been found to be classified in more subjective categories than White student offenses (Skiba et al., 2002). This subjectivity is especially likely to occur when punishment is given for vague categories such as "disobedience" and "defiance" (Skiba et al., 2011). For example, when teachers were presented with fictional case scenarios, students with stereotypical "Black" names were more likely to have their behavior perceived as severe, patterned, and worthy of future suspension (Okonofua & Eberhardt, 2015). Educators tend to perceive White youth as having psychiatric diagnoses that warrant treatment such as Attention-Deficit/Hyperactivity Disorder or behaving poorly due to external factors, whereas minority youth are less likely to receive such diagnoses and may instead be labeled as simply defiant, violent, or lacking in self-control (Safer & Malever, 2000; Coker et al., 2016). There is also a high mismatch between the demographic characteristics of the majority of United States teachers who are White middle-class women, and the much more diverse student body (Staats, 2014). These stereotypes and cultural differences have the potential to cause friction when school authority figures have their own conflicting perceptions of what constitutes normative behavior (Golann, 2015; Skiba et al., 2011), therefore fostering student mistrust in those authorities (Gregory & Weinstein, 2008).

Social factors also perpetuate unfair targeting of minority youth, such as the gradual re-segregation of schools since Brown v. Board of Education (Reardon et al., 2012), television and media portrayals of people of color as offenders rather than victims (Heitzeg, 2014), and the revenue gained by keeping the prison industrial complex full (Brewer & Heitzeg, 2008). Additionally, poverty, trauma, and mental health problems are all prevalent in urban minority

communities, and are associated with lower academic achievement and higher rates of developmental delays and delinquency (Mallett, 2017). While punishment disparities are not explained solely by poverty and delinquency rates, when schools in these communities lack resources to provide adequate education and culturally-relevant mental health services, youth are even more likely to receive punishment over treatment or socioemotional instruction (Kim et al., 2010; Fabelo et al., 2011; Gonzalez, 2012).

Multi-Tiered System of Supports

The proliferation of numerous types of alternative discipline solutions has stemmed from nationwide calls to reduce exclusionary discipline practices and eliminate disparities in disciplinary outcomes through new programs and policies, so that all children have a fair opportunity to learn (Steinberg & Lacoe, 2017; Jones, 2013). The U.S. Department of Education (2014) produced official guidelines for improving school discipline by using methods for enhancing school climate in particular, emphasizing the importance of this construct for addressing the negative effects of exclusionary punishment. To this end, the former Secretary of Education Arne Duncan under Obama argued for replacement of exclusion with prevention-based strategies to teach socioemotional skills to all students and match them with appropriate interventions when needed, therefore promoting the development of crucial competencies for success. Specifically, the DOE encouraged whole school and tiered approaches that pair safe and supportive school climate with clear, appropriate, and consistent expectations and consequences, which are associated with better safety (Gregory et al., 2012).

School climate is broadly defined as the quality and character of school life to promote youth development and learning, and typically centers on perceptions of acceptance, support, and

safety in the school environment (Cohen et al., 2009). Overlapping core components from the many operational definitions of school climate include beliefs and values, relationships and social interactions, safety, teaching and instruction, leadership, and physical environment (Rudasill et al., 2018). For this reason, school discipline approaches must address transactions that occur between and within each part of the school ecology (e.g., peers, family, classroom, school, community) using a systems view of school climate (Rudasill et al., 2018). For example, some research has shown that providing both high structure (e.g., firm limit-setting and monitoring to enforce these limits) and high emotional support (e.g., staff warmth, acceptance, and responsiveness) in an authoritative manner can foster school safety, order, and other positive school outcomes (Gregory et al., 2010). Overall, discipline strategies that are tailored across school contexts and student needs are more likely to be successful.

The framework that organizes these whole school approaches to safety and positive school climate is called Multi-Tiered System of Supports (MTSS; Figure 1). MTSS encompass a range of interventions for learning and behavior problems based on demonstrated levels of need (Batsche et al., 2005) and was established as part of the Elementary and Secondary Education/Every Student Succeeds Act (ESSA, 2015). While academic interventions as part of Response to Intervention (RTI) are included in the overall framework, only behavioral interventions are of importance for the current research. MTSS broadly include the use of proactive universal risk screening and prevention, evidence-based interventions to match to the student, and ongoing assessment to inform decision-making (Lane et al., 2014). MTSS models consist of three tiers of support, including primary interventions for the whole school population, secondary supports for individuals or groups of students with common needs, and tertiary

supports for individualized treatments (Stoiber & Gettinger, 2016). Most students exhibiting

challenging behaviors as a result of emotional disturbance are served in general education

settings (Lane et al., 2014), which emphasizes the need for both broad and tailored approaches

that are not specific to special education settings. A summary of the MTSS practices and their

applications across the three tiers of prevention and intervention are included in Table 1.

Figure 1. MTSS Framework

Indicated Prevention
Specialized individualized systems for students with high-risk behavior

Targeted Prevention
Specialized group systems for students with at-risk behavior

Universal Prevention
School-wide or classroom-wide systems for all students, staff and settings

*5% of Students

*15% of Students

*80% of Students
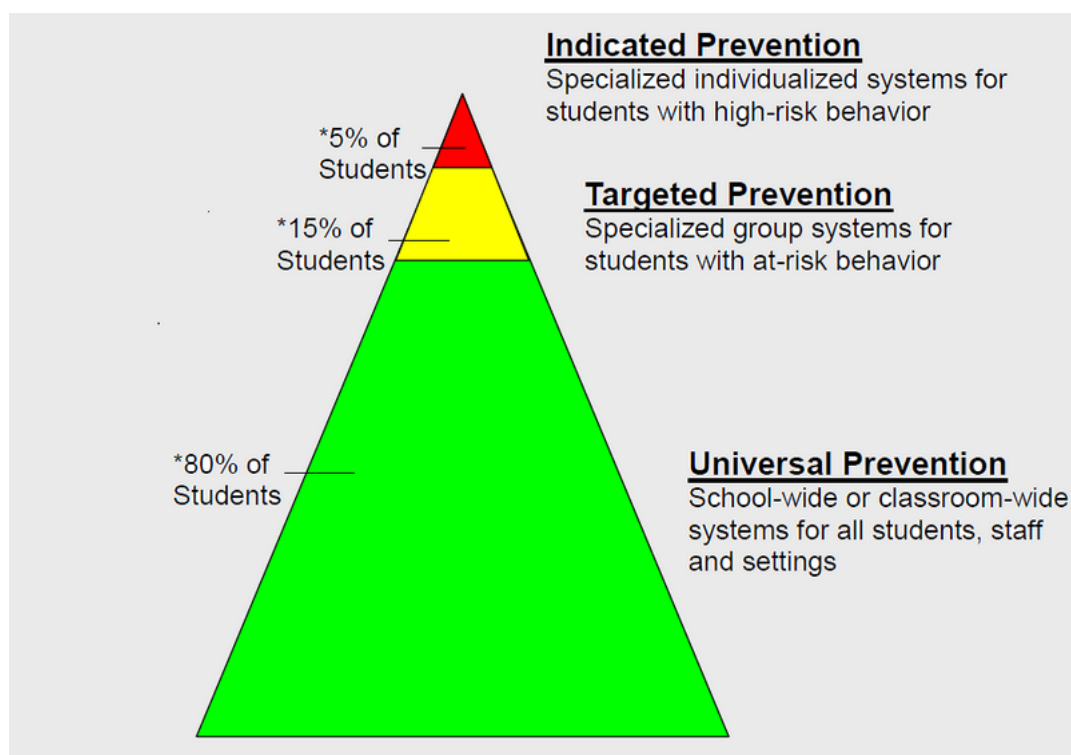
Primary Tier (Universal Supports)

The primary, or universal, interventions as part of MTSS are geared towards providing a

positive classroom environment and behavior support to all students in a population to prevent

disruptive behavior in about 80% of the population (Stoiber & Gettinger, 2016). These

interventions are considered to be proactive, such that they identify needs and address them before problems escalate (Sharkey & Fenning, 2012). These practices may include 3-5 positively stated expectations for behavior (e.g., "Be safe") set forth by the Positive Behavior Intervention Supports (PBIS) program that are posted around the school and reviewed at the classroom-level along with associated logical consequences if those expectations are not met (OSEP, 2021). Based on principles of operant conditioning and contingent reinforcement, PBIS (informally called the "carrot and stick approach") involves earning tangible and social rewards (e.g., candy, free periods, praise) through a token system for following school rules to positively reinforce and increase these behaviors in the future (Warren et al., 2006). This type of programming is most effective when at least 80% of school staff buy into the procedures for at least three years (Sugai & Horner, 2009), and school-family partnerships are encouraged to refine and make these practices culturally responsive. PBIS is currently utilized in over 25,000 schools and growing, making it a frequently implemented set of MTSS interventions.

Another potential focus of a school's primary tier of MTSS is on developing age-appropriate social-emotional competencies so that students feel cared for, motivated, and like they belong (Peterson et al., 2013). The Collaborative for Academic, Social and Emotional Learning (CASEL) defines social-emotional learning (SEL) as developing self-management skills (e.g., regulating emotions and impulses amidst stress, progress to achieve goals), self-awareness (e.g., recognizing emotions and strengths, maintaining self-confidence), social awareness (e.g., empathy, perspective-taking, resource awareness), relationship skills (e.g., cooperation, conflict management, resisting social pressure), and responsible decision-making (Durlak et al., 2011). In this way, the strengths-based instruction focuses on building youth's

internal capacity to approach distress and conflict productively so that they can focus on academic achievement and feel safe and connected. Activities can be reflective about the decisions students make (e.g., reflection sheets, role-playing, interviews), provide curriculum instruction on socioemotional skills and how to prevent problem behaviors such as bullying or substance use, or take the form of community-building practices (Durlak et al., 2011). In this tier, SEL skills are designed to become a part of daily life. Currently, 20 major school districts have partnered with CASEL since 2011 to provide SEL curricula.

Similar to SEL, Restorative Justice practices were identified by the U.S. Department of Education (2014) as a promising approach for improving safety without exclusion. Restorative Justice is an entire school philosophy centered around flexible problem-solving between all school members without any retribution, shame, or reliance on rewards (Zellerer, 2013). This philosophy encapsulates the African belief of *ubuntu*, which states that the individual is intrinsically bonded to all of humanity through a series of reciprocal relations (Davis, 2019). Restorative Justice began in the 1970s in the United States criminal justice sector as a way for victims and offenders to engage in productive mediation through reconciliation programs (Wadhwa, 2016). Similarly, educational Restorative Justice practices, broadly called Restorative Practices (RP) aim to nurture relationships between administrators, teachers, students, security staff, parents, and the community and provide a fair and equitable learning environment that emphasizes social engagement over social control (Evans & Vaandering, 2016).

At a universal level, RP utilizes teachers and staff as relational models for how to use SEL skills to handle conflict, which also involves several proactive restorative practices to build community, de-escalate, and collaboratively solve problems (Thorsborne & Blood, 2013). For

example, talking circles are implemented by passing around a talking piece to promote respectful

dialogue (Pranis, 2005), either at the beginning of the day to develop common understanding of

guidelines and values, at the end of the day to share participant experiences or current events, or

any time of day to engage in cooperative or creative activities that emphasize honesty,

responsibility, accountability, and compassion (e.g., creative writing, show-and-tell, story-

telling) (Ortega et al., 2016). Other practices can be integrated throughout daily interactions,

such as using restorative language and questions, supportive tone, or inviting body language that

contribute to an overall restorative mindset (see Table 2 for a more thorough description of these

practices).

Researchers Gregory, Skiba, and Mediratta (2017) proposed a framework for increasing

equity in school discipline, which encompasses all levels of the school's ecology. One

particularly vital locus of that ecology is changing teacher behavior. Education literature

suggests that professional development sessions for teachers on classroom management and

climate strategies can enhance staff retention, instructional time, and student engagement

(Browers & Tomic, 2000). When these professional development resources are depleted,

academic achievement decreases and disciplinary referrals rise. Primary strategies aimed at

improving environmental support for students can involve quality assessment of how effective

behavioral strategies are taught to students (Ysseldyke & Christenson, 2002). Likewise, teachers

can be taught to use more proactive strategies to minimize the need for a higher tiered

intervention. My Teaching Partner, developed by Allen et al. (2011), coaches teachers on how to

encourage SEL through positive, culturally-sensitive interactions with students, therefore

targeting perceptions and bias from cultural mismatch that might negatively influence

instruction, classroom management, and office referrals. Similarly, Fenning and Jenkins (2019) have initiated efforts to focus educators' awareness on the root causes of discipline disparities in their own schools and provide recommendations, such as a self-evaluation to recognize their own implicit bias.

Professional development topics can also involve culturally-sensitive training in the relation between trauma and behavior problems in order to reduce snap judgments, inappropriate disciplinary responses, and improve trauma-focused education. The National Child Traumatic Stress Network (NCTSN) has developed tips for educators to help youth after experiencing community trauma such as a death, disaster, or school violence. Some strategies include supportive listening, accommodation of varying trauma responses, patience with minor behavior problems or inattention, reassurance of safety, and modeling of coping strategies such as relaxation and social support activities (NCTSN Schools Committee, 2008). This is especially important training for preventing under-resourced students from developing trauma-induced behavior problems which can result in unnecessary exclusion as well as being labeled by staff as abnormal in their response to complex and chronic trauma.

Secondary Tier (Targeted Interventions)

Next, targeted, secondary supports assist youth with mild to moderate misbehavior, about 15% of the school population, through additional group or individual interventions (Bradshaw et al., 2021). For PBIS, this could include increased instruction in key SEL skills to encourage self-regulation in a more concentrated manner, such as social skills groups (Bear et al., 2015). Some students may benefit from increased adult supervision from teachers throughout the day, or academic support when disruptive behaviors function as a way to escape challenging tasks.

Knowing what motivates certain behaviors through observation is how teachers in this tier encourage replacement behaviors. They can then remind students of classroom expectations using pre-corrections, or gestures or verbal statements that signal to a whole class, group, or individual student what to do before they misbehave. Increased opportunities for positive reinforcement are also part of the secondary tier, such as a check-in and check-out procedure up to 5-7 times per day with a teacher who completes a daily progress report and provides positive adult attention (Crone et al., 2010). A screening process and a behavioral intervention team are essential for identifying students that need additional support, consider the function of a student's behavior, and monitor student progress with these supports.

Specialized SEL teachings might address appropriate peer interactions, taking turns in conversations, giving compliments, or using strategies such as "stop and think" to reduce impulsivity. This instruction is tailored to reach key social competence goals that allow for adaptive behaviors, such as following directions and engaging in group play. Broadly, targeted strategies can include modeling, guided practice, coaching, corrective feedback, and the capacity for positive peer influence (Stoiber, 2004). Manualized social skills programs are common, such as Check and Connect to prevent school drop-out (Sinclair et al., 2005) and the Social Skills Improvement System to focus on 20 essential social skills (Gresham & Elliot, 2008). Cognitive behavioral interventions can also be incorporated to treat specific internalizing or externalizing symptoms that are contributing to problem behaviors that lead to exclusionary punishment.

While the proactive discipline approaches reduce the need for more intensive services and time spent implementing them by setting students up for success, reactive discipline strategies are still needed to respond to misbehavior immediately (Zuckerman, 2007). With its

roots in many indigenous cultures around the world (e.g., Native American tribes, Maori in New Zealand, Celtic practices, Aboriginal Australians) (Hamlin & Darling, 2012), Restorative Justice is defined as "an approach to achieving justice that involves, to the extent possible, those who have a stake in a specific offense and to collectively identify and address harms, needs, and obligations, in order to heal and put things as right as possible" (Zehr, 2015, p. 48). In this way, crime is viewed as injury to people and relationships rather than rule-breaking, and provides wrongdoers a chance to take responsibility, make amends for their actions, strengthen relationships with others, and develop socioemotional skills rather than simply being punished or blamed (Mirsky & Wachtel, 2008). Distinct from the primary tier, restorative conversations, peace circles, and peer conferences in the secondary tier are utilized to create space for assuming responsibility and creating solutions, either more informally right after the incident or with planned structured meetings with the affected parties (Table 1).

Tertiary Tier (Individual Interventions)

Lastly, intensive tertiary or individual supports are aimed at the remaining 5% with high frequency and severity of behavior problems or externalizing risk factors, who would benefit from more time-intensive interventions (Stoiber & Gettinger, 2016). If students are unsuccessful in the tertiary tier, then it is likely they will be referred for special education services. For PBIS, a functional behavior assessment, administered by school counselors, teachers, parents, and/or paraprofessionals, isolates the specific behavior, its purpose, and factors that maintain the behavior (Gresham et al., 2013). Then using operant conditioning, a formal plan is made to teach appropriate behavior, strengthen contingencies for positive behavior, place problem behaviors on extinction, and apply logical consequences (e.g., time-out, detention, call home, office referral)

to eliminate severe and harmful behaviors (Sugai & Horner, 2009). Wraparound supports go

further to identify natural supports within the student's system and set of strengths to improve

their functioning, which can include evidence-based interventions (e.g., multisystemic therapy,

pharmacological treatment), community services, and interpersonal connections (Merrell &

Gueldner, 2010). Culture and context, such as family rituals, ethnicity, neighborhood

environment, and language are pertinent to consider when designing PBIS plans in this tier.

While secondary PBIS practices may require 6-20 weeks, tertiary practices last much longer and

could consist of a host of evidence-based interventions such as Olweus' Bully Prevention or the

Good Behavior Game as well as consistent monitoring of progress (Stoiber & Gettinger, 2016).

     As for individualized Restorative Practices (RP), re-entry circles and restorative

conferences help those who have received a suspension or expulsion reintegrate back into their

school or community setting (Gregory, Bell, & Pollock, 2016). In this way, even those highest

risk students are seen as a valued member of the community, which helps prevent their transition

into the school-to-prison pipeline by allowing them to reengage with more prosocial activities or

access special education services if needed. Altogether, RP seem to work best when they involve

both proactive and reactive practices to create a whole-school culture of working together based

on common values and skills that support healthy relationships and community success

(McCluskey et al., 2008).

     As mentioned above, there are ways in which teachers are being taught to recognize and

address their own cultural and racial biases, which has also been applied to the implementation

of the most severe disciplinary punishments. For example, discipline checklists have been

created so that staff have a sequential guide for decision-making before assigning exclusionary

punishment (McIntosh et al., 2014; Fenning & Johnson, 2016). In this way, staff members can openly address their implicit biases by acknowledging whether an offense truly poses a threat and warrants a suspension or expulsion or whether non-exclusionary alternatives would be more appropriate and have been fully exhausted. This method helps reduce subjectivity by providing clear guidelines for adhering to reformed policies and student codes of conduct. There are also questions that prompt the staff member to consider whether mental health or special education needs are playing a role in their behavior (e.g., trauma, substance use, bullying, disability). In this case, referral to an alternative school, therapeutic school, or counseling may be most appropriate.

Table 1. Selected MTSS Interventions Organized by Tier

Intervention	Strategies	Tier
Positive Behavior Intervention Supports (PBIS)	• 3-5 positively stated expectations for behavior that are rewarded with a token system • Increased instruction in key SEL skills to encourage self-regulation • Increased opportunities for positive reinforcement and monitoring (e.g., Check-In/Check-Out) • Apply logical consequences (e.g., time-out, detention, call home, office referral) to eliminate severe and harmful behaviors • Functional Behavior Assessment • Wraparound supports - identify natural supports within the student's system and set of strengths to improve their functioning	Primary Secondary Tertiary
Restorative Practices (RP)	• Restorative Mindset • Restorative Language • Community-Building Circle • Restorative Conversation • Peace Circle • Peer Conference • Re-entry Circle • Restorative Conferencing	Primary Secondary Tertiary

Social and Emotional Learning (SEL)	• Strengths-based instruction that focuses on building youth's internal capacity to approach distress and conflict productively so that they can focus on academic achievement and feel safe and connected • Activities can be reflective about the decisions students make (e.g., reflection sheets, role-playing, interviews), provide curriculum instruction on socioemotional skills and how to prevent problem behaviors such as bullying or substance use, or take the form of community-building practices • Specialized SEL teachings might address appropriate peer interactions, taking turns in conversations, giving compliments, or using strategies such as "stop and think" to reduce impulsivity. This instruction is tailored to reach key social competence goals that allow for adaptive behaviors, such as following directions and engaging in group play. • Broadly, targeted strategies can include modeling, guided practice, coaching, corrective feedback, and the capacity for positive peer influence (Stoiber, 2004). • Manualized social skills programs are common, such as Check and Connect to prevent school drop-out (Sinclair et al., 2005) and the Social Skills Improvement System to focus on 20 essential social skills (Gresham & Elliot, 2008).	Primary Secondary
Staff Coaching	• Encourage SEL through positive, culturally-sensitive interactions with students, therefore targeting perceptions and bias from cultural mismatch that might negatively influence instruction, classroom management, and office referrals • Professional development topics can also involve culturally-sensitive training in the relation between trauma and behavior problems in order to reduce snap judgments, inappropriate disciplinary responses, and improve trauma-focused education • Discipline checklists as a sequential guide for decision-making before assigning exclusionary punishment • Questions that prompt the staff member to consider whether mental health or special education needs are playing a role in their behavior	Primary Secondary Tertiary

Table 2. Core Elements of Restorative Practices across MTSS Tiers

Element	Definition	Tier	When to Use
Restorative Mindset	A restorative mindset describes how a person understands community and one's role in the community.	Primary	Always
Restorative Language	Restorative language encourages positive interaction. Restorative language uses "I" statements to remain non-judgmental, gives the speaker positive feedback through empathetic listening, and encourages him/her to speak using restorative questions.	Primary	Always
Talking Circle (Community-Building Circle)	Pro-actively build relationships and community among classroom or team. Talking circles may be used as daily check-ins (such as Morning Meetings), to set classroom norms and agreements, teach social and emotional skills, provide feedback, and discuss pertinent issues and topics.	Primary	Proactively to build community and SEL skills; to work collaboratively on a problem.
Restorative Conversation	After a behavior incident and when the student is calm, a Restorative Conversation can help guide him/her through reflection, problem solving, and repairing harm. Rather than chastising a student for his/her behavior, Restorative Conversations help identify root causes and place responsibility on students to understand the impact of their behavior and take steps to make things better.	Secondary	At the moment of a minor occurrence, in a safe space, and after the respondent has accepted responsibility for his/her behavior. Or, to better understand an incident that has occurred and the root causes of the behavior exhibited.

Peace Circle	A Peace Circle is a planned, structured meeting between a person or people who caused harm, the person or people who were harmed, and both parties' family and friends, in which they discuss the consequences of wrongdoing and decide how to repair harm.	Secondary	Can be used in response to repeated inappropriate behaviors; persistent disruptive behaviors, most seriously disruptive behaviors, and some very seriously disruptive behaviors
Peer Conference	A Peer Conference (sometimes called peer mediation, peer council, or peer jury) is a voluntary, student-led process in which a small group of trained Peer Conference members provide a positive peer influence as they work to empower referred students to understand the impact of their actions and find ways to repair the harm they have caused.	Secondary	In response to repeated inappropriate behaviors, persistent disruptive behaviors and some seriously disruptive behaviors
Re-entry Circle	A structured process that reintegrates students back into the school if they received an out of school suspension to re-establish connection with the community. Students who have already been expelled are provided the opportunity to take full responsibility for their actions, which resulted in the expulsion, and to make amends for the impact their actions caused toward the school community.	Tertiary	Following an OSS or expulsion
Restorative Conferencing	Restorative Conferencing is a structured formal process that involves all members of the community affected by a particular incident. Those who cause harm are held accountable for their actions, those harmed are given a voice in the process and agreements are made to address needs, repair harms and prevent future wrongdoing.	Tertiary	Formal conferences may be utilized as suspension diversion if the student/s actively participate and follow through with the resulting plan of action.

Note. Adapted from Kidde, 2017 and the Chicago Public Schools Restorative Practices Guide & Toolkit.

MTSS to Reduce Behavior Problems

As introduced above, there are a host of diverse Multi-Tiered System of Supports (MTSS) practices in the education sphere, many of which have been disseminated throughout the United States. The earliest tiered intervention system for schools, PBIS, received substantial attention and development following the reauthorization of the Individuals with Disabilities Education Act (IDEA, 1997) when positive behavioral interventions were encouraged to address problematic behaviors in both regular and special education settings. Since that time, a wide variety of interventions have been developed and show effectiveness in mitigating disruptive behaviors and enhancing prosocial outcomes across the entire school, but to varying degrees depending on the intervention type, duration, and setting.

Several systematic reviews and meta-analyses have been performed which present this evidence for PBIS and SEL. PBIS was deemed to have enough evidence to support large scale implementation (Kincaid & Horner, 2017), yet it can take two to three years to develop PBIS leadership and three to five years to achieve treatment integrity (Solomon et al., 2012). Broadly, PBIS research supports better school behavior, graduation, school climate, OSS, and achievement outcomes across all tiers (Losen, Hewitt, & Toldson, 2014; Skiba, Arrendonda, et al., 2014), yet most of the twenty primary tier studies analyzed by Solomon et al. (2012) focus on elementary schools. Similarly, a meta-analysis of 82 SEL interventions by Taylor, Oberle, Durlak, and Weissberg (2017) showed that SEL teachings foster gains in academic performance, high school and college graduation rates, and prosocial behaviors and attitudes, plus reductions in emotional and conduct problems, arrests, STIs, and teen pregnancies; although it is unclear how these outcomes vary depending on age and cultural background.

While PBIS and SEL supports represent the largest number of published studies for MTSS behavioral interventions, the Restorative Practices literature is emerging. The first systematic review to assess RP practices and violence outcomes in school consisted of ten studies summarized by Katic, Alba, & Johnson (2020). Most of the ten studies concluded that RP was effective in reducing disciplinary referrals and bullying while increasing social skills, self-esteem, and positive social relationships. Even so, there were several limitations to these works, including dosage information for only half of these studies, only two randomized controlled trials, and a wide range in training duration and treatment fidelity. Additionally, an updated review of the overall positive effects of RP done by Fronius et al. (2019) noted that these practices can be used overtly as a "suspension diversion" program to replace staff's use of suspensions for misbehavior with RP practices, but it is difficult to tell whether reduced suspensions are a result of this intention, or rather a byproduct of reduced school misconduct.

MTSS to Reduce Exclusionary Discipline Practices

While it is clear that these MTSS tiered discipline strategies have the potential to promote more positive student behaviors, they can also intend to reduce the implementation of exclusionary practices when inclusionary MTSS strategies are chosen over punitive practices to manage behavior problems. Still, the majority of the available literature reviews focus on MTSS as it relates to behavior rather than school discipline policies. Clarification is needed to better understand how MTSS supports address the disproportionate school-to-prison pipeline directly through disciplinary reform.

There are only a handful of published reviews of the literature that focus specifically on how alternative discipline interventions reduce exclusionary practices, all published in the past

three years. Valdebenito et al. (2018) performed a systematic review and meta-analysis of randomized controlled trials published through December 2015, which consisted of 37 studies of mainstream school interventions intended to reduce the rates of suspension as an alternative to exclusion across all school ages. They found a small significant drop in exclusion rates over the first six months of intervention; but this was not sustained over time, few studies represented each type of intervention, and effects were greater for expulsion and ISS than OSS. Of note, none of these studies evaluated Restorative Practices. More recently, Mielke & Farrington (2021) investigated randomized controlled trials published through December 2019 that measured the effect of interventions on arrest or suspension, but reduction in suspension or arrest did not have to be the primary intended outcome of the intervention (e.g., academic intervention that also reduces suspensions). They identified a total of 14 studies, also with small but significant reductions in suspensions and arrests for programs stronger in implementation, and with greater effects for universal over targeted MTSS interventions such as Positive Action SEL programming and school-wide PBIS that promote a supportive environment, skill-building, and coaching staff to use positive behavior practices. Importantly, this review only included one RP evaluation, with significant reductions in suspensions only for elementary school students.

Other reviews have focused on the literature for specific interventions. Gage et al. (2018a) aimed to see how PBIS affected disciplinary exclusion, identifying four experimental or quasi-experimental studies with control groups that covered 90 schools, with only one study assessing high schools instead of elementary schools. They found a moderate effect of PBIS on reducing suspensions but not office disciplinary referrals (ODRs). McDaniel et al. (2020) also examined PBIS articles through 2019 but for underrepresented populations specifically. They

found that referrals and suspensions were generally reduced, but only six of the 46 studies

discussed culturally responsive practices being implemented and did not elucidate what these

practices were (which could include implicit bias training). Lastly, Darling-Hammond et al.

(2020) performed a non-systematic review of quantitative RP research through December 2019

and highlighted evidence that teachers who use more emotional communication had a reduced

racial discipline gap in their referrals, as well as a narrowed gap for Los Angeles, Oakland, and

Denver Public Schools; but noted that RP research is limited in the United States, especially in

terms of internal validity needed to attribute this reduced disproportionality to RP specifically.

Aims of the Current Systematic Review

In sum, there are numerous alternative discipline programs that are designed to eliminate

the school-to-prison pipeline and fall within the MTSS tiers of intervention. PBIS includes

universal, targeted, and individual proactive and reactive supports to provide reinforcement for

positive behaviors. RJ is not based on contingencies, and instead strives to enhance daily

interactions and school culture through several practices aimed at strengthening school

relationships and respect and facilitating constructive and collaborative conflict resolution, both

preventatively and in response to misbehavior. Staff coaching provides culturally-sensitive

training for acknowledging and reducing discipline bias, which is crucial for schools with many

youth exposed to trauma. Several of these interventions have the potential to replace

exclusionary discipline practices, but can also be very abstract, diffuse, and time-intensive;

making implementation challenging without external resources. Strategies from different

intervention manuals are also often implemented simultaneously, causing confusion as to which

strategies should be used and when. Systematic reviews on this topic are essential for clarifying

and isolating MTSS interventions that appear to be most successful and useful in directly

reducing the prevalence of exclusionary punishments that fuel the school-to-prison pipeline.

Within all of the aforementioned reviews, there are critical limitations that preclude full

understanding of the implementation and effectiveness of the interventions available to replace

and reduce exclusionary practices in schools. Of the two reviews that examined all types of

interventions together, they were limited to randomized controlled trials and Mielke &

Farrington (2021) included only one study on RP practices despite updating four years worth of

literature that proceeded the Valdebenito et al. (2018) search. Additionally, Mielke & Farrington

(2021) did not require interventions that purposefully aim to reduce exclusionary outcomes,

making it challenging to assert whether the reduction in suspensions and arrests was driven by

behavior change, more tolerant school policies, or an unrelated intervention factor. Of the other

available reviews, they only focused on one particular type of intervention, and either had limited

studies for inclusion (Gage et al., 2018a for PBIS), did not illuminate the key practices that

reduced exclusion (McDaniel et al., 2020 for PBIS), or failed to use a systematic review process

(Darling-Hammond et al., 2020 for RJ). Because of these limitations, it is difficult to grasp a full

understanding of which MTSS supports are truly efficacious and effective in replacing exclusion

to dismantle the school-to-prison pipeline, which mechanisms or processes are contributing to

these reductions, and when exclusion is still utilized when these strategies are not successful.

The following systematic review aims to address these gaps in understanding by utilizing

careful, thorough search procedures that build upon past research and address all available types

of interventions designed to reduce rates of exclusion, beyond simply reducing behavior

problems. First, this review will assess the evidence base for any school-based intervention

studies that specifically aim to reduce exclusionary outcomes in order to more clearly identify

the "suspension diversion" programs and make these disciplinary procedures replicable. Given

the low number of randomized controlled trials in this field, this review will also expand upon

the more limited internal validity of less rigorous quantitative findings and enhance external

validity by taking qualitative research into consideration. It is notable that past reviews included

studies from mostly elementary school settings, therefore the current systematic review hopes the

inclusion of qualitative and quasi-experimental research will illuminate what works in middle

and high schools as well, since secondary students are at greatest risk of suspensions and racial

disproportion (Losen et al., 2016). This review process should also provide more intervention

context to describe specific practices across all of the tiers of MTSS (rather than just prevention),

as well as increase the chance that more amorphous interventions such as RP will be included in

the review. Lastly, this systematic review is being performed at a time when there has been a

rapid increase in strategies developed and implemented in the United States to replace exclusion

over the past decade, and is likely to include studies that were not yet available for prior reviews.

There is also considerable concern that there are too many different types of interventions

being implemented across school districts through the MTSS framework, making it challenging

for staff to choose an effective approach to reduce exclusion. For example, Sugai et al. (2016)

highlights the need for more effective and efficient adoption of extensive professional

development, unified perspectives and initiatives reaching the same goals with the same

implementation strategies, and evidence-based and culturally-relevant approaches. Similarly,

considerable time is needed when adopting a new discipline approach. For example, RJ practices

must also be embedded into the school culture so that it is accepting, respectful, supportive, safe,

and self-perpetuating (Beckman et al., 2012; González, 2012). This process can take three to five years, and without consistent financial and other supports, schools will return to more engrained exclusionary tactics (Wadhwa, 2016; Anfara et al., 2013). Especially concerning, schools with the largest populations of Black, Hispanic, and low-income students are the most likely to benefit from tiered discipline supports like RP practices, but are the least likely to implement them (Payne & Welch, 2018). This systematic review will serve the additional purpose of condensing and comparing the wide variety of practices in order to distill the most manageable and acceptable methods of inclusion.

Research Questions

      Question One. What are the various universal, secondary, and tertiary MTSS disciplinary interventions being implemented in United States schools as alternatives to replace or reduce exclusionary discipline practices?

      Question Two. For which outcomes (e.g., exclusion rates, school climate, academic variables) are these interventions effective?

      Question Three. In which settings (i.e., elementary, middle, and high school) and for which populations (e.g., underrepresented groups) are these interventions effective in reducing exclusion?

      Question Four. How are these interventions qualitatively perceived by the implementers and recipients of the alternative strategies? In other words, what do stakeholders view as the positive effects and processes through which successful strategies reduce exclusion?

CHAPTER TWO

METHOD

Systematic Review Design

The design of this systematic review was developed based on guidelines from the

Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al., 2021, version 6.2).

A convergent mixed-methods review protocol was utilized to synthesize quantitative research

(RCTs, non-randomized quasi-experimental studies, observational studies, correlational studies,

and descriptive studies), qualitative research, and mixed-methods research together in one review

(Noyes et al., 2021). This method was utilized to address gaps in the previously published

systematic reviews and meta-analyses, as identified in the research aims. A convergent mixed-

methods review synthesizes quantitative and qualitative evidence separately so that these types

of evidence can address the "what" and "how," respectively, regarding effectiveness of the

interventions in reducing exclusionary discipline practices. These findings are then synthesized

together to create more nuanced understanding of the effectiveness of interventions and the

mechanisms of action that make each intervention successful. A sequential review that builds

upon past reviews was not appropriate for the current research given the variable time frames,

interventions studied, and inclusion criteria used for those reviews. A meta-analysis, while

performed in other past reviews, was not conducted given the lack of study homogeneity and

small number of RCTs available on this topic, as well as the wide scope and variety of

intervention strategies assessed across MTSS tiers. The structure of this systematic review

follows the recommended methodology from the Preferred Reporting Items for Systematic

reviews and Meta-Analyses 2020 Statement (PRISMA; Page et al., 2021).

Study Eligibility Criteria

Studies for inclusion in this systematic review were limited to research that specifically

aimed to replace or reduce exclusion in schools through alternative non-exclusionary disciplinary

practices. It was not essential that reducing exclusionary discipline outcomes (e.g., suspensions,

expulsions) be the primary aim, but this was required to be an intended outcome of the research.

Only peer-reviewed articles were selected in order to identify high quality research that is

deemed valid and original by experts (Kelly et al., 2014). While "grey" literature such as

dissertations, non-published or conference papers, and non-peer-reviewed reports and articles

can provide valuable insight and reduce publication bias, they are also highly variable in

structure and detail and have potential for significant methodological concerns. They have also

been shown to minimally impact the key conclusions of several systematic reviews regardless of

their inclusion or omission (Hartling et al., 2017). Similarly, research without original data

outcomes (e.g., book chapters, systematic reviews, meta-analyses) were also not included.

This review only included research published in the United States because of the distinct

racial disparities that persist in the U.S. compared to other countries as a result of federal zero

tolerance policies (Skiba & Rausch, 2006). As a result, this also limited studies included in this

review to those published in English. The articles in this review were also limited by publication

date. The amendment to the Education for All Handicapped Children Act (EHA, 1975), called

the Individuals with Disabilities Education Act (IDEA, 1997), was intended to provide access to

public schools for students with disabilities and prevent schools from discriminating against and

excluding them. In this way, students with disabilities were to be kept in the least restrictive environment and encouraged to interact with non-disabled students. This law also encourages "positive behavioral interventions, strategies, and supports" to address student behavior when it interferes with learning (IDEA, 1997). Given the timing of this law, the current systematic review includes only research published in 1997 and later, so that the most relevant research was included coinciding with the national shift in policy towards implementing non-exclusionary discipline practices and keep students in mainstream education.

There were no restrictions set for study design, such that all quantitative and qualitative methods were included as long as the study contained original data. Setting was limited to elementary, middle, junior high, secondary, and high schools such that only students from kindergarten (if applicable) through grade twelve were included. This decision was made based on the wide range of discipline practices used in early childhood education and universities or colleges, representing unique settings that differentiate them from the grade school years in developmental level. Setting was further restricted to mainstream school samples, such that alternative schools, therapeutic schools, and criminal justice settings were not investigated in order to focus analyses on more typical and wide-reaching education settings. Additionally, therapeutic and alternative schools are inherently specialized and trained to implement more inclusionary practices, as they are designed to support at-risk or disabled students and are often recommended as a tertiary tier intervention for behavior problems or learning difficulties. Lastly, there were no restrictions placed on the type of participant in studies for this review, such that students or staff could participate.

Search Strategy

The search strategy was developed with the assistance of Tracy Ruppman, a Loyola University Chicago Research and Learning Librarian from the School of Education. The article search took place from November 2020 through March 2021. Systematic searches were completed using the following ten databases: Academic Search Complete, Education Research Complete, APA PsycINFO, ERIC, Professional Development Collection, Child Development & Adolescent Studies, OmniFile Full Text Select, Educational Administration Abstracts, APA PsycARTICLES, and Social Work Abstracts. These databases were selected based on their relevance to the systematic review questions and their high frequency of use within education research. The ten databases were combined for each search in order to minimize duplicates and peer-reviewed articles were selected as a filter. Databases were searched from the time of their inception to March 2021, then more specific exclusion criteria were applied during the abstract screening and full-text review process.

Search terms consisted of 14 different combinations of Boolean phrases that ranged in specificity in order to capture a wide range of articles as well as specific disciplinary interventions that may be more difficult to capture using only broad terms. These Boolean phrases included various combinations of the following key terms and their plural forms: intervention, practice, alternative, exclusion, suspension, expulsion, school, primary school, elementary school, middle school, secondary school, high school, discipline, disciplinary, restorative justice, and restorative practices. Reference sections for articles added during the full-text review process were reviewed in order to identify additional individual studies that might have been missed by the structured database searches but could still be relevant for the current

systematic review. These articles were then also uploaded for the abstract review process, discussed below.

Data Screening and Extraction Procedures

References from the above searches were imported into Covidence software for ease of organization and collaboration between reviewers on the review team. This software allowed for automatic removal of duplicates, creation of a PRISMA diagram, assigning voting roles, and ensuring blinding of raters which helps to minimize bias and protect the integrity of the review process (Kellermeyer et al., 2018). The review process was broken down into three stages: abstract screening, full-text review, and extraction. There were 16 total team members that contributed in some way to the review process, including ten undergraduate, two post-baccalaureate, two graduate students, and two experienced research professors. Of note, only four of these team members were actively involved throughout the entire review process; the majority of reviewers were involved during the extraction and quality assessment phases given the large number of articles and time intensive nature. Two independent reviewers were required for screening, full-text review, data extraction, and quality assessment in order to reduce errors and maximize judgment accuracy (Buscemi et al., 2006). The lead researcher (author of this review) took part in completing abstract screening and full-text review, and was solely responsible for resolving conflicts between raters for both of these steps. The lead researcher did not complete data extraction or quality assessment, so that she could remain blind for final data consensus.

All team members involved in abstract screening and full-text review were trained in how to understand and apply the inclusion and exclusion criteria for this review. Proportion

agreement between raters for abstract screening (both endorsing "yes" for including or "no" for excluding) was .73 at the lowest and .98 at the highest. Once an abstract was deemed relevant to the current review, it was automatically sent to the full-text review group. In this group, the full-text of each article was uploaded and reviewed more in-depth to confirm or deny its goodness of fit with inclusion criteria. Notes were added by reviewers to aid decision-making when resolving conflicts between raters. Reasons were indicated for each full-text article that was excluded. Verbal discussions were held with team members to clarify inclusion criteria and enhance their understanding of the topic. Proportion agreement between raters for full-text review (both endorsing "yes" for including or "no" for excluding) was .78 at the lowest and .94 at the highest.

Articles included during full-text review were automatically sent for extraction. A data extraction template (Table 3) was created based on Cochrane, Campbell, and PRISMA guidelines as well as aspects of the participant/study demographics, intervention implementation, study methodology, outcomes, and statistics that were deemed important to this particular review. Reviewers were asked to attempt one extraction then return to the lead researcher for questions. Notes were provided within the template to guide reviewers in copying over the most relevant parts of each article into each section (e.g., how to determine the intervention tier). The completed extractions forms for each article were then compared by the lead researcher so that the most accurate and comprehensive information was chosen or amended during consensus. Reliability across raters for extraction was not calculated given the wide range in free text responses.

Quality Assessment Procedures

Assessment of the quality of primary research is essential for uncovering biases that can lead to underestimation or overestimation of the effectiveness or generalization of findings (Higgins et al., 2021). There are a limited number of mixed-methods systematic review tools available given that most reviews have focused on quantitative work. The Quality Assessment for Diverse Studies (QuADS; Harrison et al., 2021) was selected as the quality assessment tool for this research based on review of literature. This tool was chosen because of its applicability to health services research, particularly psychology, and its ability to appraise both quantitative and qualitative research using the same criteria, therefore acknowledging their different strengths and weaknesses using consistent language with less bias favoring randomized controlled trials. It was previously called the Quality Assessment Tool for Studies with Diverse Designs and has been heavily cited in the past ten years and used in more than 80 reviews (QATSDD; Sirriyeh et al., 2012). These items were refined based on feedback from researchers (e.g., adding more explicit examples for each item, reducing bias towards quantitative data) and the resulting QuADS tool has been shown to display sufficient inter-rater reliability (kappa = .65; Chauhan et al., 2020). The Mixed Methods Appraisal Tool (MMAT; Hong et al., 2019), another highly utilized quality assessment tool for mixed methods systematic reviews, was not chosen due to its limited number of quality assessment items (i.e., 5 yes/no criteria per type of study), the inability to compare ratings equally across each type of study, and the increased difficulty that more novice researchers would have in making reliable yes/no judgments for each question (e.g., "Is the risk of nonresponse bias low?," "Are the outputs of the integration of qualitative and quantitative components adequately interpreted?").

The resulting 13 items included in this tool assess the methodological quality (e.g., appropriate sampling to address the research aims), evidence quality (e.g., strengths and limitations critically discussed), and quality of reporting (e.g., description of data collection procedure) for qualitative, quantitative, and multi-method studies (Table 4). Each item is rated on a scale from 0 to 3, with 0 indicating no mention of the quality component and 3 indicating the best possible quality. The four anchors for each item were created to provide sufficient responses for items that are more nuanced than a dichotomous scale (Harrison et al., 2021). A total score can therefore be calculated for descriptive purposes, but this was not designed for determining whether studies should be excluded, as the cut-off would be arbitrary. Each reviewer independently scored studies that they had already extracted using the QuADS criteria uploaded as a template into Covidence. The lead researcher then compared the quality ratings for each article and made the final determination. Intraclass correlation coefficients (ICC) were calculated for each quality assessment item using SPSS version 27 to determine the inter-rater reliability, or how consistent the raters' responses were with each other. Koo and Li (2016) indicate that ratings below .50 are poor, with .50-.75 indicating moderate agreement, .75-.90 indicating good agreement, and .90 and above rated as excellent.

Data Analysis

The analysis began with an overview of study characteristics followed by tabulation of extracted data and a list of MTSS interventions identified in this review (Research Question 1). To analyze the extracted data, first the quantitative findings were narratively described to understand which interventions led to significant disciplinary and related school outcomes and for which populations and settings these interventions made significant impacts (Research

Questions 2-3). A narrative approach was also used to describe qualitative perceptions and observations of the processes that made each intervention effective in reducing exclusionary discipline strategies, as well as strategies to address implementation barriers (Research Question 4). Quality assessment findings provided additional descriptive context for determining how deeply each study should be interpreted and highlighted in terms of the study's impact in answering the systematic review questions.

A thematic synthesis was utilized to produce descriptive and analytic themes to highlight common or key ingredients within and across interventions and tiers for each type of data (Thomas & Harden, 2008). The qualitative themes were integrated with quantitative findings and organized into the MTSS tier framework to allow for meta-synthesis in the discussion (Dixon-Woods, 2011). In other words, the quantitative effectiveness data and qualitative evidence for each particular intervention were categorized by their specific MTSS tier (Tier 1, 2, or 3) to easily identify the student's level of need that is most applicable for each strategy. These findings are then organized further by the type of school or the type of intervention when helpful for highlighting overlap or similarities across studies (i.e., elementary, middle, high, Restorative Practices, PBIS). Commentary is provided based on the population served by the intervention or school. Meta-synthesis included implications and recommendations for future design, implementation, and evaluation of alternative discipline interventions to reduce or replace exclusionary practices.

Table 3. Data Extraction Template

Information Extracted	Response Options
General information	
Study ID #	Free response
Title	Free response
Lead author & university	Free response
Publication year	Free response
Journal	Free response
Study characteristics	
Methods	
Aim of study	Free response
Alternative discipline practice(s) under study	Free response
How is this practice defined?	Free response
How is this practice implemented?	Free response
How is this practice measured?	Free response
What MTSS tier does this practice fit into?	Tier 1 Tier 2 Tier 3 Tiers 1 & 2 Tiers 2 & 3 Tiers 1 & 3 Tiers 1, 2, & 3
Are the practice(s) under study the main focus of this article?	Yes No
Overall study design	Quantitative Qualitative Both
Specific study design	Randomized controlled trial Non-randomized experimental study (quasi-experimental) Correlational study Systematic review Qualitative research Randomized controlled trial & qualitative research

	Quasi-experimental study & qualitative research
	Other (specify)
Start/end date of intervention	Free response
Start/end date of data collection	Free response
Study funding sources	Free response
Possible conflicts of interest for study authors	Free response

Participants

Population description (students)	Free response
Sample description (students)	Free response
Population description (staff)	Free response
Sample description (staff)	Free response
Sample type	Students
	Staff
	Both
School level	Elementary
	Middle
	High
	Elementary & middle
	Middle & high
	All 3
	Unknown
School type	Private
	Public
	Alternative
	Charter
	Unknown
	Other (specify)
Setting	Urban
	Rural
	Suburban
	Urban & rural
	Urban & suburban
	Suburban & rural
	All 3
	Unknown
Inclusion criteria	Free response

Exclusion criteria	Free response
Briefly describe recruitment strategy	Free response
Total number of participants	Free response
Total number of schools	Free response
If two or more groups were compared, what were they?	Free response

Measures

More than one time point of assessment?	Yes No
What are the time points for data collection?	Free response
What are the (relevant) outcomes being assessed?	Free response
How are these outcomes measured?	Free response

Results

Summary of the main quantitative findings and conclusions	Free response
Secondary quantitative findings relevant to evaluation of alternative practices	Free response
Were there statistically significant quantitative associations between the intervention and any demographics?	Yes No Unknown Not applicable
Were there statistically significant quantitative associations between the intervention and any outcomes (DVs)? (e.g., effect sizes)	Yes No Unknown Not applicable
What were the statistically significant quantitative associations?	Free response
What were the non-significant quantitative associations?	Free response
What were the qualitative themes and conclusions reported?	Free response
Any notable qualitative quotes (or pages to refer to) that highlight the impact or process of the intervention?	Free response
Any other helpful sections that would be worth copying over from this paper?	Free response

Table 4. Quality Assessment Scoring

QuADS Criteria	Scoring			
	0	1	2	3
1. Theoretical or conceptual underpinning to the research	No mention at all.	General reference to broad theories or concepts that frame the study. e.g. key concepts were identified in the introduction section.	Identification of specific theories or concepts that frame the study and how these informed the work undertaken. e.g. key concepts were identified in the introduction section and applied to the study.	Explicit discussion of the theories or concepts that inform the study, with application of the theory or concept evident through the design, materials and outcomes explored. e.g. key concepts were identified in the introduction section and the application apparent in each element of the study design.
2. Statement of research aim/s	No mention at all.	Reference to what the sought to achieve embedded within the report but no explicit aims statement.	Aims statement made but may only appear in the abstract or be lacking detail.	Explicit and detailed statement of aim/s in the main body of report.
3. Clear description of research setting and target population	No mention at all.	General description of research area but not of the specific research environment e.g. 'in primary care.'	Description of research setting is made but is lacking detail e.g. 'in primary care practices in region [x]'.	Specific description of the research setting and target population of study e.g. 'nurses and doctors from GP practices in [x] part of [x] city in [x] country.'

4. The study design is appropriate to address the stated research aim/s	No research aim/s stated or the design is entirely unsuitable e.g. a Y/N item survey for a study seeking to undertake exploratory work of lived experiences.	The study design can only address some aspects of the stated research aim/s e.g. use of focus groups to capture data regarding the frequency and experience of a disease.	The study design can address the stated research aim/s but there is a more suitable alternative that could have been used or used in addition e.g. addition of a qualitative or quantitative component could strengthen the design.	The study design selected appears to be the most suitable approach to attempt to answer the stated research aim/s.
5. Appropriate sampling to address the research aim/s	No mention of the sampling approach.	Evidence of consideration of the sample required e.g. the sample characteristics are described and appear appropriate to address the research aim/s.	Evidence of consideration of sample required to address the aim. e.g. the sample characteristics are described with reference to the aim/s.	Detailed evidence of consideration of the sample required to address the research aim/s. e.g. sample size calculation or discussion of an iterative sampling process with reference to the research aims or the case selected for study.
6. Rationale for choice of data collection tool/s	No mention of rationale for data collection tool used.	Very limited explanation for choice of data collection tool/s. e.g. based on availability of tool.	Basic explanation of rationale for choice of data collection tool/s. e.g. based on use in a prior similar study.	Detailed explanation of rationale for choice of data collection tool/s. e.g. relevance to the study aim/s, co-designed with the target population or assessments of tool quality.

7. The format and content of data collection tool is appropriate to address the stated research aim/s	No research aim/s stated and/or data collection tool not detailed.	Structure and/or content of tool/s suitable to address some aspects of the research aim/s or to address the aim/s superficially e.g. single item response that is very general or an open-response item to capture content which requires probing.	Structure and/or content of tool/s allow for data to be gathered broadly addressing the stated aim/s but could benefit from refinement. e.g. the framing of survey or interview questions are too broad or focused to one element of the research aim/s.	Structure and content of tool/s allow for detailed data to be gathered around all relevant issues required to address the stated research aim/s.
8. Description of data collection procedure	No mention of the data collection procedure.	Basic and brief outline of data collection procedure e.g. 'using a questionnaire distributed to staff'.	States each stage of data collection procedure but with limited detail or states some stages in detail but omits others e.g. the recruitment process is mentioned but lacks important details	Detailed description of each stage of the data collection procedure, including when, where and how data was gathered such that the procedure could be replicated.
9. Recruitment data provided	No mention of recruitment data.	Minimal and basic recruitment data e.g. number of people invited who agreed to take part.	Some recruitment data but not a complete account e.g. number of people who were invited and agreed.	Complete data allowing for full picture of recruitment outcomes e.g. number of people approached, recruited, and who completed with attrition data explained where relevant.

10. Justification for analytic method selected	No mention of the rationale.	Very limited justification for choice of analytic method selected. e.g. previous use by the research team.	Basic justification for choice of analytic method selected e.g. method used in prior similar research.	Detailed justification for choice of analytic method selected e.g. relevance to the study aim/s or comment around of the strengths of the method selected.
11. The method of analysis was appropriate to answer the research aim/s	No mention at all.	Method of analysis can only address the research aim/s basically or broadly.	Method of analysis can address the research aim/s but there is a more suitable alternative that could have been used or used in addition to offer a stronger analysis.	Method of analysis selected is most suitable approach to attempt to answer the research aim/s in detail e.g. qualitative interpretative phenomenological analysis might be considered preferable for experiences vs content analysis to elicit frequency of occurrence.
12. Evidence that the research stakeholders have been considered in research design or conduct	No mention at all.	Consideration of some the research stakeholders e.g. use of pilot study with target sample but no stakeholder involvement in planning stages of study design.	Evidence of stakeholder input informing the research. e.g. use of pilot study with feedback influencing the study design/conduct or reference to a project reference group established to guide the research.	Substantial consultation with stakeholders identifiable in planning of study design and in preliminary work e.g. consultation in the conceptualization of the research, a project advisory group or evidence of stakeholder input informing the work.
13. Strengths and limitations critically discussed	No mention at all.	Very limited mention of strengths and limitations with omissions of many key issues. e.g. one or two strengths/ limitations mentioned with limited detail.	Discussion of some key strengths and weaknesses of the study but not complete. e.g. notable omissions or lack of depth of explanation.	Thorough discussion of strengths and limitations of all aspects of study including design, methods, data collection tools, sample & analytic approach.

CHAPTER THREE

RESULTS

Overview of Studies

The search process and total number of articles included and excluded in this systematic

review are summarized in Figure 2. The initial set of searches generated 530 studies, 210 of

which were duplicates and automatically removed by the Covidence program. Of the 353 peer-

reviewed article abstracts screened, 214 were excluded during abstract and title screening

because of their clear irrelevance to reducing exclusion in schools or lack of fit with inclusion

criteria (e.g., no original data, criminal justice setting). If their fit was not easily determined

based on the abstract information, then they were included in full-text review. Over the course of

reviewing full-text articles, 33 additional studies were identified through review of references

and searches on Google Scholar then added to the abstract screening process. There were 139

total articles for which an electronic version of the article was uploaded to Covidence for more

thorough review of the entire text to assess eligibility for this systematic review. A total of 80

articles were excluded during full-text review due to a host of reasons: interventions that did not

seek to reduce or replace exclusion in schools, interventions performed outside of the United

States, no evaluation (e.g., description of an intervention), lack of original data outcomes (e.g.,

reviews, commentary), articles published before 1997, inappropriate settings (e.g., therapeutic

schools, alternative schools, school bus), lack of peer review, duplicates missed by Covidence

programming, outcomes not related to exclusion, and irrelevant populations (e.g., outside the K-12 grade range).

Fifty-nine articles met all inclusion criteria and were included in this systematic review. No study was eliminated due to methodological quality, although this is taken into account within the discussion of quality assessment findings. The inter-rater reliability findings for quality assessment are presented in Table 5 which includes the intra-class correlation for each quality assessment item. The two-way mixed ICC with average absolute agreement ranged from .31 at the lowest reliability ("Clear description of research setting and target population") to .75 at the highest ("Description of data collection procedure," "Strengths and limitations critically discussed"). The two-way mixed effect was chosen given the specific set of raters involved and an average was determined across the multiple raters (Perinetti, 2018). Absolute agreement was chosen to assess repeatability based on the exact same scores across each rating and takes into account the systematic error among the raters and their ratings (McGraw & Wong, 1996). A list of the MTSS interventions identified during the review is incorporated into Table 6 and a list of significant outcomes is in Table 7. Quality assessment ratings for each quality assessment item and the article's total score are summarized in Table 8. The demographic information for each article is summarized in Table 9. The study methods and outcomes are summarized in Table 10 and Table 11, respectively, for each article.

Of those fifty-nine studies, 24 (41%) were classified into the primary tier, 8 (14%) were in the secondary tier, 9 were in the tertiary tier (15%), and 18 (31%) represented multiple MTSS tiers of intervention. In terms of grade level, 13 (22%) took place in elementary schools, 4 (7%) took place in middle schools, 12 (20%) took place in high schools, 18 (31%) took place in

schools across all grade levels, 3 (5%) took place in elementary/middle schools, and 9 (15%)

took place in middle/high schools. In terms of study design, 8 (14%) of the studies were RCTs,

21 (36%) were quasi-experimental, 14 (24%) were correlational, 10 (17%) were qualitative, and

6 (10%) incorporated mixed quasi-experimental and qualitative methods. There were 41 (69%)

studies examining multiple schools and 29 (49%) studies performing large scale evaluations of

15 schools or more.

When summarizing the interventions that involved multiple evaluations, there were 22

(37%) studies evaluating some form of PBIS, 5 (8%) with interventions taking place in schools

already using PBIS, 17 (29%) studies evaluating some form of Restorative Practices (RP), and 6

(10%) assessing a threat assessment model. In terms of fidelity assessment, only 21 (36%) of the

studies included in this systematic review incorporated a measure of fidelity to the intervention,

with 14 assessing fidelity to PBIS, 2 assessing RP fidelity, and one study for 5 other

interventions (GSP, VSTAG, ISLA, PATHS, EBH-CRP). Some studies incorporated post-

training evaluations of commitment to or understanding of the intervention, but did not assess

implementation fidelity (i.e., Cornell et al., 2011). Regarding dosage, PBIS implementation

length ranged from 1-10 years, with 2-4 years being the most common length of time. RP

implementation ranged from 1-6 years, with only 1 year being the most common time frame.

Other interventions with known dosage amounts ranged from 5 months to 2 years. Of note, only

two interventions were considered brief, such as the Alternative to Suspension for Violent

Behavior (Breunlin et al., 2006) family training program over 8 hours and the brief empathic

mindset training for teachers for 70 minutes over 2 months (Okonofua et al., 2016).

When looking at the types of school setting, 25 (42%) of the studies took place in urban communities, with 41 (69%) of the studies conducted in public schools. In terms of the sample demographics, 34 (58%) studies incorporated school samples with over 50% racial/ethnic minority students attending, with 7 of those samples including predominantly Black students and 8 predominantly Latino students. Of these majority minority studies, 15 evaluated PBIS (with 2 majority Black, 2 majority Latino samples), 13 evaluated RP (with 2 majority Black, 4 majority Latino), and 6 evaluated other interventions (with 4 majority Black, 1 majority Latino). Additionally, 29 (49%) of studies included samples with majority of the students receiving free or reduced price lunches, meaning they were at 185% of poverty or below, classified as mid to high poverty. Of these majority poverty studies, 13 evaluated PBIS, 9 evaluated RP, and 7 evaluated other interventions. These findings are summarized in Table 12.

Figure 2. PRISMA Flow Diagram of Selection of Studies for Inclusion in the Review

Identification of studies via databases and registers

Identification

Records identified from:
    Databases (n = 530)
    Citation searching (n = 33)

→

Records removed *before screening*:
    Duplicate records removed (n = 210)

Screening

Records screened
(n = 353)

→

Records excluded
(n = 214)

Articles sought for retrieval
(n = 139)

→

Articles not retrieved
(n = 0)

Articles assessed for eligibility
(n = 139)

→

Articles excluded (n = 80):
    Wrong intervention (n = 20)
    Not in US (n = 13)
    No original data (n = 13)
    No evaluation of intervention (n = 10)
    Pre-1997 (n = 9)
    Wrong setting (n = 9)
    Duplicate (n = 2)
    Wrong outcomes (n = 2)
    Not peer reviewed (n = 1)
    Wrong population (n = 1)

Included

Studies included in review
(n = 59)

Table 5. Intra-Class Correlations for QuADS Quality Assessment Items

Item	ICC	F	df1	df2	p	CI 95%	Reliability Classification
1	0.652	2.870	69	70	<.001	[0.44 0.78]	Moderate
2	0.518	2.076	69	70	0.001	[0.23 0.7]	Moderate
3	0.311	1.452	69	70	0.061	[-0.11 0.57]	Poor
4	0.504	2.015	69	70	0.002	[0.2 0.69]	Moderate
5	0.635	2.737	69	70	<.001	[0.41 0.77]	Moderate
6	0.620	2.634	69	70	<.001	[0.39 0.76]	Moderate
7	0.701	3.342	69	70	<.001	[0.52 0.81]	Moderate
8	0.753	4.044	69	70	<.001	[0.6 0.85]	Good
9	0.644	2.813	69	70	<.001	[0.43 0.78]	Moderate
10	0.592	2.448	69	70	<.001	[0.34 0.75]	Moderate
11	0.687	3.200	69	70	<.001	[0.5 0.81]	Moderate
12	0.585	2.408	69	70	<.001	[0.33 0.74]	Moderate
13	0.751	4.019	69	70	<.001	[0.6 0.85]	Good

Note. ICC values less than .50 = poor, .50-.75 = moderate, .75-.90 = good, and .90 and above = excellent. ICC analyses were performed on all extracted articles (n = 70), then sample was reduced to 59 articles after examining extracted data.

Research Question One

The first research question for this systematic review asks, what are the various universal, secondary, and tertiary MTSS disciplinary interventions being implemented in United States schools as alternatives to replace or reduce exclusionary discipline practices? The following table provides a breakdown of the interventions identified through the systematic search (Table 6).

Table 6. List of MTSS Alternative Disciplinary Interventions

Tier	Intervention
Primary Tier	**School-Wide Positive Behavior Intervention Supports (SWPBIS)** - Bradshaw, 2010; Bradshaw, 2012; Mitchell, 2013; Gage, 2018b; Lee, 2021; Zhang, 2021; Ward, 2013; Luiselli, 2005; Morrissey, 2010; Barnhart, 2008; Muscott, 2008; Smolkowski, 2016; Childs, 2016; Kim, 2018 **Implicit bias and/or empathic mindset training for teachers** - Greet-Stop-Prompt, Classroom Check-Up, My Teaching Partner – Secondary (Cook, 2018; Gion, 2020; Gregory, 2014; Okonofua, 2016) **Restorative proactive circles** - Lustick, 2020c; Garnett, 2020; Kervick, 2020 **Relationship-building strategies** - Anyon, 2018
Secondary Tier	**Restorative Conversations, Circles, and Conferences** - Anyon, 2016; Gregory, 2018; Huang, 2020; Ortega, 2016; Lustick, 2020a; Bruhn, 2020 **ISS** - Turpin, 1997
Tertiary Tier	**Threat Assessment** - Cornell, 2009; Cornell, 2011; Cornell, 2012; Nekvasil, 2015; Cornell, 2018; Maeng, 2020 **Alternative to Suspension for Violent Behavior** - Breunlin, 2002; Breunlin, 2006 **Alternatives to Suspension** - Chin, 2012 **School Referral Reduction Protocol** - Teske, 2013

Multiple Tiers	**Positive Behavior Intervention Supports (PBIS)** - Baule, 2020; Gage, 2019; Nocera, 2014; Cruz, 2018; Freeman, 2019; Simonsen, 2012; Scott, 2004) **PBIS + trauma-informed training** - Von der Embse, 2019 **PBIS + Inclusive Skill-Building Learning Approach** - Nese, 2020 **Restorative Practices** - Hashim, 2018; Anyon, 2014; Stinchcomb, 2006; Lustick, 2020b; Sandwick, 2019 - SaferSanerSchools (Mirsky, 2007; Gregory, 2016; Mansfield, 2018; Rainbolt, 2019) **Promoting Alternative Thinking Strategies SEL program** - Osher, 2014 **Emotional and Behavioral Health – Crisis Response and Prevention** - Bohnenkamp, 2021

Research Question Two

The second aim of this systematic review was to identify for which outcomes (e.g., exclusion rates, school climate, academic variables) these interventions are effective. Relevant quantitative results for each study are highlighted and separated by broad intervention category and grade level.

School-Wide PBIS

The majority of published effectiveness trials identified that fell within the category of universal, or primary tier interventions, were designed to investigate PBIS strategies. These fourteen quantitative studies broadly assess the implementation and success of School-Wide PBIS strategies (typically referred to as SWPBIS for universal strategies) in reducing rates of exclusion, by way of reduced behavior problems, reduced use of punishment, and improved school climate factors. Of these fourteen peer-reviewed articles, twelve included large-scale effectiveness trials with multiple schools representing an entire region, state, city or district

school system, with seven of these in elementary schools specifically. In addition, two small-scale studies each looked at tier one PBIS in a single elementary school and a single high school.

Elementary schools. For example, Bradshaw et al. (2010), Bradshaw et al. (2012), and Mitchell et al. (2013) were all randomized controlled trials to investigate the effectiveness of SWPBIS in 21 Maryland public elementary schools that were receiving this intervention from 2002-2007 in five rural, urban, and suburban school districts compared to 16 schools without training. For the five years of intervention, a SWPBIS team of 6-10 staff members and an administrator provided training and action plans with the assistance of an external behavioral support coach (e.g., school psychologist, guidance counselor) for consultation and functional behavioral assessments. Universal PBIS strategies included three to five expectations for positive student behavior posted around the school, lesson plans to teach these expectations, a reward system using tangible reinforcers for positive behaviors, a school-wide system to respond to behavioral violations with consistent consequences in the classroom or an office referral, and tracking disciplinary data to inform improvements in implementation. In this way, these systems were designed to create positive change in staff behaviors, which would then alter student behaviors and the entire school environment. SWPBIS school teams received an initial 2-day training from Dr. George Sugai, one of the developers of the program, as well as yearly booster sessions and quarterly professional development meetings from the Maryland State Leadership Team.

Bradshaw et al. (2010) found that across these five years of implementation, the elementary schools implemented the universal PBIS strategies with high fidelity, which led to significantly reduced student suspensions and ODRs while non-SWPBIS schools remained

unchanged. However, schools already had lower rates of ODRs compared to the national average, therefore effect sizes were relatively small. In addition, there was a trending effect for fifth grade students in the growing number of proficient and advanced scores on the state standardized math achievement test. Together, these results suggested that SWPBIS could create sustainable change in the use of more positive behavior supports, and that these changes to the school environment reduced exclusionary discipline outcomes for elementary school students. Bradshaw et al. (2012) found further evidence suggesting significantly lower levels of behavior problems (i.e., aggression, disruptive behavior, difficulty concentrating), lower likelihood of ODRs by 33% (especially for girls), and better social-emotional functioning (i.e., emotion regulation, prosocial behavior; especially for those exposed to the intervention in kindergarten) for students in Maryland SWPBIS schools compared to control schools. Lastly, Mitchell et al. (2013) looked into school climate variables and found that across all schools, greater use of exclusionary discipline strategies was correlated with lower student-rated scores of order and discipline, whereas greater use of positive behavior supports in the classroom was associated with better order, discipline, fairness, and student-teacher relationships. Contrary to its intended purpose, exclusionary discipline conveyed a sense of disorganization in the classroom, whereas proactive strategies made students feel more supported in their school. All three of these studies had high quality assessment scores ranging between 32-35, with 39 as the highest possible score, suggesting the methodology and results are relatively trustworthy and unbiased.

Similar to Maryland research, there are two published articles evaluating Georgia public elementary schools' implementation of primary tier PBIS strategies using the support of a district PBIS leadership team and Georgia Department of Education. Gage et al. (2018b) investigated

whether SWPBIS implemented with fidelity (greater than 70% on the Benchmarks of Quality

scale) would result in fewer disciplinary incidents, ISS, and OSS compared to propensity score

matched comparison schools without SWPBIS. Researchers found medium effect sizes across all

outcomes, with the largest effect size for ISS. When implementation fidelity was higher (i.e.,

greater than 85%), there were larger treatment effects for OSS and behavioral incidents.

Likewise, Lee et al. (2021) aimed to replicate Gage et al. (2018b) within the same school year of

data collection (2015-2016), while using a more nuanced assessment of exclusionary discipline

outcomes by race, gender, and IEP status. For those schools implementing SWPBIS with

fidelity, there were significantly fewer rates for all eight types of exclusionary discipline (e.g.,

ISS, single and multiple OSS, expulsion, alternative school placement, referral to law

enforcement, arrests, and days missed for OSS), with effect sizes ranging from 0.26 to 0.70.

These moderate treatment effects for SWPBIS compared to control schools held true for males,

Black students (except for arrests), and students receiving IEP services (except for expulsions

and alternative school placement), suggesting that SWPBIS can reduce disciplinary exclusion for

some of the most vulnerable populations of students. These studies were also rated as 33 and 36

respectively for quality assessment, again signifying strong methods and outcomes.

  In addition to Georgia and Maryland school research, three additional published studies

of public elementary schools were identified in this review. Zhang et al. (2021) found that

SWPBIS implemented over two to three years in 106 Western elementary schools was more

successful when more time was allocated to professional development, collaboration, and

planning, which in turn fostered greater treatment integrity, followed by reduced suspension

rates. Like other large-scale studies above, this mediation analysis was rated high in quality with

a score of 33. Interestingly, more years of staff experience was associated with significantly lower treatment integrity and lower suspension rates.

The Safe and Civil Schools (SCS) model was evaluated by Ward and Gersten (2013), which consists of seven days of SWPBIS training derived from Sprick et al. (2002). This model is very similar to other SWPBIS approaches including a leadership team, consultant, data analysis, developing priorities for improvement in safety, behavior, and discipline, and training all other staff in this proactive discipline plan. Ward and Gersten (2013) evaluated SCS implementation through a randomized controlled trial in a large urban school district with more than 80,000 students, the majority of which were minority students receiving Free and Reduced Price lunches (FRP) and who scored low on district testing. These authors found that SCS training resulted in changes to the development and enforcement of school discipline policy (e.g., staff perceptions of safe and secure environment) in elementary schools, as well as staff perceptions of student behavior (e.g., bullying, classroom disorder) compared to waitlist control elementary schools over two years (2008-2010). Additionally, students perceived greater safety in their school and were suspended less and for fewer total days. Quality assessment analysis revealed a strong score for this study (QuADS score = 33).

Luiselli et al. (2005) focused on SWPBIS in only one urban elementary school with majority Black students (88%) and students qualifying for Free and Reduced Price lunches (FRP; 90%), as well as a history of recurrent discipline problems, academic underachievement, and poor staff morale. Researchers found that these proactive, preventive, and skill-building strategies resulted in decreasing rates of office referrals and suspensions over three years (1999-2001), as well as increasing percentile ranks for reading comprehension and mathematics. This

research shows that in one very high-need elementary school, SWPBIS has the capacity to reduce discipline problems and increase exposure to classroom instruction with less exclusion, and that it is perceived as effective to reduce disruption by teachers. However, these findings should be interpreted with some caution, given that the quality of this study was substantially lower than large-scale effectiveness studies with a score of 23.

Secondary schools. In addition to elementary schools, SWPBIS has also been implemented in secondary school settings to assess effectiveness. In one urban high school in Chicago, Morrissey et al. (2010) found that one year of SWPBIS implementation (2002-2003) was associated with decline in ODRs compared to baseline, along with fewer students receiving multiple ODRs. While this article did not report significance levels or complete descriptions of data collection or analysis (QuADS score = 14), the authors concluded that SWPBIS seemed to have a positive impact on student behaviors and they heavily involved stakeholders in the research process. Training students in positive behavior strategies seemed to produce more effective drops in exclusionary referrals when this was done in a large assembly for the whole school, with small groups used for booster education. Raffle tickets for small prizes were used to reinforce positive behavior, suggesting a potential adaptation for the high school setting.

A large-scale effectiveness trial of SWPBIS was conducted in the Los Angeles Unified School District for 64 middle and high schools with high suspension rates for students with disabilities (Barnhart et al., 2008). While this publication did not include significance levels and had limited rationale or critique for their data collection and analysis (QuADS score = 15), they did compare between SWPBIS and non-SWPBIS schools. The authors present broad statistics showing reduced OSS rates by about 5% for those with and without disabilities in SWPBIS

schools over five years (2003-2008), particularly when SWPBIS support became fully

operational (2006-2007). Uniquely, teachers with low numbers of referrals had the chance to

help support those with higher numbers, and positive effects from early adopting schools

convinced other schools to adopt SWPBIS. Of note, SWPBIS schools still had higher rates of

suspensions overall, suggesting need for continued implementation training and sustainability.

For both of these secondary school studies, the conclusions are limited by the potential for

unknown factors complicating the outcomes given their quality ratings.

All grade levels. Four large-scale effectiveness studies for SWPBIS took place in

elementary, middle, and high schools. Muscott et al. (2008) studied New Hampshire schools and

found that SWPBIS primary tier interventions led to significantly reduced ODRs, ISS, and OSS

over two years of implementation (2003-2005), with the greatest effects for middle and high

schools, although no statistical significance values are reported. In terms of benefit analysis,

schools were able to collectively recover 864 days of teaching, 1,701 days of learning, and 571

days of leadership as a result of reduced ODRs and suspensions. Despite the lack of statistical

significance, this study was deemed high quality with a rating of 34. Smolkowski et al. (2016)

evaluated the real-world effectiveness and maintenance of a scale-up of the SCS SWPBIS

approach for 74 non-randomized public elementary and secondary schools in the same large

urban school district as Ward & Gersten (2013). Over four years, schools at all grade levels

showed moderate and steady improvements in staff-reported problem behavior (e.g., bullying,

defiance), suspensions and number of days suspended, tardiness rates, and staff-perceived school

safety and enforced discipline policy. Despite the lack of randomization, Smolkowski et al.

(2016) also displayed high quality with a QuADS rating of 33.

Childs et al. (2016) examined 1,122 Florida schools to assess the relationship between SWPBIS fidelity and school-level behavioral outcomes over time (average 3.71 years of implementation, SD = 0.98). Descriptively, ODR, ISS, and OSS rates appeared to drop over time, yet growth modeling showed that fidelity to SWPBIS methods did not predict growth trajectory changes. However, schools with higher Benchmarks of Quality ratings of fidelity were associated with lower ODR, ISS, and OSS, both at the beginning and end of the four years of data collection (2010-2014). Likewise, Kim et al. (2018) also found that ODRs and OSSs decreased over time in 477 schools across ten states (average 3.65 years of implementation, SD = 2.69). More specifically, schools that had implemented SWPBIS first and sustained fidelity over time had reductions in OSS, whereas schools starting with low fidelity did not reduce exclusion. However, years implementing SWPBIS was a stronger predictor of academic achievement than fidelity. Similar to the large-scale elementary school SWPBIS studies reviewed in prior sections, greater fidelity and sustainable implementation of proactive behavior strategies appears to enhance reductions in exclusionary discipline outcomes across grade levels, with high levels of research quality (QuADS scores > 30).

PBIS and Supplemental Interventions

In addition to school-wide universal PBIS strategies, there are eight additional articles investigating PBIS across multiple tiers of intervention. These articles included specialized group and/or individual interventions for at-risk and high-risk behaviors, with four studies examining exclusionary disparities directly. Of those disparity studies, Baule (2020) looked explicitly at whether implementing PBIS reduced racial disparities in suspensions in urban, public middle and high schools in a high poverty school district over two years (2015-2017). While students were

predominantly qualifying for FRP lunch, they were also majority White with a lower Hispanic population than most urban areas. This study showed reduced rates of total suspension and students receiving more than one suspension, suggesting students were suspended less overall given that the average number of suspensions among suspended students remained the same. While a statistically significant disparity in racial disparities persisted, these disparities did reduce except for Hispanic versus White students. Of note, Baule (2020) had a low quality score of 14 due to issues with theory, aims, data collection and analysis details and rationale, discussion of strengths and limitations, and lack of stakeholder involvement.

Similar to Childs et al. (2016), Gage et al. (2019) investigated SWPBIS in 536 Florida public schools, but with focus on all tiers of intervention and specifically rates of corporal punishment, exclusion, and legal consequences for all students and vulnerable students. Interestingly, schools implementing SWPBIS with fidelity were larger and had more White students, with only 9% of them high schools and 19% middle schools. They found that schools implementing with fidelity for 3-5 years had significantly lower OSS, especially for students with disabilities and Black students. Middle and high school students had higher rates of referrals to law enforcement and expulsions. One limitation is that fidelity was measured for Tier 1 only, so it is unclear what Tier 2 and 3 interventions were adhered to and may not have been effective enough to reduce legal consequences. This study had high quality for all domains except for stakeholder involvement (QuADS = 36).

Not all studies showed reductions in exclusionary discipline disparities. Nocera et al. (2014) examined SWPBIS as it pertains to a low-performing middle school in Connecticut. Over the course of implementation across tiers over two years (2008-2010), the school showed

significantly reduced infractions for problem behavior by 40%, teacher discipline referrals by 36%, overall suspensions by 39%, and suspensions for students with disabilities by 51%, as well as improved school climate and student resilience scores on 30 out of 47 items, 25% increase in reading scores, and 11% increase in math scores. African American and Hispanic students still received disproportionate suspensions despite their rates decreasing along with others. Qualitative analysis showed that while leadership was needed to help increase staff buy-in, they also felt the reward system was highly motivating, the five-step system promoted fair referrals with an emphasis on prevention, and the monthly school data team meetings promoted focus on student success and fostered positive school climate. While this research suggests strong improvements in exclusionary discipline and school climate outcomes, it was weaker in quality of its aims, recruitment context, analysis explanation, and stakeholder involvement (QuADS = 22).

Covering more schools, Cruz et al. (2018) used a correlational design to assess the relationship between school practices, demographics, and OSS risk in 41 schools in a diverse California district from 2010-2016, 27 of those schools using SWPBIS. Contrary to hypotheses, students in schools that implemented SWPBIS actually had a higher risk of suspension, with continued overrepresentation of African American students, males, and students with disabilities and even increasing risk for Latino, African American, and males in these schools. Still, the district reduced suspensions overall and these findings may signify need for more culturally relevant discipline practices. For Latino students overall, suspensions reduced significantly, with the gap with White students closing by ten years at that rate, but African American, special education, AAPI, and male students did not show reducing rates. Of note, as parent education

increased, suspension risk decreased. This study had acceptable quality with a rating of 30 based on limitations mostly in the recruitment information and stakeholder involvement.

Two studies addressed ways to augment PBIS with additional strategies. Von der Embse et al. (2019) investigated ways to enhance primary tier PBIS strategies in a K-8 demonstration site in Philadelphia by providing trauma-informed training, coaching teachers in classroom management to prepare for secondary tier interventions, and implementation of a decision-making process to allocate secondary tier interventions. By helping teachers learn how to identify mental health risk through screening on the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS), the implementers aimed to increase a sense of safety, transparency, and collaboration at the universal level while helping students build coping skills and trusting relationships through secondary interventions, such as progressive muscle relaxation, Check-In Check-Out, or Cognitive Behavioral Intervention for Trauma in Schools (CBITS; Stein et al., 2003) for more intensive support. In this school with 90% qualifying for FRP lunch and 80% minority youth, use of the SAEBRS revealed 50.3%-68.5% of students were deemed at risk for social-emotional and behavioral problems, with risk increasing by 5-6% from fall to winter of both years of implementation. While ODRs reduced during the first year of intervention in classes with universal screening, so did SAEBRS risk by 13.6%, then again by 3% in 2018. For three-fifths of students with emotional risk, Tier 2 relaxation groups were provided and prevented this risk from increasing over fall to winter. Authors concluded that trauma-informed classroom practices (e.g., building student-teacher relationships, defining expectations, consistent responses to misbehavior) supported the environment necessary for successful

secondary tier interventions. This research scored a 26 on the QuADS quality scale, with some

lack of justification for analysis or involvement of stakeholders.

Nese et al. (2020) investigated another enhancement of PBIS with the Inclusive Skill-

Building Learning Approach (ISLA) for a pilot year in two middle schools with 77% FRP

qualifying students and 30% students of color. This approach built upon universal PBIS by

training staff how to explicitly minimize the use of exclusion, respond effectively to problem

behaviors, and use systems to discipline students equally. These systems made sure only the

most serious behavior incidents received exclusion. Less serious behavior incidents were

operationally defined and received reteaching, redirection, restitution, counseling, parent

involvement, or behavioral contracts rather than punishment. Behaviors that could not be

handled in the classroom received a structured referral to the office and phone call home,

functional behavior assessment, coaching on replacement behaviors, a restitution plan to repair

damage if needed, and a Reconnection Conversation Card to prepare for reentry into the

classroom. Schools showed decreases in ODRs, ISS, OSS, expulsions, and lost instructional

minutes at the end of the pilot year. Staff felt ISLA helped arm students with prosocial skills and

reconnect back with the classroom, but wanted to know more about what skills students were

working on through ISLA. They also noticed positive aspects such as healthier teacher-student

relationships, increased self-esteem, removed the stigma of the principal's office, and non-

judgmental space to work on behavior. This article had a solid quality score of 30 with only

weaknesses in recruitment data and stakeholder involvement.

Freeman et al. (2019) examined how PBIS implementation fidelity relates to student-

level ODRs, suspension, attendance, and GPA in 15 high schools with the largest proportion of

students being African American followed by Hispanic, with the majority qualifying for FRP lunch. They found that with greater PBIS fidelity there was a significant decrease in ODRs and suspensions over the 2015-2016 school year, as well as fewer absences and unexcused tardies. Fidelity scores were found to negatively correlate with the percentage of students qualifying for FRP lunch in that school. This research built upon the findings of Muscott et al. (2008) to again show a decrease in exclusionary discipline with PBIS and how this also affects attendance, specifically for high school students. Simonsen et al. (2012) examined SWPBS in 428 Illinois schools from 2000 to 2008. They found that as time increased, standardized reading and math scores also increased while ODRs decreased. Additionally, as implementation increased, ODRs, OSS, and total suspensions decreased. Of note, elementary schools had the highest rate of fidelity at 81% in 2008, followed by middle schools at 73%, and high schools at 31%. The QuADS score for both of these articles was 30, representing solid quality across most domains except for stakeholder involvement and recruitment or sampling data.

 With all three tiers in play, Scott et al. (2004) evaluated the cost benefits of implementing PBIS in Maryland elementary schools from 2000 to 2002. Over two years of implementation, ODRs decreased from 608 to 46, and suspensions decreased from 77 to 22. In terms of time spent on office referrals and suspensions, administrators saved 14.6 days in the first year and 16.8 days in the second year, which equated to $6,024.84 and $6,932.69 respectively. In terms of student instructional time lost due to referrals and suspensions, students saved 72.7 days in the first year and 86.2 days in the second year, costing $3,182.08 and $3,772.97 respectively. Quality assessment indicated a solid score of 29 except for limited recruitment, sampling, and

stakeholder domains, but this article also requires clarification for why suspensions were

calculated as discrete events rather than the true length of suspension.

Empathic Mindset and Implicit Bias Strategies for Teachers

In addition to highly-studied SWPBIS and PBIS strategies, there are also universal

prevention strategies that more directly address the root of disparities in exclusionary discipline

for vulnerable populations, which have been studied in four smaller-scale effectiveness

evaluations. In this way, teachers are targeted as the key implementers of discipline policies to

recognize their underlying biases that can affect the way they treat students. The GREET –

STOP – PROMPT (GSP) approach was developed and piloted within a school district already

receiving SWPBIS training as a concrete solution to address disproportionality in exclusionary

discipline (Cook et al., 2018). This approach involved proactive classroom management

strategies to prevent disruptive behavior, review of situations that can elicit a teacher's implicit

bias, and reactive strategies that evoke empathy and consistency towards all students. Cook et al.

(2018) examined three urban public elementary schools from the same Western school district

that were receiving federal and state oversight due to racial disproportionality in exclusion and

special education referrals. This single case experimental concurrent multiple baseline design

was very high in reporting and methodological quality (QuADS score = 35). Analyses revealed

reduced likelihood of Black male office referrals (rate cut by two-thirds) as well as reduced

ODRs per student per week (rate cut in half) once the GSP intervention was introduced in the

2015-2016 school year. Educators rated this intervention as both acceptable and feasible in these

three elementary schools. There was also a significant increase in school connectedness for

Black male students, suggesting that these students were subjectively happier with their school in addition to objective measures of exclusionary disciplinary reductions.

The Classroom Check-Up Model (CCU) evaluated as an RCT in Gion et al. (2020) also addressed implicit bias by helping educators see how bias affects classroom behavior expectations, learn ways students would like to be acknowledged, and receive performance feedback on ways they treat students unequally. This model involved providing motivational interviewing and evidence-based classroom management strategies to four teachers across a K-8 and K-5 school. In this sample at baseline, African American students received all exclusionary discipline from these teachers and higher rates of reprimands from three of them. Teachers' implementation of CCU resulted in increased praise and decreased reprimands for African American students which narrowed the racial gap in reprimands, as well as increased praise for all students. Overall, teachers implemented CCU with high fidelity and rated it as effective, acceptable, and a good fit for their school and classroom. The quality of this research did not raise substantial red flags for interpretation except for low justification of the analytic method and involvement of stakeholders (QuADS score = 32).

Another intervention similar to GSP in its focus on teacher behavior and disproportionality, My Teaching Partner – Secondary (MTP-S) was implemented and evaluated in five middle and high schools in a mid-sized city in the southeast (Gregory et al., 2014). This study employed a professional development program to reduce the use of exclusionary discipline for all students, with a special focus on African American students. In this intervention, teachers were encouraged to reflect on recorded videos of their instruction with an assigned coach over the course of the school year (2010-2011). Within this "coaching cycle," coaches directed

teachers to their high quality interactions with students and ways they could improve their behavior in the domains of emotional support, classroom organization, and instructional supports. They completed written prompts and built an action plan with their coach to address challenges and implement new behaviors. For teachers enrolled in MTP-S, African American students and other students had similarly low likelihood of receiving exclusionary discipline, whereas African American students of equivalent achievement, income, and gender were 2.69 times more likely to receive referrals when they had teachers without MTP-S. Interestingly, students in the same courses with the same teachers and peers throughout the entire year received more exclusionary discipline than students who switched groups halfway. This research had a solid quality score of 29, with weaknesses primarily in sampling and recruitment reporting.

Similarly to GSP's empathy component, Okonofua et al. (2016) delivered and evaluated a brief empathic mindset intervention (45-minute module then 25-minute module two months later) in five diverse public middle schools in three urban and suburban school districts in California. Researchers first conducted an experiment that confirmed teachers could be encouraged to adopt this type of non-punitive mindset towards discipline, value students' perspectives, and sustain positive relationships while emphasizing positive behaviors. The second experiment confirmed this promoted students' respect for their teachers and motivation to behave. The third experiment involved a randomized trial of a brief, online intervention to promote empathic mindset in 31 math teachers. Results showed year-long suspension rates were cut in half for teachers who received the brief intervention compared to control teachers, even when controlling for race, gender, and prior year suspension status. Students with a history of suspensions perceived improved student-teacher relationships, such that this intervention

repaired ruptures caused by a punitive mindset. Of note, this study was not as strong in its reporting and justification of its aims, sample, methods, and strengths/limitations, leading to a lower QuADS score of 16.

Restorative Practices

       Secondary tier. While there were no quantitative evaluations of primary or tertiary RP strategies alone, there were two studies that evaluated secondary RP strategies using statistical means within the same school population: Denver Public Schools. The Denver discipline policy was reformed in 2008 to include more secondary RP practices. These more formal and structured practices included restorative circles and conferences that allowed for expression of harmful experiences and collective problem-solving to repair harm in order to hold the wrongdoer accountable and mend their relationships with the community. Anyon et al. (2016) accounted for nesting within schools and controlled for covariates in order to assess the relationship between RP use and exclusionary discipline outcomes in the 2012-2013 school year, then explored the moderation of student racial background. Researchers found that as a student's RP use increased in the first semester, ODR and OSS rates decreased in the second semester. ODR and OSS rates were also lower for schools using more RP overall, suggesting school-level efforts to reduce the use of exclusionary discipline. However, racial disparities were still present for Black, low-income, and special education students regardless of RP involvement.

       Gregory et al. (2018) addressed this by studying whether RP was associated with more fair and equal OSS rates for Black and White students in Denver Public Schools. The current study focuses on 2014-2015 school year. Results showed that participation in RP by referred students was correlated with lower likelihood of receiving an OSS by 69% for all groups, but not

for Black students over others as was hoped. Additionally, school-level responses to infraction with RP practices were again associated with lower suspensions. Unfortunately RP did not lessen racial disparities in Denver Public Schools. However, RP studies were deemed high in quality with 33 and 36 out of 39 total quality assessment points, respectively.

Lastly, Huang et al. (2020) evaluated students' participation in ISS and RP interventions (e.g., circles, mediations, conferences) after conflicts occurred in 116 secondary schools from a large urban district in 2016. Students with one or more OSS or ISS reported poorer perceptions of school climate and negative attitudes towards school, but with small effect sizes. RP practices did not foster any significant changes in school climate or school attitudes across discipline groups, which researchers state could be due to lack of fidelity assessment or due to less powerful effects of RP at the secondary tier compared to primary tier with all students involved in RP (Gregory et al., 2016). Quality assessment revealed a score of 26 due to lacking information regarding the sample, recruitment, data collection, and no involvement of stakeholders. Overall, little new information was gleaned from this analysis beyond the continued negative impact of OSS and ISS on school climate.

Multiple tiers. Several studies also examined the implementation of RP across tiers of intervention simultaneously. The SaferSanerSchools model of RP, including two years of formal training in 11 elements of RP across all three tiers, was evaluated in terms of its effectiveness in three different publications identified in this review. Mirsky et al. (2007) investigated this model in two high schools and one middle school in southeastern Pennsylvania. These authors revealed reduced referrals, detentions, disruptive behavior incidents (e.g., inappropriate behavior, disrespect, classroom disruption), and OSS over the course of the four pilot years (1998-2002).

Staff members reported feeling enthusiastic about RP in improving academic performance, culture of collaboration, and school climate, but also acknowledged that lack of structure and failure to address relationships over content made implementation difficult. However, the quality of this study was very low across all domains except for description of research setting, with a score of 5 on the QuADS.

Regarding disproportionality, Gregory et al. (2016) assessed whether higher multi-tier RP implementation in two large high schools in a small city on the East Coast over two years (2010-2012) was associated with more respectful relationships between students and teachers and less use of referrals for misconduct and defiance across racial and ethnic groups. Data did show that teachers using more RP had fewer exclusionary referrals overall for misconduct or defiance, and these teachers also had a narrower gap in referrals between Asian/White and Latino/African American students. As implementation of RP increased, the student perceived the teacher as more respectful and this did not vary by race. Quality assessment indicated substantial quality (QuADS = 29) but without thorough reporting of sampling and recruitment and without stakeholder input. Mansfield et al. (2018) evaluated this model's impact on a variety of disparities in disciplinary consequences over the course of five years of gradual implementation from 2010-2015 in a large suburban public high school in Virginia. Their research showed that ISS and OSS rates declined over the five years with discipline gaps narrowing across race/ethnicity, gender, and special education status. The principal reported an 80% decrease in office referrals with a noticeable sense of change in the school culture that perpetuated the use of practices, along with a need for continued momentum and buy-in. This research had a quality

score of 23 due to limited data collection and analysis justification as well as lack of stakeholders in the design or conduct of research.

Similar to Barnhart et al. (2008) in terms of school district, Hashim et al. (2018) examined change in suspension rates as a result of policy reform that banned suspensions for willful defiance in 2011 and facilitated RP implementation in 2014 for the highest need locations out of 804 Los Angeles public schools that were already implementing SWPBIS since 2006. Prior to this suspension ban, educators were suspending marginalized students, including Black, Hispanic, male, special education, and middle/high school students, at higher rates than their peers. While suspensions were significantly decreasing overall, these marginalized groups dropped more steeply in suspensions than others as a result of the suspension ban. After plateauing for one year, they continued to drop more steeply compared to other students, with suspensions rates dropping more for schools receiving RP implementation; so much so that the gap between Hispanic and White/Asian students was eliminated. The gaps between Black and non-Black students and those with and without special education status were still present in 2015, but this was only one year after RP began. Researchers are open about the lack of causal implications or extended follow-up data in their study, leaving a quality score of 29 resulting from limited population, recruitment, and stakeholder domains.

Another study that assessed the correlation between changes in suspension policy and racial and poverty-level disparities in exclusionary discipline outcomes was Anyon et al. (2014) in Denver Public Schools. These researchers looked at ISS, RP, and behavior contracts as alternatives to suspension and found that students with behavior problems decreased their odds of OSS if they participated in ISS or RP as measured in 2011-2012, regardless of their

demographics or discipline history, but that odds increased for behavior contracts. The specific

interventions for RP were not delineated but starting in 2008 there was a more centralized

discipline system with checks and balances and law enforcement referrals only if mandated. The

student's racial background, gender, special education status, and designation as seriously

emotionally disabled were prominent risk factors for exclusionary discipline practices. Data

collection procedures and recruitment were limited in reporting, but other than lack of

stakeholder involvement, the quality of this article was substantial (QuADS = 31).

Threat Assessment

At the highest tier of school intervention and student need, the Virginia Student Threat

Assessment Guidelines (VSTAG) were developed by researchers at the University of Virginia to

prevent violence through conflict resolution that understands and solves the root of the threat and

permits the student to stay at school without suspension or expulsion (Cornell et al., 2009). This

involves a six-hour training in using a seven-step decision tree to evaluate whether the threat is

likely to be carried out (e.g., intent and means) and preventing it from happening if so (e.g.,

notifying law enforcement and the potential victims or developing an intervention plan to resolve

the initial problem). VSTAG involves paying attention to problems such as bullying and teasing

in order to prevent escalation into violent behavior and embodies a more flexible, non-punitive

approach to misbehavior. There were six studies identified through the systematic review process

that evaluated these guidelines.

The first study by Cornell et al. (2009) found that the use of the VSTAG model in

Virginia high schools over the school year 2006-2007 fostered significantly lower levels of long-

term suspensions, less bullying and teasing, lower levels of student and bullying victimization,

better learning environment, and increased willingness to ask for help from adults compared to

schools with other or no model to address threats, ranging from small to medium effect sizes.

Only short-term suspensions and aggressive discipline violations were not affected. To achieve

these positive school climate outcomes, staff conveyed a sense of respect and caring for students

and trained them in appropriate ways to seek help before the threats become more serious.

Cornell et al. (2011) extended this quasi-experimental research to follow-up two years later

(2008-2009) and compared 23 VSTAG high schools to 26 high schools without this model while

controlling for student demographics. They found that schools using the threat assessment model

had a 52% reduction in long-term suspensions and 79% reduction in bullying infractions

compared to control schools. Cornell et al. (2012) went further to show that students in

elementary, middle, and high schools randomly assigned to receive these threat assessment

guidelines were more likely to receive counseling services and a parent conference, but less

likely to receive a long-term suspension or alternative school placement than those in a wait-list

control group, with small to medium effect sizes. Implementation fidelity also contributed to

decreased long-term suspensions, suggesting a shift to less punitive approaches to address

student threat of violence across all grade levels. The quality assessment scores for all three of

these studies were 30 or higher on the QuADS scale.

Nekvasil et al. (2015) used the Virginia Secondary School Climate Survey to assess

school climate and safety in 166 middle schools using the VSTAG model in 2013, 47 schools

using no model, and 119 schools using a different threat assessment model. Overall, researchers

found that teachers at middle schools using VSTAG reported feeling safer at school across the

domains of physical safety, adequate security, and worry about school shootings. These schools

also had significantly fewer short-term suspensions (8 per 100 students) than the comparison

groups (12 per 100 students), as well as lower levels of student-reported teasing and bullying,

bullying victimization, and general victimization. There were no significant differences between

groups for student perceptions of structure (i.e., strict but fair) or support or long-term

suspension rates. This research was high in quality (QuADS score = 36), in tandem with

previous VSTAG research.

Cornell et al. (2018) used a correlational design to investigate whether Black or Hispanic

students were more likely to receive a discipline sanction (e.g., OSS, expulsions, school

transfers, and actions by law enforcement) than White students in 785 Virginia public schools

across grade levels that reported at least one threat assessment case in the 2014-2015 school year.

In this research, the threat assessment procedures were not limited to the VSTAG model. About

75% of the threat assessment sample was male, 35% received special education services, 51%

were White, 31% were Black, and 7% were Hispanic. Of importance, there were no significant

differences between the three main racial groups or gender in rates of suspension, expulsion,

transfers, or legal ramifications after threat assessment. As should happen, lower-level threats

were less likely to receive exclusionary discipline. However, students in special education were

more likely to be suspended following a threat assessment. The lower rates of disproportion for

threat assessment differed greatly from the overall disparities for Black (3.1 times more likely to

be suspended than White students) and Hispanic (1.8 times more likely) students regardless of

threat assessment. Like other studies authored by Cornell, methodological, reporting, and

evidence quality was quite high at a rating of 36.

Lastly, Maeng et al. (2020) assessed whether the students receiving a threat assessment through the evidence-based five-step decision tree from the Comprehensive Student Threat Assessment Guidelines (CSTAG) model implemented in 260 schools (formerly VSTAG) differed in exclusionary discipline outcomes compared to state DCJS guidelines implemented in 267 schools during the 2014-2015 school year. The full CSTAG threat assessment team typically involved an administrator, mental health professional, and school resource officer (SRO) to create a plan to prevent violence. Substantive threats required protective action by notifying victims and parents and potentially increasing monitoring of the student. Very serious substantive threats involved threats to kill or use a lethal weapon and received a screening from a mental health professional to understand the problem and determine the need for services, followed by investigation from law enforcement. Students may be suspended while this safety plan is developed, but often can return to school or go into an alternative placement. DCJS, on the other hand, does not directly address exclusion. Maeng et al. (2020) found that students in CSTAG schools were less likely to be suspended, expelled, or receive a legal consequence than DCJS schools. In this way, the CSTAG problem-solving approach appears to be an effective alternative to automatic exclusionary practices for student threats, including referral to law enforcement that fuels the school-to-prison pipeline. Still, Hispanic students were more likely to receive legal consequences than White students for threats and special education students were again more likely to be suspended. Again, this study of threat assessment was high in its quality assessment rating (QuADS score = 31).

Suspension Alternatives

In addition to the six studies on threat assessment, similar yet distinct interventions have been developed and evaluated to replace prolonged 10-day suspensions or legal consequences when a student is violent or at risk of being violent. Breunlin et al. (2002) developed Alternative to Suspension for Violent Behavior (ASVB) which involves a family training program for reducing resuspension rates for acts of physical violence, nonphysical violence, other misbehavior, and overall disciplinary acts. The manual titled, "Making the Smart Choice: Tools for Resolving Conflict" was implemented and evaluated in one suburban public high school from Fall 1997 through Fall 1999. This intervention involves teaching problem-solving, perspective taking, active listening, and anger management skills grounded in conflict resolution theory and mediation along with parent training. A trainer is assigned to the family when a student is suspended for violence, then they must sign a contract and arrange four 90-minute meetings through the Family Institute at Northwestern University within 48 hours in order for their child's suspension days to be reduced. In this particular school, Hispanic students were three times more likely to receive suspensions and African American students were two times more likely. The study showed nonsignificant but trending effects for reduced physical violence resuspensions (four times less likely), nonphysical violence resuspensions, resuspensions overall (almost half the rate of controls), lower disciplinary acts per year (i.e., detentions, ISS), and zero expulsions compared to seven expulsions for the control group. The quality assessment rating for this article was 33, representing solid reporting, methods, and evidence quality.

Breunlin et al. (2006) expanded these findings to examine intervention effectiveness over four years starting in 2001 at the same suburban public high school. They discuss how the use of

trainers outside of their school allowed for open expression of feelings about the suspension. There is also a booster visit where the trainer visits the school to see how the student is doing with the skills. Analyses revealed a sevenfold decrease in all types of suspensions as well as a fourfold decrease in physical violence suspensions. Unfortunately, this article resulted in a QuADS score of 11 due to issues with lack of clear aims, limited methodological description or justification, and insufficient assessment of the evidence. Therefore, Breunlin et al. (2002) and Breunlin et al. (2006) present only tentative support for the effectiveness of ASVB family therapy sessions for reducing resuspensions in youth suspended for violence.

Chin et al. (2012) also sought to replace suspension with skill-building interventions responsive to a student's emotional needs, which was called Alternatives to Suspension (ATS) and implemented within an elementary school by the PBIS team already in place. ATS involved the following strategies after a child received an infraction that would warrant suspension: individualized self-management plans (e.g., behavior monitoring), debriefing and reflection assignments, behavior contracts, natural consequences, individualized SEL training, counseling (e.g., ways to calm down), parent training or monitoring, and ISS intervention rooms to process the incident. This intervention was piloted for nine Latino ELL 4th-6th grade students with low socioeconomic status in 2010-2011 school year. While the study lacked methodological reporting quality (QuADS score = 22), descriptively there were fewer suspensions and resuspensions during the ATS year compared to previous years. A case study description showed that one female student was able to recall calm down strategies and role play scenarios to prevent future behavior problems and increase connection to school.

In a county where school referrals to the juvenile justice system were increasing, a juvenile court judge met with school administration and police to design a Memorandum of Understanding that detailed a School Referral Reduction Protocol to reduce suspension, explusions, and arrests (Teske et al., 2013). In this way, misdemeanor offenses were no longer eligible for law enforcement referral without an initial warning and conflict skills workshop after the first two offenses. Additionally, a multidisciplinary panel was developed to assess the needs of students at risk for court referral and link the child and family to services (e.g., FFT, MST, CBT, wraparound services). Referrals to the court in Clayton County Public Schools reduced by 67.4% given that arrests were no longer being made for low-level offenses. Additionally, there was a felony referral reduction of 30.8% given that warnings were now being given for felonies related to typical adolescent behavior and school police seemed to use more warnings in general without automatic opportunities for arrests. The number of students detained on school offenses reduced by 86%, youth of color referred to the court reduced by 43%, serious weapons on campus reduced by 73%, middle school OSS decreased by 8%, and graduation rates increased by 24%. Since police were not doling out as many arrests, they were more present on campus and more available to hear student concerns on campus which helped solve crimes. By providing alternatives to suspension and arrest, these rates successfully decreased substantially, although the overall research study has significantly limited description of the design, methods, analyses, and conclusion that prevent the findings from receiving much weight (QuADS = 9).

SEL and Crisis De-escalation

Whereas limited published research is specific to SEL practices as they relate to exclusionary discipline, Osher et al. (2014) evaluated Cleveland Metropolitan School District's

use of the Promoting Alternative Thinking Strategies (PATHS) SEL program, student support teams, and planning centers as an ISS/OSS alternative in elementary schools from 2008 to 2012, with 100% of students needing FRP lunch, 68% African American, and 13.2% Latino. PATHS was a classroom curriculum designed to teach about self-control, feelings and relationships, and interpersonal cognitive problem solving. Student support teams reviewed student behavior and attendance warning signs to provide reinforcing incentives or allocate mental health resources early on. Planning centers prevented behavior escalation by providing a cool down space to use coping strategies from PATHS (e.g., replacement behaviors, social skills, anger management, safety procedures), either before receiving an exclusionary referral or upon returning to school from suspension or transfer. Researchers found that these alternatives were indeed associated with significantly reduced suspendable behavior incidents for medium/high implementers and a 58.8% decrease in OSS. Other variables that were positively affected included significantly better conditions for learning (e.g., school safety and fighting) for middle school students, as well as higher ratings of student social competence, student support, and student attendance (although no statistical significance reported). Given lack of recruitment data, rationale for data collection tools and analyses, stakeholder involvement, and critique of strengths and limitations, this article is lower in quality with a rating of 22 by reviewers.

Bohnenkamp et al. (2021) used an RCT design to evaluate Emotional and Behavioral Health – Crisis Response and Prevention (EBH-CRP) to more effectively address student crises, improve school safety, and reduce negative discipline outcomes over two years of implementation (2015-2017). This intervention involves comprehensive training and support for evidence-informed strategies that include the Safe School Ambassadors Program to train

students to intervene on bullying and violence (universal; mean implementation quality = 6.93/10), identifying students in need of increased psychological support (secondary), assessment and referral to emotional and behavioral interventions through a Mobile Crisis Facilitator (tertiary), a Standardized Crisis Response Protocol and Life Space Crisis Intervention (mean implementation quality = 6.63/10) for cognitive behavioral techniques for de-escalation (tertiary), and post-crisis relapse prevention (tertiary). Of the 20 comparison schools compared to 20 intervention schools, they had higher rates of Latinx and FRP students but lower AAPI and attendance rates. Intervention secondary schools showed a significantly slower rate of suspension and juvenile justice referral growth and greater decline rate for ODRs, yet these effects were not significant for African American or Latinx students, meaning disproportionality of suspensions for African American, Latinx, and multi-racial students did not decrease significantly. Primary schools showed slower rate of bullying growth and all schools using EBH-CRP had more frequent assessment of student reports of harm and decreasing threat assessments. Of note, this research maintained a high quality score of 34 with some involvement of stakeholders.

Table 7. List of MTSS Interventions and Significant Outcomes

Intervention Type	Significant Outcomes
SWPBIS	• Reduced ODR, ISS, OSS, behavior problems, expulsions, alternative school placement, law enforcement referrals • Improved fairness, relationships, SEL skills, safety, time for learning
Teacher empathy/bias training	• Fewer reprimands, ODRs, ISS, OSS • Greater school connectedness
Restorative circles (primary)	• Reduced ODR, OSS
Threat assessment	• Reduced bullying, teasing, OSS, expulsions, alternative school placement, referrals to law enforcement • Improved safety, school climate, access to counseling and parent conferences
Suspension alternatives	• Reduced ISS, OSS, expulsions • Improved relationships with staff (including SROs)
Restorative practices (secondary, multiple tiers)	• Reduced ODRs, detentions, disruptive behavior, OSS
PBIS + supplements	• Reduced ODR, ISS, OSS, expulsion • Improved relationships and self-esteem
SEL (PATHS)	• Reduced behavior incidents and OSS
Crisis Response and Prevention (EBH-CRP)	• Fewer ODR, OSS, and juvenile justice referrals

Research Question Three

The third aim of this systematic review was to highlight which settings (i.e., elementary, middle, and high school) and populations (e.g., underrepresented groups) these interventions were found to be effective in reducing exclusion. While not every intervention was assessed across all settings and populations, key themes are highlighted below.

Primary SWPBIS strategies were shown to work effectively to reduce exclusionary discipline and behavior problems and enhance prosocial outcomes in elementary schools especially. There was some research performed in secondary schools as well, but with lower

quality assessment for these studies and without effects for the highest severity punishments like expulsions and law enforcement referrals. These strategies were also shown to foster significant reductions in exclusionary outcomes within several samples of majority racial and ethnic minority students. Notable large-scale effectiveness studies were conducted in Maryland, Georgia, New Hampshire, and Florida. Multi-tier PBIS interventions, including those with adjunctive components, demonstrated success in reducing suspensions for primarily low-income schools and some majority racial/ethnic minority schools (e.g., Florida, California). However, disparities between Hispanic and Black youth and other racial counterparts for these interventions still persisted at times with variable quality for disparity research.

Studies involving Restorative Practices also had some evidence of narrowing discipline gaps based on demographic factors of race, gender, and special education status, as well as evidence of less referrals overall for misconduct or defiance when these practices were used. For the four types of teacher empathy and bias trainings, they were all implemented in majority minority, low-income schools with evidence of reduced reprimands, referrals, and suspensions for the most vulnerable youth across a variety of grade levels, as well as increased school connectedness for Black youth in one study. Social emotional skill development was also shown to be effective in more marginalized elementary school settings.

More often than not, secondary and tertiary interventions were evaluated in the secondary school context, likely due to the increased relevance of these strategies for older populations. Tertiary threat assessment reduced disproportionate rates of exclusionary outcomes for Black, Hispanic, and male students and was often successfully implemented in under-resourced school populations. There was some evidence of suspension diversion alternatives functioning

effectively in schools with high baseline disproportionality levels, suggesting potential for addressing biases that fuel disparities. Although no definitive statements can be made about which interventions are effective for which populations and settings, there are some notable trends that suggest where and for whom interventions may be more likely to be useful.

<div align="center">**Research Question Four**</div>

Finally, the fourth research question asks how these interventions are qualitatively perceived by implementers and recipients of the alternative strategies. Stakeholder opinions regarding positive effects and notable processes that reduce exclusion are discussed below.

Restorative Practices

Primary tier. While no quantitative studies were identified through the database search that assessed the quantitative effectiveness of primary prevention Restorative Practices (RP)**,** there were three studies which explicitly addressed qualitative perceptions of proactive RP circles. First, Lustick et al. (2020c) elicited opinions of teachers and students in three majority non-White, low-income secondary schools to assess the successes and challenges of implementing community-building circles as a way of reducing exclusion and disproportionality in discipline. Teachers had at least one year of RP training prior to data collection and conducted these circles in advisory class dedicated to social and emotional skill-building and check-ins. Overall, RP implementation was described as dependent on teacher-student relationships, teacher familiarity with RP, and support provided to teachers for implementation. Teachers felt that community-building circles were helpful for allowing students to share with and learn about each other, which in turn built trust and growth. For some staff and student participants, the RP philosophy was difficult to embody. Researchers noted that trauma, bias, and cultural

responsiveness were not included in the RP training, therefore making marginalized students more vulnerable and teachers felt overwhelmed. In these situations, community-building circles were implemented less consistently or effectively for shared meaning-making. When the circles worked well, a White female teacher noticed more reconciliatory language being used to avoid conflict in the circle and one male student noticed his own transformation into a more accepting, respectful individual. This evaluation of the process of RP community-building circles received a quality assessment score of 21, particularly lacking in its description of the interview tool and recruitment of interview participants.

Garnett et al. (2020) also assessed RP community-building circles specifically, but in an urban northeastern elementary school with a large refugee population in its first year of implementation in 2018. These circles mirrored the Lustick et al. (2020c) intervention in the elements of an RP circle (e.g., centerpiece, talking piece, circle prompts, opening/closing ceremonies), but were modified for elementary school students (e.g., sitting at eye level, acknowledging students' effort). Teachers reported using community-building circles on average three times per week and they noticed a relationship between use of these circles and student SEL skill development, such as responsible decision-making, social awareness, relationship building, and self-awareness. Seventy percent of 3rd-5th graders reported sharing their feelings and experiences in these circles. Teachers highlighted the increase in meaningful relationships they had with their students. Unfortunately, these RP circles clashed with the current discipline system when they were perceived as too passive or permissive. There was also concern brought up that the circles may be too frequent along with several other interventions for SEL and

behavior. Garnett et al. (2020) had very strong quality of reporting, methods, and evidence with a QuADS score of 37.

Using the same sample as Garnett et al. (2020), Kervick et al. (2020) also reviewed the process of the first school year of RP circle implementation (2017-2018) in an urban elementary school with a predominant refugee population in a northeastern state, but with a focus on the structural facilitators that enhanced implementation and how this fit in with existing discipline procedures. There were no systematic expectations for these RP circles in terms of their frequency or format (e.g., restorative question sequence) and staff were not told to alter previous discipline practices to align better with RP circles. Staff identified sustained professional development, a strong culture of collaboration, and the swift actions of leadership as facilitators of RP implementation. With regard to other interventions, teachers noted that RP circles were compatible with PBIS and SEL initiatives, but it was not clear how to describe this alignment to families and students. Lastly, staff voiced a need for a systematic approach to promote fidelity and reduce disciplinary referrals, such as aligning existing disciplinary structures with the RP philosophy beyond community-building circles, or creating a consistent curriculum across circles. Concerningly, there was a higher level of ODRs during this year of implementation compared to the three years prior. This study, like Garnett et al. (2020), had a very strong quality rating of 38.

Secondary tier. Secondary RP practices were also addressed in terms of the process by which they were carried out in schools. Ortega et al. (2016) explored experiences of 35 African American students and 25 primarily African American staff participating in Restorative Circles as a response to student conflict and disruptive behavior in a southeast urban high school in the

2012-2013 school year. Three nonprofit facilitators assist in preparing pre-Circles to build connections, identify needs, and obtain consent. Then the facilitator promotes dialogue between the parties involved while instilling reflection of each other's words and responsibility for their choices as well as agreements for moving forward. Post-Circles then assess how progress has been going towards the agreed upon actions. Students and teachers reported frustration and disappointment with students and peers who were lying or combative in circles rather than committing to the process or allowing themselves to be vulnerable. Oppositely, Restorative Circles seemed to promote a shift to less exclusionary methods of discipline by replacing suspensions with circles. They also reported better relationships of students to other students and teachers despite conflict. Students identified new ways of managing conflict with talking and staff noted honing their reflective listening skills, both of which were deemed important to make students feel heard and address the underlying issues. Students even noticed their own changes in daily use of Circles on their own, which was evident in staff's recognition of students' increased maturity. Quotes from participants highlighted the circles as "revolutionary" for reducing aggression between one pair of female students, feeling surprised at the new sense of kinship and trust they felt towards the other person, or noticing the power of truly listening and feeling understood. This qualitative study received a QuADS score of 32, in line with several rigorous quantitative works.

Lustick (2020a) describes a year-long (2013-2014) ethnographic multi-case study looking at field notes and semi-structured interviews at three NYC secondary schools using RP to reduce suspensions and transform the punitive culture. This implementation of culturally responsive RP meant holding the cultural, systemic, and interpersonal aspects of a conflict and promoting

understanding of individual, structural, and systemic harm. Students were referred to youth court, peer mediation, restorative conversation, or suspension/detention. The three cases investigated the cultural responsiveness of youth court, a restorative circle for chronic lateness and absence, and a restorative circle after an off-campus fight. The majority of students at each school were minorities and qualified for FRP, whereas majority of teachers were White. Several qualitative themes were identified, including reinforcement of belonging and accountability to a larger group which made the circle effective and supportive (e.g., having one's behaviors normalized by a peer with similar experiences) and White teachers seeking to improve their own understanding throughout each circle. The importance of critical consciousness was a frequent theme, such as assessing how a student's behavior relates to larger systems of power (e.g., violent or gendered language) or informing students of the systemic issues related to the school-to-prison pipeline and the subsequent value of conflict resolution. While all of the RP interventions prevented suspensions, culturally-relevant factors were not always taken into account, such as when a student's concerns were dismissed as psychosomatic or excuses in a circle without advice or help with solutions; making the circle feel more like a formality than a true restorative process therefore perpetuating implicit bias and inequality. Overall, there was commitment to reducing racial disproportionality in discipline but with high variability among the implementation of practices based on the educator's approach. This article received a score of 27 with high quality scores on most items except for recruitment, sampling, and stakeholder involvement.

Lastly, Bruhn (2020) focused on leadership qualities rather than reactive RP circles for an urban charter northeast high school using RP to reduce racial disparities in exclusion for

predominantly Black, low-income, and diverse language students. The authors aimed to see how two administrators (e.g., principal and RP coordinator) moved from zero tolerance policies to a restorative culture (e.g., advisory and responsive circles) in the 2016-2017 school year, given the challenges associated with this kind of transformation. The research showed restraint, persistence, and respect were essential for establishing networks of support for staff, upholding the responsibility to prioritize and educate others about relationships and restorative ethos, and valuing the perspectives and hard work of their students and teachers. The leaders scaffolded teachers, such as starting circles in advisory periods then transitioning to having them lead responsive circles. The study represented high quality with a score of 33 and only lacking in recruitment data and discussion of strengths and limitations.

Multiple tiers. There were four pieces of qualitative evidence identified in this review that covered all three tiers of implementation for RP. As discussed above, the SaferSanerSchools Model of RP was evaluated in Rainbolt et al. (2019) to assess teachers' perceptions and beliefs around implementation and efficacy in one high school in the mid-Atlantic in 2015-2016 with predominantly White and Black students. Out of 43 staff, 92% were trained in RP, half used RP often, and 19% used these practices daily, but only 3% said they did not use RP. All respondents indicated students had at least some knowledge or experience with RP, with 75% stating that only some students participated actively in practices. Specific practices were ranked in their relative efficacy from highest to lowest: affective statements, fair process, small impromptu conferences and circles, restorative questions, restorative conferences, restorative approach with families, proactive circles, restorative staff community, responsive circles, reintegrative management of shame, and understanding fundamental hypothesis. Based on Likert ratings,

teachers generally felt that RP curbed misbehavior, school culture reflected emphasis on respect and relationships, and teacher-student relationships were the most respectful. Barriers included teachers not being told about the restorative process happening, teachers perceiving students as manipulating the system to receive less punitive consequences, and significant patience and persistence in learning and applying strategies. Teachers expressed desire for staff to know the philosophy of RP and why it was important to reduce the school-to-prison pipeline. Overall, there was a downward trend in exclusion and discipline gaps for this school. Quality assessment was strong with a rating of 30 except for issues with justifying qualitative analyses and assessing strengths and weaknesses of the work, but did involve stakeholders in the process unlike most other studies.

Stinchcomb et al. (2006) explored how RP elements were implemented in three St. Paul, Minnesota elementary and middle schools. Schools reported reduced behavior referrals, suspensions, expulsions, and increased attendance from 1997 to 2001, but without comparison schools this could not be attributed solely to restorative justice implementation. Students shared positive views of circles in facilitating novel methods of conflict resolution, repairing of friendships, empathy-building, and equal treatment of participants. Staff expressed some concern with the intensive time for a restorative and patient problem-solving attitude and process (e.g., several circles). They also endorsed transformational changes in culture, such as reduced tensions between students, less violent and more control over behavior, and more positive language. Training in the restorative justice philosophy was provided to all staff, but often existing discipline policies prevented the use of RP for more serious offenses. Mentoring, support, visionary leadership, and reflective opportunities were highlighted as necessary to

promote the continued use of new RP strategies by staff, make sure they were consistently and holistically applied in the classroom and school, and enhance staff cohesion. Despite the positive outcomes of this case study, its quality is in the lower range with a score of 16 due to limited discussion of the population, data collection and analysis procedure and justification, and stakeholder involvement.

To further understand how RP can culturally transform three NYC secondary schools, Lustick (2020b) assessed what promoted the use of punitive over RP by principals in 2013-2014. These schools had the lowest suspension rates of NYC schools with majority students of color qualifying for FRP lunch. Restorative coordinators and deans represented both restorative and traditional discipline policies. The author found that RP reduced the need for suspension but that RP did not permeate the school culture given that principals still prioritized symbolically maintaining the perception of order to prevent criticism from stakeholders and promote accountability, rather than building relationships, trust, and shared responsibility. Suspensions were used to prevent or respond to physical fights, but could also be applied to lower-level behaviors. This study received a quality score of 28 due to limited recruitment data, no stakeholder involvement, and no assessment of strengths and limitations.

Lastly, Sandwick et al. (2019) interviewed a range of stakeholders at five diverse NYC middle and high schools ranging in one to five years of implementation during the 2016-2017 school year to see what they perceived to be key practices for building schoolwide RP culture, perceptions of these practices, strategies for implementation, and challenges to implementation. Overall, some major themes that emerged included the importance of strong community ties among staff, students, and families, sufficient resources and infrastructure, shift in school

hierarchy and power dynamics, and supporting student leadership to foster buy-in and effective RP processes. Most interviewees preferred RP to punishment-based models and perceived decreased suspensions and increased attendance. Benefits they cited included conflict resolution and de-escalation, addressing root causes of conflict, learning from mistakes, minimizing future harm, limiting negative effects of punishment, and promoting empathy, relationships, and accountability. However, some interviewees noted stakeholder perceptions of RP as "soft" or "enabling" without consequences, but exposure to RP and communication helped implementers understand it as a method to address incidents and eventually become proponents. This research has strong quality across all domains (QuADS = 34) except for more limited discussion of strengths and weaknesses of the evidence.

Relationship-Building Strategies

Working in a bottom-up fashion, Anyon et al. (2018) used a phenomenological approach to identify instead 33 low-suspending Denver public schools and elicit from 198 educators the discipline strategies they used to achieve these low suspension rates overall (3%) and for Black students. Students were majority students of color with FRP, whereas teachers were predominantly female and White. Across the proactive strategies highlighted in the interviews for reducing racial disparities in exclusion, a major theme included building relationships with students through home visiting, greetings, morning meetings, advisory periods, increased adult visibility off and on campus, and contact with families. Educators also reported that once these relationships were built, they became more aware of students' strengths, triggers, coping resources, and challenges, which in turn helped illuminate the motivation for their misbehavior, such as a reading disorder, unstructured classroom, or trauma. As a result, consequences were

better tailored to the individual and became a means for growth rather than solely punishment. Quotes revealed efforts to involve parents in promoting positive behaviors and signify to students that they care. Again, this qualitative research was rated very high in quality at a QuADS score of 36.

ISS as an OSS Alternative

Although ISS is also an exclusionary practice, the systematic search revealed one qualitative study where it was used to reduce the degree of exclusion faced by students with disruptive behavior problems. Turpin et al. (1997) performed an observational case study to better understand a new ISS program in a rural high school in Louisiana for improving discipline outcomes. In this way, the ISS served as a positive alternative by giving students an opportunity to complete class assignments instead of being unsupervised in the community for a day or more. Of the 47 students that had been in ISS once and the 40 that had ISS more than once, these students and their teachers felt that the ISS program helped reduce OSS and student misbehavior by removing those students from the classroom and students changing their behavior in order to prevent returning. However, there were no reductions in lost instructional days or OSS with the implementation of the new ISS program and students often used this time for sleeping. Given the moderate quality score of 21 and limited reporting quality of the data collection process and evidence, these findings should be interpreted with caution.

Table 8. Quality Assessment Ratings by item

Lead author (year)	1	2	3	4	5	6	7	8	9	10	11	12	13	Total
Turpin (1997)	2	2	3	3	0	1	2	1	2	2	3	0	0	21
Breunlin (2002)	3	3	3	3	3	3	3	3	1	2	3	0	3	33
Scott (2004)	2	2	3	3	1	3	3	2	0	3	3	1	3	29
Luiselli (2005)	2	1	3	3	1	2	3	3	1	0	1	0	3	23
Breunlin (2006)	2	0	3	0	0	1	0	1	1	0	0	3	0	11
Stinchcomb (2006)	3	3	1	2	0	2	1	0	0	1	1	0	2	16
Mirsky (2007)	1	0	3	0	0	0	1	0	0	0	0	0	0	5
Barnhart (2008)	2	1	2	1	3	1	2	1	0	0	1	0	1	15
Muscott (2008)	3	3	3	3	3	3	3	2	2	3	3	0	3	34
Cornell (2009)	3	3	3	3	3	3	3	3	3	2	3	0	3	35
Bradshaw (2010)	3	3	3	3	3	3	3	3	1	3	3	0	3	34
Morrissey (2010)	2	1	3	1	0	0	1	1	0	0	1	3	1	14
Cornell (2011)	3	3	2	3	3	3	3	3	0	3	3	0	3	32
Bradshaw (2012)	2	3	1	3	2	3	3	3	3	3	3	0	3	32
Chin (2012)	3	2	2	3	1	1	1	0	0	3	3	1	2	22
Cornell (2012)	3	3	3	3	2	2	3	2	1	3	3	0	2	30
Simonsen (2012)	2	3	2	2	3	3	3	3	0	3	3	0	3	30
Mitchell (2013)	3	3	3	3	2	3	3	3	3	3	3	0	3	35
Teske (2013)	2	2	1	1	1	0	0	0	0	0	1	1	0	9
Ward (2013)	3	3	2	3	3	3	3	2	1	3	3	1	3	33
Anyon (2014)	3	3	3	3	2	3	3	1	1	3	3	0	3	31
Gregory (2014)	3	3	2	3	1	3	3	1	1	3	3	0	3	29
Nocera (2014)	2	1	2	2	3	2	2	2	0	2	1	1	2	22
Osher (2014)	3	2	3	3	2	1	2	2	1	1	2	0	0	22
Nekvasil (2015)	3	3	3	3	3	3	3	3	3	3	3	0	3	36
Anyon (2016)	3	3	3	3	3	3	3	3	0	3	3	0	3	33
Childs (2016)	3	3	2	3	1	3	3	3	0	3	3	0	3	30

Gregory (2016)	3	3	2	3	1	3	2	2	1	3	3	0	3	29
Okonofua (2016)	2	1	2	3	0	1	1	1	2	1	2	0	0	16
Ortega (2016)	3	3	2	3	2	3	3	2	1	3	3	1	3	32
Smolkowski (2016)	3	3	2	3	2	3	3	3	1	3	3	1	3	33
Anyon (2018)	3	3	3	3	3	3	3	3	1	3	3	3	2	36
Cook (2018)	3	3	2	3	2	3	3	1	3	3	3	3	3	35
Cornell (2018)	3	3	3	3	3	3	3	3	3	3	3	0	3	36
Cruz (2018)	2	3	2	3	2	3	2	3	1	3	3	0	3	30
Gage (2018)	3	3	3	3	3	3	3	2	1	3	3	0	3	33
Gregory (2018)	3	3	3	3	3	3	3	3	3	3	3	0	3	36
Hashim (2018)	3	3	1	2	3	3	2	3	0	3	3	0	3	29
Kim (2018)	3	3	2	3	3	3	3	3	3	3	3	0	2	34
Mansfield (2018)	3	2	3	2	3	1	2	1	0	1	3	0	2	23
Freeman (2019)	3	3	2	3	1	2	3	3	1	3	3	0	3	30
Gage (2019)	3	3	3	3	3	3	3	3	3	3	3	0	3	36
Rainbolt (2019)	3	3	2	3	2	2	3	3	2	1	3	2	1	30
Sandwick (2019)	3	3	3	3	3	3	3	2	2	3	3	2	1	34
von der Embse (2019)	3	3	3	2	2	2	3	2	0	1	1	1	3	26
Baule (2020)	1	1	3	3	0	0	2	1	0	0	2	0	1	14
Bruhn (2020)	3	3	3	2	3	3	3	3	1	3	2	3	1	33
Garnett (2020)	3	3	2	3	3	3	3	3	2	3	3	3	3	37
Gion (2020)	3	3	3	3	2	2	3	3	3	1	3	0	3	32
Huang (2020)	3	3	2	3	0	3	3	1	0	3	3	0	2	26
Kervick (2020)	3	3	3	3	3	3	3	3	3	3	3	3	2	38
Lustick (2020a)	3	3	3	3	1	2	2	2	1	2	2	0	3	27
Lustick (2020b)	3	3	2	2	3	3	3	2	1	3	3	0	0	28
Lustick (2020c)	3	2	3	3	2	0	2	1	0	1	3	0	1	21
Maeng (2020)	3	2	3	2	3	2	3	2	3	3	3	0	2	31
Nese (2020)	3	3	3	3	2	3	3	2	0	2	3	0	3	30
Bohnenkamp (2021)	3	3	2	3	3	2	3	2	3	3	3	1	3	34
Lee (2021)	3	3	3	3	3	3	3	3	3	3	3	0	3	36
Zhang (2021)	3	3	2	2	3	3	3	2	3	3	3	0	3	33

Table 9. Study Demographics

First author (year)	Sample N	Grade level	Race/ethnicity	Other Demographics	School type	School setting
Turpin (1997)	1 school 233 students 15 staff	HS	Students (pop., $n = 364$): 57% Black 42% White 1% Other	Population: 48% male 32% in poverty	NR	Rural
Breunlin (2002)	1 school 165 students	HS	74% Caucasian 12% Hispanic/Latino 10% AA 2% Asian/PI 2% Other	82% male	Public	Suburban
Scott (2004)	1 school	ES	NR	NR	NR	Urban
Luiselli (2005)	1 school 660 students 30 staff	ES	Students: 88% Black 5% White 4% Hispanic 2% Asian/PI 1% other	90% FRP 11% SpEd 10% ESL	Public	Urban
Breunlin (2006)	1 school 3,700 students (pop.)	HS	Population: 90% White 6% Hispanic 3% AA 1% Other	NR	Public	Suburban

Stinchcomb (2006)	3 schools Students, staff NR	ES, MS	NR	NR	Public	Urban
Mirsky (2007)	3 schools 2,146 students (pop.) Staff NR	MS, HS	NR	NR	NR	Urban, Suburban

Barnhart (2008)	59 targeted schools
	58 compar. schools
		MS,
		HS
			Target schools:
			  71.2% Latino
			  17.0% AA
			  7.2% White
			  2.7% Asian
			  1.9% Other
			
			Comparison schools:
			  66.3% Latino
			  11.8% AA
			  12.3% White
			  5.6% Asian
			  4.0% Other
				Target schools:
				  51.0% male
				  34.9% ELL
				  74.8% FRP
				  12.9% SWD
				
				Comparison schools:
				  50.5% male
				  29.2% ELL
				  69.8% FRP
				  12.0% SWD
					Public	Urban

Muscott (2008)	27 schools
	15,386 students
		ES,
		MS,
		HS
			86.8-99.5% Caucasian
			  ($M = 95.7$, $SD = 2.8$)
			0.0-7.0% Hispanic
			  ($M = 1.2$, $SD = 1.3$)
			0.0-6.3% Asian
			  ($M = 1.6$, $SD = 1.4$)
			0.0-4.0% AA
			  ($M = 1.2$, $SD = 1.0$)
			0.0-1.3% NA
			  ($M = 0.3$, $SD = 0.4$)
				2.1-47.5% FRP
				  ($M = 23.9$, $SD = 13.0$)
				0.0-9.6% LEP
				  ($M = 1.6$, $SD = 2.4.0$)
					Public	NR

Cornell (2009)	280 schools 7,318 students	HS	63% White/Caucasian 23% Black/AA 5% Latino/Hispanic 3% Asian American 1% AI 5% Other	Sample: 51% male Population, $n = 335{,}720$: 0-100% FRP ($M = 34$, $SD = 26$)	Public	Urban, Suburban, Rural
Bradshaw (2010)	21 targeted schools 16 compar. schools 17,995 students Staff NR	ES	Target schools: 53.81% White (M) ($SD = 33.16$) Comparison schools: 67.51% White (M) ($SD = 28.99$)	Target schools: 42.93% FRP (M) ($SD = 19.22$) 13.24% SpEd (M) ($SD = 4.27$) Comparison schools: 36.25% FRP (M) ($SD = 20.93$) 15.08% SpEd (M) ($SD = 6.66$)	Public	Suburban, Rural
Morrissey (2010)	1 school Approx. 1,800 students	HS	NR	NR	Public	Urban
Cornell (2011)	23 targeted schools 26 compar. schools	HS	48% minority (M)	20% FRPM (M)	Public	NR
Bradshaw (2012)	37 schools 12,344 students	ES	46.1% White 45.1% AA 4.3% Asian/PI 3.9% Hispanic 0.6% AI/AN	52.9% male 49.4% FRP 12.9% SpEd	Public	Urban fringe, Suburban, Rural

Chin (2012)	1 school
	9 students
		ES	100% Latino/a	Sample:
				  0% SpEd
				  100% ELL
				
				Population, $n = 553$:
				  92% FRP
				  79% ELL
					NR	NR

Cornell (2012)	40 schools
	201 students
	59 staff
		ES, MS, HS	Students:
			  24% White
			  73% AA
			  3% Hispanic
				District population:
				  46% FRP
					Public	Urban, Suburban

Simonsen (2012)	428 schools	ES, MS, HS	NR	NR	Public	Urban, Suburban, Rural

Mitchell (2013)	21 targeted schools
	16 compar. schools
	1,902 students
	93 staff
		ES	Students:
			  40.2% White
			  33.5% Black
			  26.3% Other
			Staff 87.1% White
				51.2% male
				2.7% ELL
					Public	Urban fringe, Suburban, Rural

Teske (2013)	NR	ES, MS, HS	NR	NR	NR	NR
Ward (2013)	32 schools	ES	Students: Approx. 87% minority	Approx. 90% FRP	Public	Urban

Anyon (2014)	183 schools 87,997 students	ES, MS, HS	58% Latino 20% White 15% Black 3% Asian <1% PI 3% Multiracial	49% male 42% Non-native English 67% FRP 2% Homeless 12% SpEd 1% ED	Public	Urban
Gregory (2014)	5 schools 979 students 82 staff	MS, HS	Students: 59% AA 30% White 8% Hispanic 3% Asian Staff: 62% White 28% Black	Population: 20-40% FRP	NR	NR
Nocera (2014)	1 school Approx. 300 students 6 staff	MS	Students (pop., n ≈ 750): 40% minority	Population >50% FRP 40% minority 15% SpEd	Public	NR
Osher (2014)	100 schools 41,000 students	ES, MS, HS	68.0% AA 14.6% White 13.2% Latino	District population: 100% FRP 7% Homeless >33% Poverty-driven mobility	Public	Urban

Nekvasil (2015)	332 schools 29,203 students 6,298 staff	MS	Students: 51% White 20% Black 3% Asian 2% AI/AN 16% Multiracial 8% Other 13% Hispanic/Latino	Population: 2-99% FRP ($M = 44$, $SD = 20.5$)	NR	Urban, Suburban, Rural
Anyon (2016)	180 schools 9,921 students	ES, MS, HS	58.7% Latino 25.2% Black 10.6% White 1.5% Asian 0.99% NA 0.14% PI 2.9% Multiracial	65.8% male 84.8% FRP 37.8% ELL 20.28% SpEd 4.0% ED	Public	Urban
Childs (2016)	1,122 schools 946,968 students	ES, MS, HS	NR	1-100% FRP ($M = 54.88$, $SD = 31.79$)	NR	NR
Gregory (2016)	2 schools 412 students 29 staff	HS	Students: 44% White 21% Latino 5% AA 3% AI 2% Asian 25% Mixed Race Staff 100% White	53% male	NR	Urban

Okonofua (2016)	5 schools
	1,682 students
	31 staff
		MS	Students (pop., n = 3,027):
			  60% Latino
			  17% Asian
			  15% White
			  3% Black
			  5% Other
				Population:
				  59% FRP
					NR	NR

Ortega (2016)	1 school 35 students 25 staff	HS	All students and majority of staff identified as AA	43% male	NR	Urban
Smolkowski (2016)	74 schools	ES, MS, HS	73-87% minority	68-90% FRP	Public	Urban

Anyon (2018)	33 schools
	198 staff
		ES,
		MS,
		HS
			73% White
			10% Black
			12% Latino
			5% Other
				Population:
				  56% FRP
				  14% LEP
				  9% SWD
					Public,
					Charter,
					Innovation
						Urban

Cook (2018)	3 schools
	869 students
	40 staff
		ES	Students:
			  12-66% AA
			    ($M = 31$, $SD = 30$)
			  12-52% Latino
			    ($M = 26$, $SD = 22$)
			  2-51% White
			    ($M = 24$, $SD = 25$)
			  2-11% Asian
			    ($M = 8$, $SD = 5$)
			Staff 82% White
				26-96% FRP
				  ($M = 64$, $SD = 35$)
					Public	Urban

Cornell (2018)	779 schools 1,836 students	ES, MS, HS	51% White 31% Black 7% Hispanic 3% Asian 9% Other	Sample: 75% male 35% SpEd Population: 44.26% FRP (*M*) (*SD* = 24.01)	Public	Urban, Suburban, Rural
Cruz (2018)	41 schools 59,675 students	ES, MS, HS	52.9% Latino 25.7% White 15.8% Asian 3.9% AA <1% AI/AN	45.1% FRP 25.5% ELL 8.9% IEP/SpEd	Public, Alternative (2), Early Learning Center (1)	Urban, Suburban, Rural
Gage (2018)	119 target schools 119 compar. schools 125,795 students	ES	Target schools: 44.1% White (*M*) (*SD* = 29.6) 32.7% Black (*M*) (*SD* = 27.1) 16.2% Latino/a (*M*) (*SD* = 15.8) Comparison schools: 47.3% White (*M*) (*SD* = 29.4) 29.4% Black (*M*) (*SD* = 27.9) 17.3% Latino/a (*M*) (*SD* = 22.3)	Target schools: 68.9% FRP (*M*) (*SD* = 23.6) 11.2% SpEd (*M*) (*SD* = 3.1) 12.2% LEP (*M*) (*SD* = 15.3) Comparison schools: 69.0% FRP (*M*) (*SD* = 27.6) 11.5% SpEd (*M*) (*SD* = 3.5) 12.0% LEP (*M*) (*SD* = 19.1)	Public	NR

Gregory (2018)	5 schools 9,039 students	ES, MS, HS	57.7% Latino 24.9% Black 11.4% White 1.4% Asian 0.9% NA 0.2% PI 3.6% Multiracial	68% male 39% ELL 87% FRP 21% SpEd 4% ED	Public	Urban
Hashim (2018)	785 schools	ES, MS, HS	NR	NR	Public	Urban
Kim (2018)	477 schools	ES, MS, HS	43.85% minority (M) ($SD = 30.59$)	53.65% FRP (M) ($SD = 23.92$)	NR	Urban, Suburban, Rural
Mansfield (2018)	1 school 1,400 students (pop.) Staff NR	HS	Students (population): 54% White 32% Black Staff NR	NR	Public	Suburban
Freeman (2019)	15 schools 12,127 students	HS	47.6% AA 24.0% Hispanic 10.9% White 6.0% Asian 1.5% NA	52.0% male 65.9% FRP 21.6% SpEd	NR	Urban, Suburban, Rural

Gage (2019)	593 target schools 593 compar. schools	ES, MS, HS	Target schools: 48.9% White (*M*) (*SD* = 25.2) 20.7% Black (*M*) (*SD* = 21.4) 23.7% Hispanic (*M*) (*SD* = 18.6) Comparison schools: 48.7% White (*M*) (*SD* = 27.9) 21.4% Black (*M*) (*SD* = 24.2) 23.5% Hispanic (*M*) (*SD* = 20.8)	Target schools: 51.7% male (*M*) (*SD* = 3.2) 60.7% FRP (*M*) (*SD* = 22.1) Comparison schools: 51.6% male (*M*) (*SD* = 2.9) 60.0% FRP (*M*) (*SD* = 24.0)	Public	Urban, Suburban, Rural
Rainbolt (2019)	1 school 43 staff	HS	NR	NR	Public	Urban, Suburban
Sandwick (2019)	5 schools 44 students 32 staff	MS, HS	Students (pop., *n* = 2,052) 26-73% Latinx (*M* = 51, *SD* = 21) 23-70% Black (*M* = 44, *SD* = 22) Staff NR	Population: 38-54% male (*M* = 48, *SD* = 7) 74-92% in poverty (*M* = 82, *SD* = 7) 2-25% ELL (*M* = 9, *SD* = 10) 20-33% SpEd (*M* = 25, *SD* = 5)	Public	Urban
von der Embse (2019)	1 school 570 students (pop.)	ES, MS	District population nearly 80% minority	District population nearly 90% FRP	Public	Urban

Baule (2020)	4 schools 3,895 students	MS, HS	63.3-65.6% White 20.6-21.6% Black 2.4-3.5% Hispanic 0.6-0.8% Asian 0.3-0.5% NA 9.6-10.4% Multiracial	74.6-75.8% FRP	Public	Urban
Bruhn (2020)	1 school 2 staff	HS	Students (pop., $n \approx 1,100$) 55% Black 21% Hispanic 13% White 8% Asian 3% Mixed/Other Staff: 1 White, 1 Black	Population approx. 2/3 low income, nearly 1/2 from homes that speak languages other than English	Charter	Urban
Garnett (2020)	1 school 107 students 14 staff	ES	Students (pop., $n = 300$) 50% minority Staff NR	Population: 41% FRP 24% ELL	NR	Urban
Gion (2020)	2 schools 785 students (pop.) 4 staff	ES, MS	Students: 49% White 20% AA 17% Hispanic/Latinx 1% Asian Staff NR	59% FRP	Public	Urban
Huang (2020)	116 schools 30,799 students	MS, HS	57.9% Latino 20.7% White 13.4% Black 3.5% Asian 4.4% Other/multiracial	49.8% male 67.7% FRP 9.4% SpEd	Public	Urban

Kervick (2020)	1 school 17 staff	ES	Students (pop., $n = 296$) 50% minority Staff NR	Population: 41% FRP 24% ELL	Public	Urban
Lustick (2020a, b, c)	3 schools 1,160 students (pop.)	MS, HS	Students (population): 28.0-57.5% Hispanic ($M = 43.8, SD = 14.9$) 16.9-44.7% Black ($M = 33.6, SD = 14.7$) 1.8-15.1% White ($M = 7.9, SD = 6.7$) 0.8-37.5% Asian ($M = 13.3, SD = 21.0$) 0.8-2.5% Asian ($M = 1.4, SD = 0.9$)	67.1-84.9% FRP ($M = 73.2, SD = 10.2$)	Public	Urban
Maeng (2020)	527 schools 1,138 students	ES, MS, HS	NR	NR	Public	Urban, Suburban, Rural
Nese (2020)	2 schools 1,134 students 17 staff	MS	Students 30% minority	77% FRP	Public	Suburban, Rural

| Bohnenkamp (2021) | 20 target schools 20 compar. schools 28,726 students | ES, MS, HS | Target schools: 52.73% White (*M*) (*SD* = 12.15) 25.06% AA/Black (*M*) (*SD* = 12.03) 6.14% Latinx (*M*) (*SD* = 2.38) 4.88% Multiracial (*M*) (*SD* = 1.22) Comparison schools: 45.65% White (*M*) (*SD* = 21.15) 32.90% AA/Black (*M*) (*SD* = 17.93) 10.36% Latinx (*M*) (*SD* = 7.28) 5.54% Multiracial (*M*) (*SD* = 3.17) | Target schools: 44.4% FRP (*M*) (*SD* = 12.9) Comparison schools: 67.0% FRP (*M*) (*SD* = 18.3) | Public | Urban, Suburban, Rural |

Lee (2021)	112 target schools	ES	Target schools:	Target schools:	Public	Urban,
	112 compar. schools		43.55% White (*M*)	12.02% IEP (*M*)		Suburban,
	151,683 students		(*SD* = 29.11)	(*SD* = 1.57)		Rural
			32.02% AA/Black (*M*)	8.80% LEP (*M*)		
			(*SD* = 26.35)	(*SD* = 11.29)		
			16.94% Latino/a (*M*)	69.16% econ. dis. (*M*)		
			(*SD* = 15.90)	(*SD* = 23.17)		
			Comparison schools:	Comparison schools:		
			41.65% White (*M*)	11.95% IEP (*M*)		
			(*SD* = 41.65)	(*SD* = 4.28)		
			32.76% AA/Black (*M*)	8.43% LEP (*M*)		
			(*SD* = 26.67)	(*SD* = 12.34)		
			17.48% Latino/a (*M*)	72.34% econ. dis. (*M*)		
			(*SD* = 19.19)	(*SD* = 29.28)		
Zhang (2021)	106 schools	ES	18-99.9% minority	2.7-49.0% ELL	NR	NR
	54,166 students (pop.)		(*M* = 60.31)	(*M* = 16.81)		
				3-47.1% SpEd		
				(*M* = 14.96)		

Note. NR = not reported; ES = elementary school, MS = middle school, HS = high school; AA = African American, NA = Native American, PI = Pacific Islander, AI = American Indian, AN = Alaskan Native; ELL = English language learner, LEP = limited English proficiency, FRP = free and reduced-price meals (proxy for low-income), SWD = students with disabilities, IEP = Individualized Educational Program, SpEd = special education, ED = emotional disturbance, Econ. dis. = economic disadvantage.

Table 10. Study Methods

First author (year)	Intervention	Implementers	Dosage	Tier	Study Design
Turpin (1997)	ISS program	Secretary, assistant principal, or principal	1 year	2	Qualitative
Breunlin (2002)	Alternative to Suspension for Violent Behavior	Trainers at The Family Institute	1.5 years 6hr program	3	Quasi-experimental
Scott (2004)	PBS	School PBS team, staff	2 years	1	Quasi-experimental
Luiselli (2005)	Whole-school PBS	Behavior support teams (teachers, administrators, other school personnel)	1.5 years	1	Quasi-experimental
Breunlin (2006)	Making the Smart Choice	Trainers at The Family Institute	4 years 8hr program	2	Quasi-experimental; Case study
Stinchcomb (2006)	RJ	School staff	3 years	1, 2, 3	Quasi-experimental; Case study
Mirsky (2007)	SaferSanerSchools (model of RP)	School staff	Varied by school, 1-3 years	1, 2, 3	Quasi-experimental; Qualitative
Barnhart (2008)	SWPBIS	School discipline teams, teachers	2 years	1	Quasi-experimental
Muscott (2008)	SWPBIS	School-based Universal Leadership Teams, internal PBIS coaches, school faculty	2 years	1	Quasi-experimental

Cornell (2009)	Virginia Student Threat Assessment Guidelines	School-based threat assessment teams	NR, variable by school	3	Correlational
Bradshaw (2010)	SWPBIS	School SWPBIS teams, external behavioral support coaches, staff	4 years	1	Randomized controlled trial
Morrissey (2010)	Schoolwide PBS	School PBS team, staff	2 years	1	Quasi-experimental
Cornell (2011)	Virginia Student Threat Assessment Guidelines	School-based threat assessment teams (school administrator, law enforcement officer or SRO, and mental health professionals)	Varied by school, 1-2 years	3	Quasi-experimental
Bradshaw (2012)	SWPBIS	School SWPBIS teams, external behavioral support coaches, staff	4 years	1	Randomized controlled trial
Chin (2012)	Alternatives to Suspensions program, within PBIS school	A PBIS team including the school psychologist, school psychology graduate students, and university faculty	1 year	3	Quasi-experimental; Case study
Cornell (2012)	Virginia Student Threat Assessment Guidelines	School-based threat assessment teams (school administrator, law enforcement officer or SRO, and mental health professionals)	NR, variable by school	3	Randomized controlled trial
Simonsen (2012)	SWPBS	School SWPBS team (faculty, staff, parents, and students), external coaches	7 years	1, 2, 3	Quasi-experimental

Mitchell (2013)	SWPBIS	School SWPBIS teams, external behavioral support coaches, staff	4 years	1	Correlational
Teske (2013)	School Referral Reduction Protocol; Collaborative Child Study Team	Mental health professional, school social worker and counselor, social services professional, juvenile court officer, child service providers, court provided facilitator	NR	3	Correlational
Ward (2013)	Safe and Civil Schools (model of SWPBIS)	School a leadership teams (administrator, 3+ gen. ed. teachers, 1 SpEd teacher, 1-2 other personnel)	Varied by cohort, 1-2 years	1	Randomized controlled trial
Anyon (2014)	Restorative approaches; behavior contracts; ISS	NR, variable by school/practice	NR, variable by school/practice	1, 2, 3	Correlational
Gregory (2014)	My Teaching Partner−Secondary	Classroom teachers	1 year	1	Randomized controlled trial
Nocera (2014)	SWPBS	Teachers, school staff	2 years	1, 2, 3	Quasi-experimental
Osher (2014)	Promoting Alternative Thinking Strategies (PATHS); student support teams; planning centers	Classroom teachers, administrators, support staff members, planning center instructional aides	Varied by school and program	1, 2, 3	Quasi-experimental
Nekvasil (2015)	Virginia Student Threat Assessment Guidelines	School-based threat assessment teams	NR, variable by school	3	Quasi-experimental

Anyon (2016)	RI	School staff	1 year	2	Quasi-experimental
Childs (2016)	SWPBIS	School staff	Varied by school, $M = 3.71$ years $(SD = 0.98)$	1	Correlational
Gregory (2016)	RP (SaferSanerSchools)	RP trainers, teachers, administrators, other staff	1 year	1, 2, 3	Correlational
Okonofua (2016)	Empathic mindset	Math teachers	1 year, 70 mins modules	1	Randomized controlled trial
Ortega (2016)	RP restorative circles	External Restorative Circle program staff	2 years	2	Qualitative
Smolkowski (2016)	Safe and Civil Schools (model of SWPBIS)	School leadership teams (administrator, 3+ gen. ed. teachers, 1 SpEd teacher, 1-2 other personnel)	2 years	1	Randomized controlled trial
Anyon (2018)	Proactive relationship-building strategies	School administrators, teachers, social workers, psychologists	10 years of discipline reform	1	Qualitative
Cook (2018)	Greet−Stop−Prompt within PBIS school	Teachers, administrators, professional support staff	5 months	1	Quasi-experimental
Cornell (2018)	Threat assessment	NR, variable by school	NR, variable by school	3	Correlational
Cruz (2018)	SWPBIS	NR, variable by school	NR, variable by school	1, 2, 3	Correlational
Gage (2018)	SWPBIS	NR, variable by school	NR, variable by school	1	Quasi-experimental

Gregory (2018)	RI	NR, variable by school	NR, variable by school	2	Correlational
Hashim (2018)	RJ, suspension ban for willful defiance in SWPBIS schools	District-wide ban; RJ/SWPBIS implementers NR	SWPBIS, 9 years; suspension ban, 3 years; RJ, 1 year	1, 2, 3	Quasi-experimental
Kim (2018)	SWPBIS	NR, variable by school	Varied by school, 39% 0-2 years 61% 3+ years	1	Correlational
Mansfield (2018)	SaferSanerSchools (model of RP)	Teachers, administrators	Gradual increase over 6 years	1, 2, 3	Quasi-experimental
Freeman (2019)	PBIS	NR, variable by school	NR, variable by school	1, 2, 3	Correlational
Gage (2019)	SWPBIS	NR, variable by school	Varied by school, 1-10 years	1, 2, 3	Quasi-experimental
Rainbolt (2019)	SaferSanerSchools (model of RP)	Teachers, administrators	5 years	1, 2, 3	Qualitative
Sandwick (2019)	RJ	Staff facilitators, RJ coordinators, social workers, counselors, administrators, external community-based partners, teachers, student leaders, peer mediators	Varied by school, 1-5 years	1, 2, 3	Qualitative
von der Embse (2019)	Trauma-informed PBIS	Teachers, PBIS team, school therapeutic support staff	2 years	1, 2	Quasi-experimental

Baule (2020)	SWPBIS	NR, variable by school	2 years	1	Quasi-experimental
Bruhn (2020)	RJ	RJ coordinator, student support team, teachers, principal	1 year	2	Qualitative
Garnett (2020)	RP community building circles	Classroom teachers, student "circle keepers"	1 year	1	Quasi-experimental, qualitative
Gion (2020)	Classroom Check-up	Classroom teachers	Varied by participant, 1.5-8 weeks	1	Quasi-experimental
Huang (2020)	RP; ISS	NR, variable by school/practice	NR, variable by school/practice	2	Correlational
Kervick (2020)	RP	Teachers/staff	1 year	1	Qualitative
Lustick (2020a)	RP	Restorative coordinator, school social worker, teachers	1 year	2	Qualitative
Lustick (2020b)	RP	Restorative coordinators, administrators, teachers	1 year	1, 2, 3	Qualitative
Lustick (2020c)	RP	Teachers	1 year	1	Qualitative
Maeng (2020)	Comprehensive Student Threat Assessment	School-based threat assessment teams (administrator, law enforcement officer or SRO, mental health professionals)	NR, variable by school	3	Correlational

Nese (2020)	Inclusive Skill-Building Learning Approach, within PBIS school	Educational support staff (e.g., educational assistants, behavioral support staff)	1 year	1, 2, 3	Quasi-experimental, qualitative
Bohnenkamp (2021)	Emotional and Behavioral Health–Crisis Response and Prevention	School staff, mobile crisis facilitator	2 years	1, 2, 3	Randomized controlled trial
Lee (2021)	SWPBIS	NR, variable by school	NR, variable by school	1	Quasi-experimental
Zhang (2021)	SWPBIS	NR, variable by school	Varied by school, 2-3 years	1	Correlational

Note. PBS = positive behavior support, PBIS = positive behavior interventions and supports, SWPBIS = school-wide positive behavior interventions and supports, RJ = restorative justice, RP = restorative practices, RI = restorative interventions, RC = restorative circles, ISS = in-school suspension.

Table 11. Study Outcomes

First author (year)	Intervention	Outcomes Measured	Type of Measurement	Data Collection Timepoints	Major Findings
Turpin (1997)	ISS program	Student/teacher perceptions of ISS program	Staff structured interviews; student questionnaires	1995-1996	Perceived lower OSS/misbehavior, but no objective reduction
Breunlin (2002)	Alternative to Suspension for Violent Behavior	OSS; disciplinary acts (total detentions and ISS)	School disciplinary data	1997-1998 1998-1999 (fall semester)	Fewer expulsions; no sig. difference in disciplinary acts or any types of OSS
Scott (2004)	PBS	Disciplinary incidents (ODR, suspensions); administrator time; instructional time; monetary cost	School disciplinary data; time analyses of past discipline records; district fiscal data	1999-2000 (baseline) 2000-2001 (year 1) 2001-2002 (year 2)	Decrease in ODR and suspensions; resulting increase in time/cost savings
Luiselli (2005)	Whole-school PBS	ODR; suspensions; academic performance; social validity	School disciplinary data; test scores; social validity questionnaires	Aug. 1999 - Jan. 2000; monthly (pre-int.) Feb. 2000- Jun. 2001; monthly (intervention) Aug. 2001 - Apr. 2002; monthly (follow-up)	Decreases in student discipline problems; improvements in academic performance

Breunlin (2006)	Making the Smart Choice	OSS; suspensions for physical violence	School disciplinary data	1993-1994 through 1996-1997; yearly (pre-int.) 1997-1998 through 2000-2001; yearly (intervention)	Decrease in total OSS and suspensions for physical violence
Stinchcomb (2006)	RJ	Suspensions; expulsions; behavior referrals; attendance	School disciplinary data	1998-1999 through 2000-2001; yearly	Decreases in referrals, suspensions, and expulsions; improved attendance
Mirsky (2007)	SaferSaner Schools (model of RP)	ODR; detentions; incidents of disruptive behavior; OSS	School disciplinary data	1998-1999 through 2001-2002; yearly	Decreases in incidents of misconduct and disciplinary action
Barnhart (2008)	SWPBIS	OSS among SWD	School disciplinary data	2003-2004 through 2007-2008; yearly	Reductions in suspension rates overall and among SWD
Muscott (2008)	SWPBIS	Disciplinary incidents (ODR, ISS, OSS); time savings; academic achievement	School disciplinary data; time estimate surveys; test scores	2003-2004, 2004-2005	Reductions in ODR and suspensions; resulting recovery of instructional time
Cornell (2009)	Virginia Student Threat Assessment Guidelines	Suspensions; aggressive discipline violations; student perceptions of school climate (victimization, bullying, help seeking, learning environment)	School disciplinary data; School Climate Survey	2006-2007	Fewer long-term suspensions; lower rates of victimization and bullying; more favorable ratings of learning environment and help seeking

Bradshaw (2010)	SWPBIS	ODR; suspensions; academic achievement	School disciplinary data; test scores	Baseline, Years 1-4; yearly; dates NR	Reductions in ODR and suspensions
Morrissey (2010)	Schoolwide PBS	ODR	School disciplinary data	2002-2003 through 2004-2005; yearly	Decline in total ODR and receipt of multiple ODR
Cornell (2011)	Virginia Student Threat Assessment Guidelines	Suspensions; assaults, threats, and bullying of other students	School disciplinary data	2006-2007 (baseline) 2008-2009 (follow-up)	Reductions in long-term suspensions and bullying
Bradshaw (2012)	SWPBIS	ODR; OSS; aggressive and disruptive behaviors; concentration problems; prosocial behaviors; emotion regulation	Teacher Observation of Classroom Adaptation—Checklist (TOCA-C)	2002-2003 (fall and spring) 2003-2004 through 2006-2007; yearly (spring)	Lower odds of ODR; lower levels of aggressive/disruptive behaviors and concentration problems; better emotion regulation and higher levels of prosocial behaviors
Chin (2012)	Alternatives to Suspensions program, within PBIS school	Suspensions	School disciplinary data	2005-2006 through 2009-2010; yearly (pre-intervention) 2010-2011 (intervention)	Reductions in total suspensions and re-offenses

Cornell (2012)	Virginia Student Threat Assessment Guidelines	Provision of mental health services; parental involvement; long-term suspensions; alternative school placements	Documentation forms completed by principals	NR	Higher odds of parent involvement and receiving counseling; lower odds of long-term suspension and alternative placement
Simonsen (2012)	SWPBS	ODR; suspensions; academic achievement	School disciplinary data; test scores	2000-2001 through 2007-2008; yearly	Fewer ODR and suspensions; improved scores in reading and math
Mitchell (2013)	SWPBIS	Student perceptions of school climate (fairness, order and discipline, student-teacher relationships, achievement motivation)	School Climate Survey	May (year NR)	Higher ratings of fairness, order and discipline, and student-teacher relationships
Teske (2013)	School Referral Reduction Protocol; Collaborative Child Study Team	Referrals to juvenile court; OSS; graduation rates	NR	NR	Reduction in referrals and OSS; increase in graduation rates
Ward (2013)	Safe and Civil Schools (model of SWPBIS)	Suspensions; student behavior (bullying, violence, defiance and disrespect, classroom disorder); school discipline policy; academic achievement	PBS Assessment staff survey; California Healthy Kids survey; test scores	2007-2008 through 2009-2010; yearly	Decline in suspension rates; improvements in school discipline policy lower odds of classroom disorder and bullying

Anyon (2014)	Restorative approaches; behavior contracts; ISS	ODR; suspensions; referrals to law enforcement; expulsions	School disciplinary data	2011-2012	Lower odds of OSS for RP/ISS; higher odds of OSS/law enforcement referrals for behavior contracts
Gregory (2014)	My Teaching Partner−Secondary	ODR	School disciplinary data	2010-2011	Lower odds of ODR; improvements in disproportionality
Nocera (2014)	SWPBS	ODR; suspensions; behavioral infractions; student perception of school climate; faculty perspectives; academic achievement	School disciplinary data; School Climate Survey; qualitative interviews; state mastery test scores	2007-2008 (baseline) 2008-2009, 2009-2010 (intervention)	Reductions in discipline, including among SWD; improvements in school climate; higher test scores
Osher (2014)	Promoting Alternative Thinking Strategies (PATHS); student support teams; planning centers	Suspendable disciplinary incidents; student perceptions of safety, connectedness and support, challenge, and peer social emotional competence; teacher perceptions of student (K-5) social competence	School disciplinary data; Conditions for Learning survey; teacher surveys	2008-2009 through 2011-2012; yearly	Fewer suspendable offenses and overall decrease in OSS; improved conditions for learning for MS; greater social competence for ES
Nekvasil (2015)	Virginia Student Threat Assessment Guidelines	Suspensions; student perceptions of school climate; teacher perceptions of school safety; bullying/victimization	School disciplinary data; School Climate Survey	Spring 2013	Lower rates of short-term suspensions and bullying/victimization; more favorable ratings of school safety and climate

Anyon (2016)	RI	ODR; OSS	School disciplinary data	2012-2013, fall (intervention) 2012-2013, spring (follow-up)	Lower odds of ODR and OSS
Childs (2016)	SWPBIS	ODR; OSS; ISS	Outcome Data Summary forms	2010-2011 through 2013-2014; yearly	Fewer ODR, ISS, and OSS at outset and conclusion
Gregory (2016)	RP (SaferSaner Schools)	Teacher use of ODR; student perceptions of teacher-student relationships	School disciplinary data; student surveys	2011-2012	Fewer ODR issued by teachers; more positive teacher-student relationships
Okonofua (2016)	Empathic mindset	Suspensions; student perceptions of teacher-student relationships	School disciplinary data; student surveys	Mid-fall, 2 months later; dates NR	Lower odds of suspension; greater perceptions of teacher respect among previously suspended students
Ortega (2016)	RC	Staff and student perceptions, experiences, outcomes	Individual semistructured interviews	Spring 2013	Shift to less exclusion in discipline; improved student-teacher and peer relationships; gains in students' conflict management skills and maturity
Smolkowski (2016)	Safe and Civil Schools (model of SWPBIS)	Problem behavior (bullying, disrespectfulness); suspensions; chronic absenteeism and tardiness	Administrative records; PBIS Assessment staff surveys	2005-2006 through 2011-2012; yearly	Reductions in problem behavior and suspensions

Anyon (2018)	Proactive relationship-building strategies	Educator strategies essential to non-exclusionary discipline success	Semistructured interviews; focus groups	NR	Emphasis on relationship building with students and families; use of home visits, staff visibility and communication with parents, morning meetings and advisory periods with students
Cook (2018)	Greet−Stop−Prompt within PBIS school	ODR; student perceptions of school connection/belonging; intervention acceptability and feasibility	School disciplinary data; student surveys; teacher surveys	Pre- and post-intervention; weekly; dates NR	Reductions in ODR, particularly among Black males
Cornell (2018)	Threat assessment	OSS; expulsions; school transfers; law enforcement action (arrest, court charges, and/or incarceration)	School disciplinary data; school safety audit surveys	2014-2015	Racial parity across outcomes; disparity for SpEd students in suspensions
Cruz (2018)	SWPBIS	OSS	School disciplinary data	2010-2011 through 2015-2016; yearly	Higher risk of OSS; overall reduction in suspensions for Latino students
Gage (2018)	SWPBIS	Disciplinary incidents; ISS; OSS	School disciplinary data	2015-2016	Fewer disciplinary incidents and suspensions
Gregory (2018)	RI	OSS	School disciplinary data	2014-2015	Lower odds of OSS

Hashim (2018)	RJ, suspension ban for willful defiance in SWPBIS schools	Suspensions	School disciplinary data	2003-2004 through 2005-2006; yearly (baseline) 2006-2011; yearly (SWPBIS) 2011-2014; yearly (suspension ban) 2014-2015 (RJ)	Reduction in suspensions, particularly among marginalized students; further reduction with addition of RJ
Kim (2018)	SWPBIS	ODR; OSS; academic achievement	School disciplinary data; test scores	2012-2013 through 2014-2015; yearly	Decrease in ODR and OSS; higher academic achievement with longer implementation
Mansfield (2018)	SaferSaner Schools (model of RP)	ISS; OSS	School disciplinary data; individual interviews with school and district administrators	2010-2011 through 2014-2015; yearly	Decline in suspension rates (including among SWD) and recidivism
Freeman (2019)	PBIS	ODR; suspensions (combined ISS, OSS); absences; academic achievement	School reported discipline and GPA data via Qualtrics online platform	2015-2016	Reductions in ODR and suspensions; fewer absences and unexcused tardies
Gage (2019)	SWPBIS	Corporal punishment; ISS; OSS; expulsions; referrals to law enforcement; school-related arrests	School disciplinary data	2013-2014	Fewer OSS (including among Black students and SWD)

Rainbolt (2019)	SaferSaner Schools (model of RP)	Teacher implementation experiences and perceptions of efficacy	Teacher survey (rating scales, multiple choice, open-ended questions)	2015-2016	High levels of overall use of RP; perceptions of RP reducing misbehavior and improving relationships
Sandwick (2019)	RJ	Staff, student, and family perceptions; key strategies for implementation and addressing challenges	Interviews; focus groups	2016-2017	Overall preference for RP over punitive discipline; emphasis on community ties, infrastructure, and student involvement
von der Embse (2019)	Trauma-informed PBIS	Social, academic, and emotional behavior; PTSD symptoms; ODR	SAEBRS risk screener; Direct Behavior Ratings; CBITS Trauma Exposure Checklist; school disciplinary data	2016-2017, 2017-2018; fall, winter, spring	Reduction in proportion of at-risk students; decreases in ODR
Baule (2020)	SWPBIS	Suspensions; critical discipline incidents; incidents requiring reports to CPS; referrals to juvenile probation; custodial arrests	School disciplinary data	2013-2014 through 2016-2017; yearly	Reduction in suspensions, critical incidents, and arrests
Bruhn (2020)	RJ	Leadership qualities	Interviews; observation	2016-2017	Restraint, persistence, and respect

Garnett (2020)	RP community building circles	Staff and student experiences and perceptions	Semi-structured interviews; online staff surveys; student surveys (modified RP Student Use Scale)	Dec. 2017, Jun. 2018 (staff surveys) Spring 2018 (staff interviews, student surveys)	Increased student SEL skills, meaningful student-teacher relationships, clash with other behavior interventions
Gion (2020)	Classroom Check-up	Teacher use of praise and reprimands; social validity	Observations; coaching logs; Primary Intervention Rating Scale	Baseline and intervention; varied by participant; year NR	Increased praise and decreased reprimands overall and particularly for Black students
Huang (2020)	RP; ISS	Student perceptions of school climate (disciplinary structure, student support, school disengagement, school bonding, and overall safety)	Student Satisfaction Survey	Spring 2016	Poorer perceptions and attitudes toward school for ISS/OSS; no sig. difference for RP
Kervick (2020)	RP	Staff members' personal definitions of RP, training received, and implementation experiences; parents' understanding of RP and personal experiences with RP by themselves or their children	Interviews; staff surveys	Dec. 2017, Jun. 2018 (staff surveys) Spring 2018 (interviews)	Emphasis on collaboration and leader support as facilitators; need for consistency and alignment with existing discipline

Lustick (2020a)	RP	Processes for culturally responsive RP to transform punitive culture	Semi-structured and informal interviews; observations	2013-2014	Belonging and accountability, critical consciousness, inconsistent educator use of culturally relevant factors
Lustick (2020b)	RP	Principals' disciplinary decision making	Interviews	2013-2014	Conflict between ideology and accountability; focus on discipline rather than culture
Lustick (2020c)	RP	Student and staff perceptions of successes and challenges of community-building circles, with attention to equity and inclusion	Naturalistic observation; analysis of school policy documents relevant to discipline; semi-structured interviews with students, teachers, and principals	NR	Impact dependent on teacher-student relationships and familiarity/training, relationship-building, less effective for marginalized students
Maeng (2020)	Comprehensive Student Threat Assessment	OSS; expulsion; placement change; legal action (arrest, incarceration, court charges)	School Safety Audit Survey	2014-2015	Lower odds of suspension, expulsion, and legal action

Nese (2020)	Inclusive Skill-Building Learning Approach	ODR; ISS; OSS; expulsions; instructional time; social validity; staff perceptions of effectiveness and feasibility	School disciplinary data; adapted Primary Intervention Rating Scale: Teacher Version; staff focus group; Google doc record of time spent out of class	2014-2015 (pre-int.) 2015-2016 (post-int.)	Decreases in ODR, ISS, OSS, and expulsion; less instructional time lost to disciplinary incidents; favorable impressions of feasibility
Bohnenkamp (2021)	Emotional and Behavioral Health–Crisis Response and Prevention	Suspensions; ODR; bullying; juvenile justice diversion program referrals; threat assessments and follow-up procedures	School records	2014-2015 (baseline) 2015-2016, 2016-2017 (intervention)	Greater decline in ODR; higher rates of threat assessment and follow-up; slower rate of growth in bullying, suspensions, and juvenile justice referrals
Lee (2021)	SWPBIS	ISS; OSS; expulsions; school transfers; referrals to law enforcement; school-related arrests	School records	2015-2016	Fewer incidents across all outcomes
Zhang (2021)	SWPBIS	Suspensions	School disciplinary data	2018-2019	Lower suspension rates with higher fidelity and greater planning time

Note. ODR = office disciplinary referrals, OSS = out-of-school suspensions, ISS = in-school suspensions, SWD = students with disabilities.

Table 12. Shared Study Characteristics

Category	# of Studies	% of 59 Studies
Primary Tier	24	40.7
Secondary Tier	8	13.6
Tertiary Tier	9	15.3
Multiple Tiers	18	30.5
Elementary Schools	13	22.0
Middle Schools	4	6.8
High Schools	12	20.3
All Grade Levels	18	30.5
Elementary/Middle Schools	3	5.1
Middle/High School	9	15.3
RCT	8	13.6
Quasi-experimental	21	35.6
Correlational	14	23.7
Qualitative	10	16.9
Mixed Quasi-Experimental and Qualitative	6	10.2
Multiple Schools	41	69.5
10 Schools or More	29	49.2
PBIS	22	37.3
Interventions added to PBIS schools	5	8.5
Restorative Practices	17	28.8
Threat Assessment Model	6	10.2
Fidelity Measure Used	21	35.6
Urban Communities	25	42.4
Public Schools	41	69.5
Over 50% Minority Students	34	57.6
Over 50% Free Reduced Price Lunches	29	49.2

CHAPTER FOUR

DISCUSSION

This mixed methods systematic review aimed to collate available data regarding the effectiveness of MTSS supports as alternative non-exclusionary discipline practices that aim to replace or reduce exclusionary outcomes in schools. Previous research and literature reviews suggest there are a host of practices that have promising potential to reduce exclusionary outcomes and disproportionate exclusion for vulnerable student populations, such as Positive Behavioral Intervention Supports (PBIS) and Restorative Practices (RP). However, this class of interventions is rapidly expanding in both implementation and evaluation. By including qualitative research in this review, this work identifies both the outcomes as well the processes and perspectives associated with each MTSS strategy, such as subjective impact, acceptability, feasibility, and implementation barriers and facilitators across multiple types of stakeholders (i.e., staff, students, parents). Studies included in this systematic review were limited to peer-reviewed research in United States general education K-12 settings that was published in 1997 and later in order to capture the most relevant, high-quality work on this topic.

The main research questions for this systematic review worked in tandem to form the recommendations discussed below. First, the researchers asked *what* are the various universal, secondary, and tertiary non-exclusionary MTSS disciplinary interventions helping to reduce the school-to-prison pipeline and treat students equitably. The outcomes of these interventions were

organized by MTSS tier in order to provide a clear sense of when and for whom these interventions should be delivered (i.e., universally, small groups, or individualized interventions). They were then further broken down by school setting, developmental level, and student/staff population to further understand who benefits and who does not. The researchers also asked *how* these interventions are perceived by the implementers and recipients of these alternative strategies to better understand the impact and process by which they work. This systematic review incorporates school-level, staff-level, and student-level effects across multiple types of study design, therefore highlighting prevalent themes across various practices so that these strategies can be more easily replicated in under-resourced settings with access to less support. In order to ensure more unbiased assessment of the evidence, the quality of each study's reporting, methodology, and evidence are discussed to frame how strongly these findings should be taken into consideration when drawing conclusions.

Overview of the Literature

Of the 59 studies identified in this systematic process, the largest number of studies fell into the primary tier of MTSS supports, with the next highest number evaluating all MTSS tiers together in each study. The majority of interventions were implemented either in all grade levels simultaneously or in elementary schools alone, but with middle and high schools close behind with forty-two percent of the data. This suggests a growing body of intervention research on exclusion in secondary schools compared to prior literature reviews (Gage et al., 2018a; Mielke & Farrington, 2021), which is important given that secondary students historically have the highest rates of exclusion and disproportionality (Losen et al., 2016). While RCT designs were unsurprisingly the least prevalent type of quantitative data given the challenges of implementing

this design in schools (Styles & Torgerson, 2018), quasi-experimental studies represented just less than half of the research, suggesting significantly increased efforts by researchers over time to compare intervention effects using more rigorous experimental procedures when random assignment was not possible. Just under a third of studies included qualitative methods, which were not addressed in previous reviews of this topic. The fact that just under half of the studies comprised of large-scale multi-school effectiveness evaluations means that alternative disciplinary interventions are being implemented on a grander scale with more resources and attention than ever before. There also appear to be rising efforts to address the high rates of exclusion and police officers in Black and low-income schools (Welsh & Little, 2018), as exhibited by over half of the studies assessing schools with majority minority student populations and half of the research samples with majority free and reduced priced lunches (representing mid-high poverty levels).

Also as expected, the highest number of studies assessed the effects or process of PBIS, with Restorative Practices (RP) falling closely behind with only five fewer studies. This contrasts sharply with the presence of only one RP study in the Mielke and Farrington (2021) review and none in Valdebenito et al. (2018). Of note, five different interventions took place in schools already implementing PBIS, suggesting that supplemental interventions can operate simultaneously within a structured framework like PBIS. A fidelity assessment was performed in only about one third of the published research, with the majority of those being PBIS studies, likely because of its longer history and more structured approach than other interventions. PBIS had the longest dosage time periods (most often two to four years) which is consistent with studies stating it can take up to three years to develop PBIS leadership and up to five years to

achieve treatment integrity (Solomon et al., 2012). Whereas RP practices are also said to require up to five years to permeate the school culture (Wadhwa, 2016), it is notable that most RP interventions were evaluated after only one year, suggesting the potential for either premature evaluation or rapid induction and acceptance of these practices into the school setting. Brief interventions were the least common, with teacher mindset trainings (Okonofua et al., 2016; Cook et al., 2018) and family violence prevention sessions (Breunlin et al., 2006) being the shortest in duration.

Primary Tier Findings

The majority of quantitative effectiveness trials for the universal tier of MTSS implementation evaluated SWPBIS, with most of these being large-scale multi-school trials and four out of these fourteen studies utilizing RCT designs. Within the eight elementary school studies, there were small to medium effect sizes for a range of exclusionary discipline outcomes (i.e., ODR, OSS, behavior problems) and prosocial outcomes (e.g., fairness, relationships, SEL skills) after SWPBIS was implemented with fidelity, with notable locations in Maryland and Georgia. These studies showed SWPBIS can work in majority minority school populations, including some evidence of reduced ISS, expulsions, alternative school placement, and law enforcement involvement for marginalized students. Similarly, secondary schools had lower ODR, ISS, and OSS rates in Chicago and Louisiana, but low quality precluded strong conclusions about secondary schools. Other large-scale studies across grade levels identified reduced problem behavior, tardiness, ODR, ISS, and OSS rates in states like New Hampshire and Florida, with greater perceived safety and more time for learning. These findings suggest that

students can benefit from universal strategies that take into account the classroom context to reinforce positive behaviors.

There were also four small-scale effectiveness studies looking at ways to train teachers in noticing their own implicit biases and treating or reacting to students with empathy and consistency. The Greet-Stop-Prompt (GSP), My Teaching Partner – Secondary (MTP-S), Classroom Check-Up (CCU), and brief empathic mindset interventions were all implemented in majority minority and low-income schools, with evidence of greatly decreased reprimands (cut in half), ODRs (cut by two-thirds), and suspensions that reduced discipline gaps and promoted school connectedness. While only MTP-S had a comparison group, and the brief mindset training had lower study quality, all studies showed how changing teacher behaviors with more low burden interventions interrupted the coercive cycle of inequity and addressed the root of exclusionary disparities across grade levels.

Only four qualitative studies pertained to universal MTSS strategies, with three of those referring to proactive RP circles. While these studies did not assess effectiveness directly, they did discuss implementation perspectives in mostly minority and low-income schools and found that community-building circles emphasized SEL skill-building and provided avenues to build trusting relationships and shared meaning-making. While some students and teachers made noticeable transformations, more marginalized students could be left feeling vulnerable and teachers feeling overwhelmed. When alternative practices were assessed from the ground up, relationship-building efforts were powerful for identifying students' strengths and triggers and helping them grow with tailored consequences and coping resources.

Secondary Tier Findings

Most of the secondary tier MTSS strategies in the literature involved some form of RP. In Denver, Restorative Circles and Conferences seemed to reduce ODR and OSS rates by allowing for expression of harmful experiences and collective problem-solving to repair harm. However, not all studies had significant effects and racial disparities persisted. Qualitatively, these practices promoted increased helpful dialogue, reflective listening, belonging, and mature conflict resolution without using suspension, but were not successful when the circles were not culturally responsive or embraced by participants. Important leadership qualities included restraint, persistence, and respect in order to prioritize and scaffold teachers in the restorative ethos. In contrast, ISS in a rural high school did not succeed in reducing exclusion.

Tertiary Tier Findings

At the highest level of need and more often in high schools, research shows that exclusion is reduced when staff are heavily involved in making systematic decisions regarding non-exclusionary consequences for student violence or threats. The Virginia Student Threat Assessment Guidelines were evaluated in six quantitative studies (including one RCT, five comparison groups) that found small to medium effects for lower suspensions, bullying, teasing, expulsions, alternative school placements, and referrals to law enforcement, along with better perceptions of safety and school climate and increased access to counseling and parent conferences. In this way, the underlying issue prompting the threat could be solved flexibly with resources, resulting in reduced staff biases for most groups except for special education students. Other inclusive strategies included Alternative to Suspension for Violent Behavior, Alternatives to Suspension, and School Referral Reduction Protocol; all of which strove to reduce

resuspensions by teaching social skills, conflict resolution, and emotion regulation strategies instead of an automatic referral or arrest. These programs involved either family training sessions with a therapist, reflection and skill-building during school, or linkage to specialized treatment by staff members and resulted in reduced suspension rates, expulsions, and ISS as well as better relationships with staff such as SROs.

Multiple Tier Findings

Like the findings for SWPBIS, PBIS across tiers was also successful in reducing suspension rates in primarily low-income schools and some majority minority schools (e.g., Florida, California), with racial disparities diminishing in some, but not all eight studies. Hispanic and Black differences in exclusion rates seemed to persist over other differences and the quality of this research is variable. When a structured tool was used to identify mental health risk (i.e., SAEBRS) and train teachers in how to address trauma reactions, teachers were better able to assign secondary tier interventions within PBIS and teach coping skills like relaxation. Similarly, when an Inclusive Skill-Building Learning Approach helped non-judgmentally redirect or remove students briefly and reconnect, ODRs, ISS, OSS, and expulsions dropped, whereas relationships and self-esteem flourished.

There were six quantitative and three qualitative studies examining RP across tiers, with four of these evaluating the SaferSanerSchools model. Despite variable quality, similar findings revealed reduced ODRs, detentions, disruptive behavior, and OSS, as well as palpable shifts in school culture towards respect, equity, and relationship-building plus narrowing discipline gaps across demographics. This was also true when a ban on suspensions for willful defiance was in place in Los Angeles, whereas behavior contracts in Denver actually worsened problems. Out of

11 RP elements, the more universal strategies (e.g., affective statements, fair process, small impromptu circles) that may take less time seemed to stand out as the most efficacious per staff implementers, as only half of those trained in SaferSanerSchools said they used RP often and others practitioners stated RP was time intensive, required patience and mentoring, and clashed with existing policies. In terms of leadership in NYC schools, when leaders did not prioritize relationships and shared responsibility over accountability and order, or staff perceived the practices as too permissive, the culture and hierarchy did not shift entirely.

Lastly, a focus on socioemotional skills appeared useful for teaching about self-control and interpersonal cognitive problem solving in substantially reducing behavior incidents and OSS for low-income, minority students who received the Promoting Alternative Thinking Strategies (PATHS) curriculum. They also benefited on a more intensive level from support teams to identify warning signs, incentivize positive behaviors, and allocate mental health resources. Planning centers provided opportunities for behavior de-escalation and use of replacement behaviors or coping skills for anger to prevent a disciplinary referral or ease re-entry after exclusion. The Emotional and Behavioral Health – Crisis Response and Prevention (EBH-CRP) program also provided students with psychological support at the secondary level and crisis de-escalation and relapse prevention through cognitive behavioral techniques at the tertiary level, but focused more on peers preventing bullying and violence as ambassadors on the universal level. This set of MTSS supports went even further than PATHS to reduce the growth of juvenile justice referrals in addition to OSS, but did not function as well for students of color.

Recommendations for Future Administrators, Educators, and School Practitioners

Overlapping themes were identified across the quantitative and qualitative findings from each of the 59 articles summarized above in order to synthesize key recommendations for future implementers of non-exclusionary discipline practices who would like to minimize the rates of exclusion and improve equity at their school(s). These recommendations are numbered below:

1) Reducing or replacing exclusionary discipline practices requires significant and sustained change to the culture and relationships of a school in order to transform more traditional punitive mindsets throughout all systems affecting that school.

Relationship-building has repeatedly been identified as a foundational aspect of systems-wide change in disciplinary culture. With respect to the systems view of school climate (Rudasill et al., 2018), school discipline approaches must address transactions that occur between and within each part of the school ecology. Relationships that are formed and strengthened between all members of the school community will maximize chances of altering perspectives of discipline across and within these varying ecological influences. For example, Anyon et al. (2018) documented that increased frequency of contact between staff, students, and their families through repeated greetings, morning meetings, home visits, and contact to families helped staff better understand triggers for behavior problems and root causes of difficulties (e.g., trauma, learning disability), resulting in low suspension rates for impoverished youth, including Black students. Relationship-building also allows for increased opportunities to notice a student's strengths and offer them tailored coping tools and consequences that promote growth rather than simply removing them from the learning environment. In this way, forming strong relationships with students can serve as a seemingly basic solution to impact a wide range of MTSS tiers of

intervention, especially if internal school resources are low. Additionally, relationships with parents can reinforce other discipline strategies being implemented in school, such as telling a parent to praise the same positive behaviors they are working on in school.

As a concrete set of relationship-building tools, Restorative Practices (RP) is perceived by implementers and recipients as effective in improving school relationships and promoting a respectful and collaborative school ethos. Community-building circles have been shown to build trust and shared meaning-making through check-ins and socioemotional skill-building (Lustick et al., 2020c). Rainbolt et al. (2019) and Sandwick et al. (2019) observed that RP seemed to prevent misbehavior by resolving and de-escalating conflicts, encouraging learning from mistakes, and fostering empathy and accountability during interactions with others. When relationships become more strained, Restorative Circles and Conferences can further enhance relationships through open dialogue to discuss a problem, reflective listening to show others they are understood, sharing personal accounts that bring each other closer emotionally, and eventually more spontaneous use of circles as a conflict resolution strategy; all of which aim to prevent detentions or suspensions from needing to be used (Ortega et al., 2016; Anyon et al., 2016; Gregory et al., 2018). Although not all of this available research is high in quality, after foundational relationships have been built through universal RP strategies, secondary and tertiary processes have been shown to facilitate novel ways of resolving tension, repairing friendships, and fostering equitable treatment and kinship with others (Stinchcomb et al., 2006). However, relationship-building is the most important component of implementation to keep in mind, such that focusing on RP content instead of relationships will make implementation more challenging (Mirsky, 2007).

Forming strong bonds through RP can also reduce disproportionality in exclusionary discipline. When more RP elements are used by teachers, Gregory et al., (2016) found that relationships were rated more respectful between students and teachers and the racial gap indicating higher exclusion for Latino and Black students decreased. Similar higher rates of exclusion for marginalized groups have been witnessed to drop steeply in other longitudinal research after RP was implemented (Mansfield et al., 2018; Hashim et al., 2018), but it is unclear which specific strategies or relationships may have been driving those reductions. Luckily, RP practices have the potential to be self-sustaining once they are in motion, such that an entire shift in the school culture perpetuates the use of further RP practices, further illuminating the importance of relationship-building (Mansfield et al., 2018). Also of note is the number of studies with significant findings after implementing some form of RP for only one year, suggesting that a restorative transformation of school climate can have positive effects on exclusion over a shorter period of time than was initially thought (Wadhwa, 2016).

2) Behavioral strategies through Positive Behavior Intervention Supports (PBIS) provide structured MTSS techniques to reward positive behaviors and reduce exclusionary outcomes with high levels of fidelity on a large scale.

PBIS interventions have a wide range of quantitative evidence displaying a significant relationship between high fidelity to these practices and fewer behavior problems, in-school and out-of-school suspensions, expulsions, alternative school placements, referrals to law enforcement, and arrests, particularly in large Eastern school districts like Maryland (Bradshaw et al., 2012), Florida (Childs et al., 2016), and Georgia (Gage et al., 2018). While qualitative research was not available for this type of tiered support, there were small to medium effect sizes

suggesting that: teaching positive behaviors in the classroom and rewarding them throughout the day, giving consistent logical consequences for negative behaviors, and tracking disciplinary data to improve this system can be effective on a large scale with fidelity (e.g., Kim et al., 2018) and in under-resourced settings at high risk for exclusion (e.g., Lee et al., 2021). There is also evidence that when these universal strategies are combined with secondary and tertiary strategies such as Check-In/Check-Out or a functional behavior assessment, they can improve academic scores and attendance while reducing ODRs, OSS, and racial and disability disparities for OSS (Baule et al., 2019; Gage et al., 2019). However, these decreases in disproportionality over time are not consistent across all studies and disparities have been shown to persist despite intervention, especially for Hispanic and Black youth (Gage et al., 2019; Nocera et al., 2014; Cruz et al., 2018).

Of all the interventions assessed in this review, PBIS represented the most rigorous statistical evidence from RCTs and the most frequent use of fidelity assessment. There are also data suggesting that fidelity can be achieved within the first year of implementation, signifying that its training methods are effective for translating the principles into practice (Kim et al., 2018). This body of research speaks to the utility of the PBIS approach for disrupting the school to prison pipeline, but these practices should be understood in terms of their limitations as a solely behavioral intervention. In a more rare, combined study with both behavior contracts (a PBIS strategy) and RP in place, the odds of OSS decreased when participating in RP and ISS as an alternative, but increased when using a behavior contract (Anyon et al., 2014). It may be that in certain schools that are low-income and predominantly communities of color, a more relationship-focused, flexible, and responsive intervention may be more effective in addressing

behavioral problems and increasing understanding between predominantly White teachers and minority students than relying solely on reinforcement and contingencies. It is also important to acknowledge the role that SROs can play in influencing the frequency of law enforcement referrals and arrests on campus, given that these outcomes were less likely to decrease as a result of PBIS with SRO presence, and how this presence should be taken into consideration within PBIS protocols (ACLU, 2019).

3) Teacher-focused trainings in empathy, implicit bias, and culturally-responsive approaches should be implemented in schools with high disproportionality in exclusionary discipline in order to promote equity amongst the most vulnerable groups of students.

Another theme across the research in this review was the strong connection between staff training interventions and reductions in disproportionate exclusion rates for under-resourced populations. To this end, these universal supports addressed teachers and staff as an influential system of school climate that can unintentionally or intentionally perpetuate discrimination and disparities among race, gender, and disability status (Rudasill et al., 2018). These programs tried to alter that system by promoting empathy towards students' perspectives and reflection on ways staff may be biased, less consistent, or not rewarding positive behavior. This is in line with the Vulnerable Decision Points model by Smolkowlski et al. (2016b), which discusses how subjectivity based on situational factors like time of day can influence a teacher's decision making. Four interventions to choose from include Greet-Stop-Prompt (GSP; Cook et al., 2018), My Teaching Partner – Secondary (MTP-S; Gregory et al., 2014), a brief empathic mindset intervention (Okonofua et al., 2016), and Classroom Check-Up (CCU; Gion et al., 2020). Whether it was more subtle behaviors such as reprimands, or grander reactions such as

suspensions, these interventions substantially reduced discipline gaps in majority minority, low-income schools. Fortunately, the commitment for these interventions is relatively low and teachers felt GSP and CCU were acceptable and feasible, with duration across the four programs lasting between seventy minutes, 1.5-8 weeks, five months, or one year.

Whereas whole-school interventions like RP and PBIS have promising effects on disproportionality, they may not be enough to reduce disparities without culturally responsive components embedded, such as race-conscious approaches that address the broader contexts for how disparities in discipline arise (e.g., biases in perceptions of behavior, differential access to rigorous instruction, mismatch between adult and student race or ethnicity) (Gregory et al., 2018). For example, in a study of RP community building circles in three disadvantaged high schools, teachers felt more overwhelmed and implemented circles less effectively with marginalized students (Lustick 2020c). It was noted that they did not have training in how to address bias, institutional racism, or cultural responsiveness within the RP curriculum. In comparison, culturally-responsive circles in Lustick (2020a) helped reinforce belonging, accountability, normalization of experiences, and critical consciousness of systems of oppression as they came together to build relationships and resolve conflict. When the circle was not culturally-responsive, it became more of a formality with little perceived impact from the perspective of the student. In this way, the process by which the circles were implemented was more important than the fact that circles were held. It is possible this lack of responsiveness has also contributed to continued objective disparities in some RP (Anyon et al., 2016) and PBIS research (Cruz et al., 2018).

Teacher training studies corroborate this in the data showing the subjective effects of implicit bias and empathy training on Black male students' increased happiness (Cook et al., 2018), motivation to follow the student code of conduct, and repaired relationships after prior suspensions (Okonofua et al., 2016). As such, targeting teacher bias and interactions with students can have positive reciprocal impacts on both the likelihood that teachers will utilize an exclusionary punishment for a particular student, but also the likelihood a student will misbehave. Both the GSP and Inclusive Skill-Building Learning Approach (Nese et al., 2020) were implemented in schools already utilizing PBIS, suggesting that existing behavioral approaches can be enhanced by bringing more explicit attention to how minority students are treated differently from others. This is important when students are in the same classroom over the entire school year because teacher empathy may burn out faster in these prolonged settings (Gregory et al., 2014).

4) For schools in neighborhoods with high exposure to psychosocial stressors (e.g., community violence, poverty), trauma-informed training and mental health interventions across tiers are necessary as a component of cultural responsiveness to ensure equity in discipline practices.

Exposure to trauma and resulting mental health problems are known risk factors for increased behavior problems and exclusionary discipline outcomes in students. When trauma is not addressed appropriately through disciplinary interactions or referrals to mental health support, the intervention then has potential for iatrogenic effects and the student may not feel comfortable sharing their concerns or experiences (e.g., Lustick 2020c). Trauma-informed strategies were developed to address this gap in disciplinary interventions like PBIS (von der

Embse et al., 2019). Through this trauma-informed approach, staff first identified mental health risk using a screening tool (i.e., SAEBRS), then teachers were coached in classroom secondary tier strategies they can use to address trauma and teach coping skills (e.g., relaxation), which then reduces behavioral and emotional risk over time. By incorporating universal and secondary trauma-informed classroom practices into PBIS, von der Embse et al. (2019) showed enhanced student-teacher relationships, clear expectations, and consistent responses to misbehavior.

Additional tertiary interventions can provide mental health support within the context of reducing exclusionary outcomes by recognizing what behaviors and emotions are more likely to occur as a result of trauma (e.g., dysregulation, violence) and treating the root cause. The four-session family training program called Alternative to Suspension for Violent Behavior and the skill-building program called Alternative to Suspension both utilized violence prevention strategies and SEL training to replace and prevent the need for suspensions (Breunlin et al., 2002; Chin et al., 2012). Similarly, the Emotional and Behavioral Health – Crisis Response and Prevention (EBH-CRP) program (Bohnenkamp et al., 2021) and the PATHS model for SEL (Osher et al., 2014) made sure to incorporate crisis response protocols that facilitate de-escalation and use of replacement behaviors or mental health interventions in lieu of exclusionary options or as a means of effective re-entry after exclusion.

Given that schools considered to be large, urban, and with larger populations of Black and low-income students are more likely to use exclusionary tactics (Marchbanks et al., 2018; Curran, 2019; Welsh & Little, 2018) and less likely to have adequate mental health counselors available (ACLU, 2019), it is crucial that these settings incorporate appropriate responses to the most serious behavior incidents so that students aren't automatically expelled or arrested without

a chance for recovery. Threat assessment models like the Virginia Student Threat Assessment Guidelines (VSTAG) include systematic and careful decision trees that require 6-8 hours of training to implement and assign exclusion only if absolutely necessary. This calculated approach assesses the intent and means of someone making a threat and prevents that person from following through by notifying the intended victim and offering suitable mental health interventions if appropriate. These models have been shown to prevent bullying and victimization and reduce OSS, expulsion, and law enforcement referrals with small to medium effects (e.g., Cornell et al., 2012) as well as eliminate gaps in suspension, expulsion, transfer to alternative schools, and law referrals for race and gender (Cornell et al., 2018).

5) Regardless of intervention tier or type, ongoing professional development, intervention trainings, and support for educators are necessary for effective implementation, especially in under resourced schools that are more likely to experience "initiative fatigue." Appropriate resources must be devoted to this process.

The available evidence shows that when MTSS tiered supports are higher in fidelity of implementation, their effects are stronger and more sustainable over time, especially for PBIS (e.g., Kim et al., 2018). This is also evidenced by the continuously decreasing exclusion rates over time for studies without a fidelity measure. The fact that only two out of fifty-nine studies revealed any worsening exclusion as a result of alternative discipline interventions (Anyon et al., 2014 for behavior contracts; Cruz et al., 2018 for SWPBIS) suggests that these MTSS interventions have promising potential under the right conditions for implementation, although publication bias is likely to reduce the chance that nonsignificant or negative effects of interventions were included in this review. Importantly, while these trends are in the right

direction, there is evidence that fidelity is higher in larger schools with more White students

(Gage et al., 2019) and that fidelity can be poorer in schools with greater poverty levels and the

lack of fidelity slows the rate of improvement (Freeman et al., 2019). Lustick (2020c) found that

RP implementation was enhanced by teacher-student relationships, teacher familiarity with RP,

and support provided to teachers for implementation. Given that multi-tiered interventions can

take several years to develop implementation teams and embed practices into the school culture

(Solomon et al., 2012; Wadhwa, 2016), it is essential for administrators and school district

officials to build in continual opportunities for professional development and training to promote

full understanding and commitment to the non-exclusionary practices and philosophies.

One reason why fidelity can be lower in under resourced schools is the time intensive

nature of alternative practices, as opposed to automatically assigning an exclusionary

punishment. Restorative processes can require more brief and informal use of restorative

language, but higher tier strategies require dedicated time to hold a conversation, circle, or

conference and effectively build or repair relationships. These strategies are also often

implemented several times per week (e.g., Garnett et al., 2020), thus in a high conflict school

setting this can become very onerous and overwhelming. For example, Payne and Welch (2018)

found that schools with greater Black, Hispanic, and low-income students were the most likely to

benefit from tiered discipline supports like RP, but the least likely to implement them. There is

also concerning data suggesting that only 14.2% of behavior incidents in Gregory et al. (2018)

received a restorative intervention on its own, such that all other responses either paired RP with

exclusion at the same time or relied on exclusion. These findings coincide with the notion that

when the number of education initiatives increases but the amount of educator resources and

time remains the same, each new initiative will receive less attention than the last (Reeves, 2010). Therefore, if a new discipline initiative is expected to take hold, then substantial resources are needed to combat this "Law of Initiative Fatigue" (Reeves, 2010).

6) Across the many alternative discipline practices available for educators to choose from, common factors and best practices are identified in order to simplify more complicated and time intensive interventions and reduce burden on implementers.

While the interventions identified throughout these 59 articles were classified by a host of different labels, there were several salient overlapping themes across most interventions. These themes represent overarching best practices that may be more easily adopted and enveloped into school policies when more comprehensive multi-tiered approaches are not as feasible. This synthesis mirrors the common elements approach to evidence-based treatment, which focuses on identifying the therapeutic intervention techniques that are common across treatment packages for multiple problems, then matching those practice elements to the child's presentation using algorithms as guidance (Chorpita & Daleiden, 2009). In this view, the actual strategy used is more important than the label of the whole intervention, therefore receiving training in multiple treatment approaches is no longer necessary.

Kervick et al. (2020) highlights that across prevention-oriented tiered models of school-wide support, they have common goals of reducing exclusion, fostering a positive school community, and increasing time for learning. Of the many universal tier interventions in service of those goals, one unifying factor is the emphasis on **reinforcing positive behaviors** rather than punishing negative behaviors. For example, both RP and PBIS practices use the classroom to promote acceptable school behaviors within the school's culture. Another overlapping primary

intervention is the **use of social and emotional instruction** to build students' internal capacity to regulate emotions and solve problems. Additionally, **relationship-building** and its related skills and strategies (e.g., reflective listening) has been shown to be a key method across interventions to reduce exclusionary discipline and a key component in enhancing interventions that do not specifically target relationships. A **strengths-based and culturally responsive approach** appears to augment the chances that an interaction will promote a positive school climate, including being trauma-informed, capitalizing on resilience factors, and addressing implicit bias as cause of discipline disparities. The **fair implementation of discipline** with empathy, accountability, and respect for all students and staff was shown to be critical for positive school climate across studies as well. Lastly, **systematic and collaborative decision-making** was vital within the secondary and tertiary tiers to provide de-escalation techniques and mental health referrals when appropriate and take into account the student's perspective and context when serious behavior problems arise.

Kidde (2017) posits that RP can be used to complement effective behavioral initiatives, so it appears that selecting the most pertinent strategies from these two similar tiered systems could be beneficial to foster both healthy behavior and equity in disciplinary decisions. Swain-Bradway et al. (2015) provides a model for combining PBIS and RP, such that strategies include community-building circles and schoolwide expectations, restorative conferencing and check-in/check-out, and re-entry circles and wraparound supports. PBIS has also been combined with implicit bias training and SEL skill-building with relatively brief supplemental components. Likewise, there is research to suggest that multi-tiered interventions can be successful with only one tier of strategies. For example, Rainbolt et al. (2019) found that more informal RP primary

strategies such as affective statements, small impromptu circles, and restorative questions were rated as most efficacious by implementers. Similarly, Huang et al. (2020) found that secondary tier RP circles may be less effective than primary tier circles, which may relate to the fact that peace circles can be more time consuming and affect fewer students overall than community-building circles. In terms of common factors across continued exclusion, using less severe exclusionary interventions such as ISS do not appear to be effective on their own without skills-training or reflection also being provided.

7) MTSS supports should be tailored to each school and child's developmental level and family structure.

Although no definitive conclusions can be drawn, certain interventions seem to work more consistently for particular grade levels, and this is important to note when choosing alternative disciplinary strategies to fit a school's ecology. For instance, PBIS has the strongest support in elementary schools, Restorative Circles seem to be most implemented in secondary school settings, threat assessment and crisis de-escalation procedures seem most needed and relevant for high school settings with higher risk levels, and My Teaching Partner has specific modules for secondary schools. Implicit bias and trauma-informed trainings should include information specific to how trauma responses and other emotional and behavioral problems manifest in the school's population and within each grade. For students with accessible parents, it is important to create relationships with these parents and utilize them to reinforce strategies at home or incorporate them into higher tier interventions such as family training programs to prevent violence engagement.

Whereas behavioral strategies may be effective across all school levels, early adolescents and adolescents are more cognitively advanced and able to benefit from supplemental nuanced RP or skill-building strategies than elementary school students who might struggle to tolerate extended problem-solving discussions or may have difficulty integrating multiple concepts and expectations at once. Conversely, elementary school students may be more likely to benefit from the structure and consistency of a PBIS intervention first before trying other MTSS supports. More minor differences in implementation can also help tailor strategies to age while integrating different innovations together, such as incorporating more independent or peer-to-peer practice of skills for older students, using a developmentally-appropriate circle format for younger kids, and adapting trainings and reinforcement systems (e.g., schoolwide assemblies for high schoolers, tickets as token reinforcement) to fit how those students learn best. Lastly, continued use of exclusion in tandem with non-exclusionary practices can be more or less detrimental depending on the age of the child being referred or suspended.

8) Stakeholders should be involved as much as possible in the development, implementation, and evaluation of alternative discipline interventions.

The qualitative research discussed in this review highlights the importance of involving both implementers and recipients of the interventions in the development and refinement of non-exclusionary strategies, including students, teachers, staff, and administrators. These individuals were able to notice aspects of implementation that were unique to their own setting, so their opinions would be valuable in tailoring or limiting alternative discipline strategies to fit their school's history, capacity, and population. They can also speak more to the specific processes by which change in exclusionary outcomes occurs, as well as the likelihood that certain practices

will be accepted and embraced and help work through implementation barriers. Although divergent findings were not common in the available literature, students receiving Restorative Circles in Lustick (2020a) did indicate negative experiences with the intervention when culturally-relevant factors were not addressed appropriately, suggesting teachers perceived themselves to be effective when they were not, according to students.

Of note, only 22 out of the 59 studies in this systematic review involved stakeholder input to some degree in the design of the research evaluation. Case studies, while often lower in research quality, had more involvement of stakeholders in the design of the research to make sure the right questions were being asked or tools were being used; emphasizing the value of listening to individual schools during large-scale implementation. For example, given that elementary schools tend to show higher levels of fidelity to MTSS strategies (e.g., Simonsen et al., 2012), secondary schools would benefit from more involvement of stakeholders to better understand their challenges. In terms of implementation, it also seems that involving more and more stakeholders in the practices can enhance the rate at which discipline culture shifts in the school (e.g., Bruhn et al., 2020) and maximizes the chance that students will use strategies independently (e.g., Ortega et al., 2016). For example, giving peers a voice has been identified as important for community-based after school programs to increase retention through enhanced connectedness towards the program (Love et al., 2018).

Implementation Science and Policy Recommendations

Implementation science is defined as "the scientific study of methods to promote the systematic uptake of research findings and other evidence-based practices into routine practice, and, hence, to improve the quality and effectiveness of health services" (Eccles & Mittman,

2006). This type of research focuses on the strategies used to implement evidence-based

practices, such as clinician behavior or systems-wide changes that help promote adoption,

fidelity, or sustainability of innovations (Bauer & Kirchner, 2020). As discussed above,

educators in low socioeconomic settings are more likely to encounter "initiative fatigue" which

fosters feelings of overwhelm and impatience which can lead to abandonment of promising

interventions if not carefully selected and introduced (Reeves, 2010). This is similar to the idea

of equifinality such that multiple pathways or innovations can lead to desired discipline results

(Eckert, 2017). As reviewed in the current research, the proliferation of alternative discipline

strategies resulting from Obama-era guidance can be effective, but these strategies may not be

implemented in the contexts where they are most needed when a school's attention is already

devoted to other innovations (Fixsen et al., 2005). Relevant avenues for implementation policy

are suggested below.

One recommendation that is likely to enhance implementation of key practices is the use

of professional development consultants or onsite intervention coordinators. These external (e.g.,

nonprofit) assistants can help make an intervention more effective or involve stakeholders in

decision-making when administrators and teachers don't have the time to do so. In order to

reduce the likelihood that low resourced schools will resort to exclusionary discipline,

Stinchcomb et al. (2006) proposes using explicit modeling of RP and social and emotional

instruction so that staff are better equipped to repeat those practices. They also suggest providing

ongoing support for specific techniques (e.g., circles) so that staff are more consistent and feel

connected to one another, and this would also be helpful in determining which restorative

practices are absolutely necessary. For example, a restorative justice coordinator can provide

explicit trainings, materials, coaching, and tracking of progress to uphold accountability in the use of these strategies to adhere to district policies. Schools with staff coming from an engrained dominant behavioral mindset may view RP as too permissive, avoidant, conflicting with existing policies, or time consuming in managing behavior problems (Kervick et al., 2020; Garnett et al., 2020; Rainbolt et al., 2019). Therefore, these schools may need enhanced training in the philosophy of RP for reducing disparities (Rainbolt et al., 2019) and a prominent leadership role to provide scaffolding and encouragement to reframe RP as less inconsistent with the use of positive contingencies (Bruhn et al., 2020).

Given how long school-wide systems change can take, schools need concrete examples of what strategies should be used on a daily basis and to maximize the number of adults and students implementing these strategies. As discussed above, discipline strategies that permeate all parts of the school's ecology have the best chance of positively impacting school climate. At the simplest innovation level, this should involve frequent check-ins from all staff members with students in order to build strong connections with each student, better understand how trauma or resilience factors may be influencing their behavior at school, and increase opportunities for reinforcing their good behaviors and offering social and emotional instruction equitably across students. In addition to frequency of contact, implementers can enhance relationships by showing empathy through therapeutic techniques such as active listening and validating feelings, then using "I feel" statements to communicate rather than judgments. Strong leaders that value relationship-building and shared responsibility over maintaining the perception of order are more likely to shift the traditional power dynamics and increase buy-in for relationship-focused strategies (Lustick et al., 2020b; Sandwick et al., 2019). It is also important to increase

collaboration across staff, not just between staff and students, so that teachers are informed of ongoing non-exclusionary processes and can teach those who may be less strong in these practices (Rainbolt et al., 2019; Barnhart et al., 2008). This could involve structured, collaborative relationships between implementers such as a professional learning community or team to analyze data and set decision-making limits and goals, but this may not be the best option if certain staff members do not have the time available for these collaborative discussions or they feel forced to participate.

Standardization is also an incredibly important method for allowing evidence-based interventions to become a reality. Kervick et al. (2020) staff argued for a more systematic approach to better align RP with existing discipline policies and procedures and provide clear expectations to staff and families for how RP elements should be implemented in order to minimize iatrogenic effects. Whereas PBIS has several measures of fidelity that work across different school settings and populations, more ambiguous and subjective practices like RP do not have statistically validated or reliable tools yet available for measuring adherence to these practices beyond general perceptions of how well teachers and students are embodying the philosophy or using the practices. It is recommended that MTSS practices continue to be standardized to allow for smooth replicability of the practices and that educators seek out measures to facilitate organized implementation and track how this process progresses. For a diverse range of supports like RP, this may be easiest to do first for more concrete, structured practices like community-building or restorative circles. Some school districts have created their own curriculum and evaluation measures which can be validated, such as the implementation rubric, mindset survey, and impact survey from the Chicago Public Schools' Restorative

Practices Guide & Toolkit (CPS, 2017) with support from the non-profit organization UMOJA

Student Development Corporation.

Strengths

This mixed methods systematic review has several strengths that address weaknesses or

gaps from past literature reviews on this topic and enhance our current understanding of

discipline practices that replace or reduce exclusionary outcomes in United States schools. First,

this systematic review utilizes evidence-based guidelines from Cochrane, Campbell, and

PRISMA to thoroughly update the literature on MTSS for exclusion that has been amassing at a

fast pace since IDEA in 1997 and subsequent calls for reduction of exclusion for vulnerable

groups (APA, 2008). Specifically, this review includes data on real-world effectiveness that were

not provided in older reviews that focused on randomized controlled trials (Valdebenito et al.,

2018; Mielke & Farrington, 2021). This review also investigates reduction in exclusionary

outcomes as a primary aim of the research, rather than a supplemental aim resulting from an

intervention targeted at reducing non-disciplinary outcomes (e.g., academic interventions that

also reduce suspensions). In this way, the reviewers were able to more directly synthesize the

outcomes of increasingly tolerant policy and procedure changes to reduce exclusion, rather than

interventions aimed to reduce behavior problems only.

Likewise, this systematic review examines themes in qualitative findings to illuminate

how the interventions actually work to reduce exclusion in each setting from the perspectives of

the stakeholders. As a result, the external validity of this review is improved, whereas only the

past systematic review of RP by Katic et al. (2020) included qualitative findings. This helps to

confirm whether quantitative effectiveness findings are supported or contradicted in actual

implementation, explains findings that might not be statistically significant, and provides

tangible examples for future practitioners. The mixed methods design also addresses challenges

with rigorous whole-school quantitative evaluations given multiple potential confounds

contributing to effects, high variability in implementation across settings, and the difficulty in

capturing all that goes on in schools. Future systematic reviews should continue to incorporate

qualitative studies, when possible, in order to better understand what is driving reductions in

exclusion so that these strategies can be replicated, as discussed in the recommendations above.

In terms of relevance for future implementation, this systematic review identified

commonalities across different types of interventions so that these practices can be simplified

and burden on under-resourced teachers and staff can be reduced, rather than focusing on one

particular brand of intervention (e.g., PBIS, RP). It also includes research across all K-12 grade

levels and MTSS tiers to clarify who benefits more from different specific strategies and further

understand when exclusion may still be warranted and how these practices can co-exist or

interfere. Lastly, the incorporation of multiple study designs allows for a wider range of

viewpoints, settings, and routes of analysis for newer promising interventions. Overall, this

systematic review offers direction for future implementers and researchers hoping to advance the

field of alternative discipline practices.

Limitations

While there are many strengths, there are also several limitations that preclude full

understanding of which non-exclusionary strategies are most effective in reducing exclusion and

promoting equity. The results in this review were sometimes grouped by intervention tier, but

sometimes the most applicable tier was less clear based on the study's reported methods and was

therefore assigned more subjectively. Although the literature search was systematic and careful, it is possible that important studies were not captured in the database searches performed. While effect sizes appeared to range from small to medium, the low number of randomized controlled trials and consistent methods prevented a meta-analysis from being conducted, which means researchers could not coherently synthesize results into an overall estimate of effect or provide robust suggestions regarding the relative effectiveness of various practices. Likewise, not all studies included the same outcome measurement or included a comparison or control group, making the quality of overlapping themes reported more variable.

As discussed previously, the omission of unpublished works and research published in other countries limits the conclusions that can be made due to publication and setting bias. For example, there were only four studies that assessed interventions involving SROs as part of a multidisciplinary team to reduce exclusion for the highest risk individuals (Cornell et al., 2011; Cornell et al., 2012; Maeng et al., 2020; Teske et al., 2013). Making a strong assertion about SRO involvement based on these studies is problematic given that SROs do not receive extensive training in recognizing developmentally appropriate student behaviors and implicit biases like teachers and staff do, which can lead to continued discrimination towards minority and special education students and escalation of behaviors (Maeng et al., 2020). Some research has even shown that SROs exacerbate crime and social disturbances and that it may be best to remove them entirely from school systems (Fisher & Devlin, 2019; Devlin & Fisher, 2021). Therefore, it is important that recommendations from this systematic review be understood in the context of the emerging patterns of available research.

While dosage was included as a variable in the compiled tables, it should be noted that dosage varied greatly in how it was measured. This value ranged in clarity from a rough estimate of when an intervention began implementation to a definite marker of when an implementation was adopted. In addition, some studies did not report the length of intervention at all or indicated that the number of strategies increased over time without clear delineation of which strategies started when. Likewise, some longitudinal studies assessed change over time before and during implementation, whereas other research utilized clear baseline assessments and only assessed change during intervention application. For these reasons, assertions about minimum dosage needed to observe changes in exclusion or disparities cannot be formally made. Especially for correlational studies, this also creates difficulty in assessing whether reductions in exclusionary outcomes are due to natural downtrends in exclusion based on national or state guidance and protocols for improving school climate, or whether the reductions are due to the interventions implemented at that school. To this effect, there is concern that some schools may be changing their record-keeping practices or switching out forms of punitive discipline in order to make it appear that their rates of exclusion are decreasing (Harper, 2019).

Lastly, there were several team members who were less experienced in evaluating research critically and may have led to inaccurate quality assessment or missed findings. The quality assessment tool (QuADS; Harrison et al., 2021) was limited in that it did not account for quality of outcome reporting beyond the strengths and limitations assessment, such that studies could omit statistical values and still be high in quality. This tool also had four options for each item, which provided more nuance in quality ratings but also led to more subjective interpretation. As such, the questions pertaining to appropriateness of study design, sampling,

data collection, and analysis was more difficult to determine and not all items had acceptable

reliability across raters.

Future Research Directions

Of note, MTSS supports intervene on several school variables through numerous

mechanisms across many systems of school climate. For example, interventions can reduce

exclusion by both improving student behavior and reducing staff implementation of exclusion as

a response to misbehavior through policy changes and suspension diversion programs. However,

it is difficult to distinguish between reductions in exclusionary practices versus outcomes and

whether the same processes work for reducing disproportionality when these mechanisms are not

isolated in the current research. Future research should aim to parse apart key mechanisms or

mediators that drive reductions in exclusion given that there are multiple systems of school

climate involved in implementation and multiple tiers of intervention at play. This is also

important to understand since disproportionality in exclusion persisted in several studies despite

intervention, suggesting that discipline systems may be more biased than can be addressed at the

staff level. Additionally, it is still not fully clear when exclusion can be most useful (e.g., ISS to

support reflection, suspensions while staff locate appropriate interventions) or when it is paired

with non-exclusionary practices. Studies that compare and contrast multiple types of

interventions at once can help understand which practices can be feasibly combined and which

approaches may have more weight on particular outcomes than others (e.g., a behavioral versus

humanistic approach). Further moderation analysis would also allow for identifying additional

factors that make exclusionary reductions stronger or weaker, such as fidelity (which was

addressed in several studies), teacher and student perceptions of practices and systems of school

climate that may need more attention, external resources and implementation time available, and school size and student demographics.

Regarding specific gaps in the literature, further research is needed to assess RP and teacher implicit bias and trauma-informed trainings using comparison groups to more causally confirm that changes in exclusionary outcomes were a direct result of these interventions. Of note, there was a dearth of research on tertiary tier restorative interventions. Qualitative PBIS studies are also rare and would illuminate how these practices are perceived and tailored to the school, especially in under-resourced schools that may benefit less from a purely behavioral approach. While some types of strategies were implemented with consistent fidelity measures, these would be important to include in future evaluations of more amorphous interventions like RP so that integrity of implementation can be addressed as a moderator. Although secondary school research is increasing in amount, it is still limited in comparison to research completed with primary schools. Secondary schools are a more difficult school setting to evaluate but the damage to school climate from exclusionary discipline is widespread and these schools should continue to be prioritized. As mixed methods systematic reviews become more prevalent and quality assessment measures continue to improve, these updated methods should be utilized to ensure all aspects of research quality are assessed.

Conclusions

In searching for a unifying theory for these recommendations, it appears that disability critical race theory (DisCrit; Annamma et al., 2018) is well-suited to address the intersectionality at play within the continued disparities in exclusionary outcomes for marginalized youth. The DisCrit theory acknowledges how a history of oppression based on both race and ability level

have perpetuated inequities and disadvantage in achievement and discipline. As such, a student with both special education and racial minority status will likely experience compounded risk for exclusionary punishment, such as a Black male student with ADHD automatically being labeled or perceived as deviant, aggressive, or unintelligent (Annamma et al., 2018). These demographic variables can work independently or synergistically to exacerbate experiences of discrimination or microaggressions. In the current systematic review, the continued outcome gaps in exclusion for special education students and Black and Latinx students across several studies speak to the overwhelming oppression that these students face. For these youth, it is critical that educators and school clinicians recognize the accumulation of stressors in their lives that make minority students and students with disabilities more susceptible to discrimination through racism and ableism. Exclusion has the potential to retraumatize or alienate children exposed to toxic stress, whereas the current available research reinforces the importance of supportive relationships, mental health treatment, and social and emotional skills in providing protective effects for disadvantage children and adolescents (Harper, 2019).

Overall, this systematic review provides a comprehensive examination of both qualitative and quantitative research on interventions designed to reduce exclusionary discipline practices in order to understand both the impact that these interventions have had on students and staff as well as the process by which these interventions are implemented and received. While the research is still growing on this topic, several interventions have contributed to sustainable reductions in exclusion and disparities in exclusionary outcomes based on race, gender, and disability status, both in terms of statistical changes and perceived positive impacts of MTSS. Future implementers hoping to reduce exclusion can benefit from a wide range of strategies

including: promoting school-wide culture change through relationship-building; using behavioral strategies to incentivize positive behaviors; implementing brief teacher-focused trainings to enhance empathy, cultural responsiveness, and awareness of implicit bias; offering trauma-informed trainings to enhance contextual awareness; providing ongoing professional development to prevent "initiative fatigue"; selecting core common practices across interventions to simplify implementation and reduce burden; tailoring interventions to fit students' developmental level; and involving stakeholders in development, implementation, and evaluation of these interventions. Primary, secondary, and tertiary supports can be isolated for specific circumstances (e.g., grade level, resources) and combined in various ways to best address a school's ecology and needs. Although many interventions have been evaluated by their label, it is important to address the form and process of addressing student misbehavior (e.g., cultural responsiveness, relationship-building), as this can be just as meaningful for reducing exclusion as the content of interventions that educators implement. While it is still unclear what the *most* effective interventions are for whole-school change, there are a wide range of tiered supports available to curb the school-to-prison pipeline that should continue to be investigated to promote equity in disciplinary policy and practices.

APPENDIX A

GLOSSARY OF KEY TERMS

KEY TERMS

MTSS   Multi-Tiered System of Supports

> A continuum of evidence-based approaches to educational and behavioral strategies in schools which comprises 3 tiers of practices tailored to students' levels of need. Tier 1 (universal) practices are school- or classroom-wide for all students. Tier 2 (targeted) practices are offered to at-risk students, often in small groups. Tier 3 (indicated) practices are individualized to target students with intensive needs.

PBS, PBIS   Positive Behavior (Interventions and) Supports

> Evidence-based framework for preventing problem behavior, providing instruction and support for positive and prosocial behaviors, and supporting social, emotional and behavioral needs for all students. Schoolwide implementation of PBIS requires training, coaching, and evaluation for school staff to consistently implement the key components that make PBIS effective for all students.

RJ   Restorative Justice

> A philosophy which emphasizes achieving justice by repairing harm, accepting accountability, and building relationships rather than delivering punishment.

RP   Restorative Practices

> Educational RJ practices that aim to nurture relationships between administrators, teachers, students, security staff, parents, and the community and provide a fair and equitable learning environment that emphasizes social engagement over social control.

SEL   Social Emotional Learning

> Development of intra- and interpersonal capabilities including but not limited to self-regulation and skills for building/maintaining relationships with others.

SWPBS, SWPBIS   School-Wide Positive Behavior (Interventions and) Supports

> A subset of PBS/PBIS which focuses on universal/Tier 1 practices implemented across the entire school.

ODR   Office Disciplinary Referral

ODRs are defined as representing an event where (a) a student engaged in behavior that violated a rule or social norm in the school, (b) the problem behavior was observed or identified by a member of the school staff, and (c) administrative staff delivered a consequence through a permanent (written) product that defined the whole event.

ISS   In-School Suspension

Form of punishment that keeps students in school and doing work, but isolates them from the rest of the student body.

OSS   Out-of-School Suspension

The removal of a student from the school environment for a period not to exceed ten days.

SRO   School Resource Officer

Career law enforcement officer with sworn authority who is deployed by an employing police department or agency in a community-oriented policing assignment to work in collaboration with one or more schools. NASRO recommends that agencies select officers carefully for SRO assignments and that officers received at least 40 hours of specialized training in school policing before being assigned.

FRP   Free or reduced-price lunch (FRPL)

The percentage of students eligible for free or reduced-price lunch (FRPL) under the National School Lunch Program provides a proxy measure for the concentration of low-income students within a school.

REFERENCE LIST

Advancement Project (2010). *Test, Punish and Push-Out: How "Zero Tolerance" and High Stakes Testing Funnel Youth into the School-to-Prison Pipeline.* http://www.advancementproject.org/issues/stopping-the-school-to-prison-pipeline

Allen, J. P., Pianta, R. C., Gregory, A., Mikami, A. Y., & Lun, J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. *Science, 333*(6045), 1034-1037.

American Civil Liberties Union (2017). Bullies in blue: The origins and consequences of school policing. https://www.aclu.org/sites/default/files/field_document/aclu_bullies_in_blue_4_11_17_final.pdf.

American Civil Liberties Union (2019). Cops and no counselors: How the lack of school mental health staff is harming students. https://www.aclu.org/issues/juvenile-justice/school-prison-pipeline/cops-and-no-counselors

American Psychological Association, Task Force on Resilience and Strength in Black Children and Adolescents. (2008a). *Resilience in African American children and adolescents: A vision for optimal development.* http://www.apa.org/pi/cyf/resilience.html

American Psychological Association Zero Tolerance Task Force. (2008b). Are zero tolerance policies effective in the schools? An evidentiary review and recommendations. *The American Psychologist, 63*(9), 852-862.

Anderson, K., Ritter, G., & Zamarro, G. (2017). *Understanding a vicious cycle: Do out-of-school suspensions impact student test scores?.* Department of Education Reform, University of Arkansas

Anfara, V. A., Jr., Evans, K. R., & Lester, J. N. (2013). Restorative justice in education: What we know so far. *Middle School Journal*, *44*(5), 57-63.

Annamma, S. A., Anyon, Y., Joseph, N. M., Farrar, J., Greer, E., Downing, B., & Simmons, J. (2019). Black girls and school discipline: The complexities of being overrepresented and understudied. *Urban Education, 54*(2), 211-242. https://doi.org/10.1177/0042085916646610

Annamma, S. A., Ferri, B. A., & Connor, D. J. (2018). Disability critical race theory: Exploring the intersectional lineage, emergence, and potential futures of DisCrit in education. *Review of Research in Education*, *42*(1), 46-71.

Anyon, Y., Atteberry-Ash, B., Yang, J., Pauline, M., Wiley, K., Cash, D., Downing, B., Greer, E., & Pisciotta, L. (2018). "It's all about the relationships": Educators' rationales and strategies for building connections with students to prevent exclusionary school discipline outcomes. *Children & Schools, 40*(4), 221-230. https://doi.org/10.1093/cs/cdy017

Anyon, Y., Gregory, A., Stone, S., Farrar, J., Jenson, J. M., McQueen, J., Downing, B., Greer, E., & Simmons, J. (2016). Restorative interventions and school discipline sanctions in a large urban school district. *American Educational Research Journal*, *53*(6), 1663-1697. https://doi.org/10.3102/0002831216675719

Anyon, Y., Jenson, J. M., Altschul, I., Farrar, J., McQueen, J., Greer, E., Downing, B., & Simmons, J. (2014). The persistent effect of race and the promise of alternatives to suspension in school discipline outcomes. *Children and Youth Services Review, 44*, 379-386. https://doi.org/10.1016/j.childyouth.2014.06.025

Balfanz, R., Byrnes, V., & Fox, J. H. (2015). Sent home and put off track. Closing the school discipline gap: Equitable remedies for excessive exclusion, 17-30.

Barnhart, M. K., Franklin, N. J., & Alleman, J. R. (2008). Lessons learned and strategies used in reducing the frequency of out-of-school suspensions. *Journal of Special Education Leadership, 21*(2), 75-83.

Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., Reschly, D. J., Schrag, J. & Tilly, W. D. (2005*). Response to intervention: Policy considerations and implementation.* National Association of State Directors of Special Education.

Bauer, M. S., & Kirchner, J. (2020). Implementation science: what is it and why should I care? *Psychiatry Research*, *283*, 112376.

Baule, S. M. (2020). The impact of Positive Behavior Intervention Support (PBIS) on suspensions by race and ethnicity in an urban school district. *AASA Journal of Scholarship & Practice, 16*(4), 45-56.

Bear, G. G, Whitcomb, S. A., Elias, M. J., & Blank, J. C. (2015). SEL and Schoolwide Positive Behavioral Interventions and Supports. In J. Durlak, T. Gullotta, C. Domitrovich, P. Goren, & R. Weissberg (Eds.), *Handbook of social and emotional learning* (pp. 453-467). Guilford Press.

Beckman, K., McMorris, B., & Gower, A. (2012). Restorative interventions implementation toolkit. *Minneapolis, MN: University of Minnesota, Healthy Youth Development—Prevention Research Center*.

Bohnenkamp, J. H., Schaeffer, C. M., Siegal, R., Beason, T., Smith-Millman, M., & Hoover, S. (2021). Impact of a school-based, multi-tiered emotional and behavioral health crisis intervention on school safety and discipline. *Prevention Science, 22*(4), 492-503. https://doi.org/10.1007/s11121-020-01195-3

Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of Schoolwide Positive Behavioral Interventions and Supports on student outcomes. *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. (2021). A randomized controlled trial of MTSS-B in high schools: Improving classroom management to prevent EBDs. *Remedial Special Education, 42*(1), 44-59. https://doi.org/10.1177/0741932520966727

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of School-Wide Positive Behavioral Interventions and Supports on child behavior problems. *Pediatrics, 130*(5), e1136-e1145. https://doi.org/10.1542/peds.2012-0243

Breunlin, D., Cimmarusti, R. A., Bryant-Edwards, T. L., & Hetherington, J. S. (2002). Conflict resolution training as an alternative to suspension for violent behavior. *Journal of Educational Research, 95*(6), 349-357.

Breunlin, D. C., Cimmarusti, R. A., Hetherington, J. S., & Kinsman, J. (2006). Making the Smart Choice: A systemic response to school-based violence. *Journal of Family Therapy, 28*(3), 246-266.

Brewer, R. M., & Heitzeg, N. A. (2008). The racialization of crime and punishment: Criminal justice, color-blind racism, and the political economy of the prison industrial complex. *American Behavioral Scientist, 51*(5), 625-644.

Brookmeyer, K. A., Fanti, K. A., & Henrich, C. C. (2006). Schools, parents, and youth violence: A multilevel, ecological analysis. *Journal of clinical child and adolescent psychology, 35*(4), 504-514.

Browers, A., & Tomic, W. (2000). A longitudinal study of teacher burnout and perceived self-efficacy in classroom management. *Teaching and Teacher Education, 16*(2), 239-253.

Bruhn, S. (2020). "The child is not broken": Leadership and restorative justice at an urban charter high school. *Teachers College Record, 122*(8).

Buscemi, N., Hartling, L., Vandermeer, B., Tjosvold, L., & Klassen, T. P. (2006). Single data extraction generated more errors than double data extraction in systematic reviews.

Journal of Clinical Epidemiology, 59(7), 697-703.
https://doi.org/10.1016/j.jclinepi.2005.11.010

Carter, P., Fine, M., & Russell, S. (2014). Discipline disparities: A research-to-practice collaborative. *The Equity Project at Indiana University*.

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

Chin, J., Dowdy, E., Jimerson, S., & Rime, W. J. (2012). Alternatives to suspensions: rationale and recommendations. *Journal of School Violence, 11*(2), 156-173. https://doi.org/10.1080/15388220.2012.652912

Cohen, J., McCabe, L., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *Teachers College Record, 111*(1), 180-213.

Coker, T. R., Elliott, M. N., Toomey, S. L., Schwebel, D. C., Cuccaro, P., Emery, S. T., Davies, S. L., Visser, S. N., & Schuster, M. A. (2016). Racial and ethnic disparities in ADHD diagnosis and treatment. *Pediatrics, 138*(3), e20160407.

Cook, C. R., Duong, M. T., McIntosh, K., Fiat, A. E., Larson, M., Pullmann, M. D., & McGinnis, J. (2018). Addressing discipline disparities for black male students: Linking malleable root causes to feasible and effective practices. *School Psychology Review, 47*(2), 135-152. https://doi.org/10.17105/SPR-2017-0026.V47-2

Cornell, D. G., Allen, K., & Fan, X. (2012). A randomized controlled study of the Virginia Student Threat Assessment Guidelines in kindergarten through grade 12. *School Psychology Review, 41*(1), 100-115. https://doi.org/10.1080/02796015.2012.12087378

Cornell, D. G., Gregory, A., & Fan, X. (2011). Reductions in long-term suspensions following adoption of the Virginia Student Threat Assessment Guidelines. *NASSP Bulletin, 95*(3), 175-194. https://doi.org/10.1177/0192636511415255

Cornell, D., Maeng, J., Huang, F., Shukla, K., Konold, T., & Bottiani, J. (2018). Racial/ethnic parity in disciplinary consequences using student threat assessment. *School Psychology Review, 47*(2), 183-195. https://doi.org/10.17105/SPR-2017-0030.V47-2

Cornell, D., Sheras, P., Gregory, A., & Fan, X. (2009). A retrospective study of school safety conditions in high schools using the Virginia Threat Assessment Guidelines versus alternative approaches. *School Psychology Quarterly, 24*(2), 119-129. https://doi.org/10.1037/a0016182

Crone, D. A., Hawken, L. S., & Horner, R. H. (2010). *Responding to problem behavior in schools: The Behavior Education Program* (2nd Ed.) Guilford Press.

Cruz, R. A., & Rodl, J. E. (2018). Crime and punishment: An examination of school context and student characteristics that predict out-of-school suspension. *Children & Youth Services Review*, 95, 226-234. https://doi.org/10.1016/j.childyouth.2018.11.007

Cuellar, A. E., & Markowitz, S. (2015). School suspension and the school-to-prison pipeline. *International Review of Law and Economics*, *43*, 98-106.

Curran, F. C. (2016). Estimating the effect of state zero tolerance laws on exclusionary discipline, racial discipline gaps, and student behavior. *Educational Evaluation and Policy Analysis*, *38*(4), 647-668.

Curran, F. C. (2019). The law, policy, and portrayal of zero tolerance school discipline: Examining prevalence and characteristics across levels of governance and school districts. *Educational Policy*, *33*(2), 319-349.

Darling-Hammond, S. Fronius, T. A., Sutherland, H., Guckenburg, S., Petrosino, A., & Hurley, N. Effectiveness of Restorative Justice in US K-12 schools: A review of quantitative research. *Contemporary School Psychology, 24*, 295-308. https://doi.org/10.1007/s40688-020-00290-0

Davis, F. E. (2019). *The little book of race and restorative justice: Black lives, healing, and US social transformation.* Simon and Schuster.

Deakin, J., & Kupchik, A. (2018). Managing Behaviour: From Exclusion to Restorative Practices. In *The Palgrave International Handbook of School Discipline, Surveillance, and Social Control* (pp. 511-527). Palgrave Macmillan, Cham.

Devlin, D. N., & Fisher, B. W. (2021). An examination of cShool Resource Officers as an approach to reduce social disturbances in schools: Evidence from a national longitudinal study. *Journal of School Violence*, *20*(2), 228-240.

Dixon-Woods, M. (2011). Using framework-based synthesis for conducting reviews of qualitative studies. *BMC Medicine, 9*(1), 1-2. https://doi.org/10.1186/1741-7015-9-39

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*(1), 405-432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Eccles, M. P., & Mittman, B. S. (2006). Welcome to implementation science. *Implementation Science, 1.*

Eckert, J. (2017, September 13). The "No" in "Innovate." *Edutopia.*

Evans, K., & Vaandering, D. (2016). *The little book of restorative justice in education: Fostering responsibility, healing, and hope in schools*. Simon and Schuster.

Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). https://www.congress.gov/bill/114th-congress/senate-bill/1177

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

Fenning, P., & Jenkins, K. (2018). Racial and Ethnic Disparities in Exclusionary School Discipline: Implications for Administrators Leading Discipline Reform Efforts. *NASSP Bulletin, 102*(4), 291-302.

Fenning, P. A., & Johnson, M. B. (2016). Developing prevention-oriented discipline codes of conduct. *Child. Legal Rts. J.*, *36*, 107.

Fisher, B. W., & Devlin, D. N. (2020). School crime and the patterns of roles of school resource officers: Evidence from a national longitudinal study. *Crime & Delinquency*, *66*(11), 1606-1629.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network. (FMHI Publication No. 231)

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.

Fronius, T., Darling-Hammond, S., Persson, H., Guckenburg, S., Hurley, N., & Petrosino, A. (2019). *Restorative justice in U.S. schools: An updated research review*. WestEd.

Fuentes, A. (2014). The schoolhouse as jailhouse. In A. J. Nocella II, P. Parmar, & D. Stovall (Eds.), *From education to incarceration: Dismantling the school-to-prison pipeline*. New York, NY: Peter Lang.

Gage, N. A., Grasley-Boy, N., Peshak George, H., Childs, K., & Kincaid, D. (2019). A quasi-experimental design analysis of the effects of School-Wide Positive Behavior

Interventions and Supports on discipline in Florida. *Journal of Positive Behavior Interventions, 21*(1), 50-61. https://doi.org/10.1177/1098300718768208

Gage, N. A., Whitford, D. K., & Katsiyannis, A. (2018a). 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%2F0022466918767847

Gage, N. A., Lee, A., Grasley-Boy, N., & Peshak George, H. (2018b). The impact of School-Wide Positive Behavior Interventions and Supports on school suspensions: A statewide quasi-experimental analysis. *Journal of Positive Behavior Interventions, 20*(4), 217-226. https://doi.org/10.1177/1098300718768204

Garnett, B., Kervick, C., Moore, M., Ballysingh, T., & Smith, L. (2020). School staff and youth perspectives of tier 1 restorative practices classroom circles. *School Psychology Review*. https://doi.org/10.1080/2372966X.2020.1795557

Gastic, B. (2017). Disproportionality in school discipline in Massachusetts. *Education and Urban Society*, *49*(2), 163-179.

Gion, C., McIntosh, K., & Falcon, S. (2020). Effects of a multifaceted classroom intervention on racial disproportionality. *School Psychology Review*. https://doi.org/10.1080/2372966X.2020.1788906

Golann, J. W. (2015). The paradox of success at a no-excuses school. *Sociology of Education, 88*(2), 103-119.

González, T. (2012). Keeping kids in schools: Restorative justice, punitive discipline, and the school to prison pipeline. *JL & Educ.*, *41*, 281.

Government Accountability Office. (2018, March). *Discipline disparities for Black Students, Boys, and Students with Disabilities.* Report to Congressional Requesters. https://www.gao.gov/assets/700/690828.pdf.

Gregory, A., Allen, J. P. Mikami, A. Y., Hafen, C. A., & Pianta, R. (2014). Eliminating the racial disparity in classroom exclusionary discipline. *Journal of Applied Research on Children, 5*(2), 12.

Gregory, A., Clawson, K., Davis, A., & Gerewitz, J. (2016). The promise of restorative practices to transform teacher-student relationships and achieve equity in school discipline. *Journal of Educational and Psychological Consultation, 26*(4), 325-353.

Gregory, A., Huang, F. L., Anyon, Y., Greer, E., & Downing, B. (2018). An examination of restorative interventions and racial equity in out-of-school suspensions. *School Psychology Review, 47*(2), 167-182.

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.

Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The achievement gap and the discipline gap: Two sides of the same coin?. *Educational Researcher*, *39*(1), 59-68.

Gregory, A., & Weinstein, R. S. (2008). The discipline gap and African Americans: Defiance or cooperation in the high school classroom. *Journal of School Psychology, 46*(4), 455-475.

Gresham, F. M., & Elliott, S. N. (2008). *Social Skills Improvement System: Rating Scales*. Pearson Assessments.

Gresham, F. M., Hunter, K. K., Corwin, E. P., & Fischer, A. J. (2013). Screening, assessment, treatment, and outcome evaluation of behavioral difficulties in an RTI model. *Exceptionality, 21*(1), 19-33. https://doi.org/10.1080/09362835.2013.750115

Haight, W., Gibson, P. A., Kayama, M., Marshall, J. M., & Wilson, R. (2014). An ecological-systems inquiry into racial disproportionalities in out-of-school suspensions from youth, caregiver and educator perspectives. *Children and Youth Services Review, 46*, 128-138.

Hamlin, J. B., & Darling, J. (2012). Use of peace circles in large-scale community conflict: A case study. *Conflict Resolution Quarterly, 29*(4), 403-419.

Harper, K. (2019, November 20). The need for a comprehensive approach to address school discipline & discipline disparities by race & disability in Pennsylvania. *Child Trends.* https://www.childtrends.org/publications/the-need-for-comprehensive-approach-address-school-discipline-discipline-disparities-by-race-disability-in-pennsylvania

Harrison, R., Jones, B., Gardener, P., & Lawton, R. (2021). Quality assessment with diverse studies (QuADS): An appraisal tool for methodological and reporting quality in systematic reviews of mixed-or multi-method studies. *BMC Health Services Research, 21*(1), 1-20. https://doi.org/10.1186/s12913-021-06122-y

Hartling, L., Featherstone, R., Nuspl, M., Shave, K., Dryden, D. M., & Vandermeer, B. (2017). Grey literature in systematic reviews: a cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of meta-analyses in child-relevant reviews. *BMC Medical Research Methodology, 17*(64). https://doi.org/10.1186/s12874-017-0347-z

Hashim, A. K., Strunk, K. O., & Dhaliwal, T. K. (2018). Justice for all? Suspension bans and restorative justice programs in the Los Angeles Unified School District. *Peabody Journal of Education, 93*(2), 174-189. https://doi.org/10.1080/0161956X.2018.1435040

Heitzeg, N. A. (2014). Criminalizing education: Zero tolerance policies, police in the hallways, and the school to prison pipeline. *Counterpoints*, *453*, 11-36.

Higgins, J.P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., & Welch, V.A. (Eds). (2021). *Cochrane Handbook for Systematic Reviews of Interventions version 6.2*. John Wiley & Sons.

Hockenberry, S., & Puzzanchera, C. (2014). *Juvenile court statistics, 2011*. Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice.

Hong, Q. N., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M. P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M. C., Vedel, I., & Pluye, P. (2018). The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *Education for Information, 34*(4), 285-291. https://doi.org/10.3233/EFI-180221

Huang, F. & Anyon, Y. (2020). The relationship between school disciplinary resolutions with school climate and attitudes toward school. *Preventing School Failure, 64*(3), 212-222.

Huang, F. L., & Cornell, D. G. (2017). Student attitudes and behaviors as explanations for the Black-White suspension gap. *Children and Youth Services Review*, *73*, 298-308.

Individuals with Disability Education Act Amendments of 1997 [IDEA]. (1997). https://www.congress.gov/105/plaws/publ17/PLAW-105publ17.pdf

Jones, M. C. (2013). *White and Hispanic Texas middle school students' discipline consequence type and academic achievement: A statewide analysis*. Sam Houston State University.

Justice Policy Institute. (2011). *Education under arrest: The case against police in schools*. Justice Policy Institute.

Kang-Brown, J., Trone, J., Fratello, J., & Daftary-Kapur, T.(2013). *A generation later: What we've learned about zero tolerance in schools*. New York, NY: Vera Institute of Justice, Center on Youth Justice.

Katic, B,, Alba, L. A., & Johnson, A. H. (2020). A systematic evaluation of restorative justice practices: School violence prevention and response. *Journal of School Violence, 19*(4), 579-593. https://doi.org/10.1080/15388220.2020.1783670

Kellermeyer, L., Harnke, B., & Knight, S. (2018). Covidence and Rayyan. *Journal of the Medical Library Association, 106*(4), 580-583. http://dx.doi.org/10.5195/jmla.2018.513

Kelling, G. L., & Coles, C. M. (1997). *Fixing broken windows: Restoring order and reducing crime in our communities.* Simon and Schuster.

Kelly, J., Sadeghieh, T., & Adeli, K. (2014). Peer review in scientific publications: Benefits, critiques, & a survival guide. *eJIFCC, 25*(3), 227-243.

Kervick, C. T., Garnett, B., Moore, M., Ballysingh, T. A., & Smith, L. C. (2020). Introducing restorative practices in a diverse elementary school to build community and reduce exclusionary discipline: Year one processes, facilitators, and next steps. *School Community Journal, 30*(2), 155-183.

Kidde, J., & Alfred, R. (2011). *Restorative justice: A working guide for our schools.* Alameda County School Health Services.

Kim, C. Y., Losen, D. J., & Hewitt, D. T. (2010). *The school-to-prison pipeline: Structuring legal reform*. NYU Press.

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., & Horner, R. (2017). Changing systems to scale up an evidence-based educational intervention. *Evidence-Based Communication Assessment and Intervention, 11*(3-4), 99-113. https://doi.org/10.1080/17489539.2017.1376383

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kurtz, H., Lloyd, S., Harwin, A., & Osher, M. (2018). School Policing: Results of a National Survey of School Resource Officers. *Editorial Projects in Education*.

Lane, K. L., Oakes, W., Menzies, H. M. (2014). Comprehensive, integrated, three-tiered models of prevention: Why does my school—and district—need an integrated approach to meet students' academic, behavioral, and social needs? *Preventing School Failure, 58(3)*, 121-128. https://doi.org/10.1080/1045988X.2014.893977

Leachman, M., & Mai, C. (2014). *Most states funding schools less than before the recession*. Washington, DC: Center on Budget and Policy Priorities.

Lee, A., Gage, N., Mcleskey, J., & Huggins-Manley, A. (2021). The impacts of School-Wide Positive Behavior Interventions and Supports on school discipline outcomes for diverse students. *The Elementary School Journal*. https://doi.org/10.1086/712625

Losen, D. L., Hewitt, D., & Toldson, I. (2014). *Eliminating excessive and unfair discipline in schools: Policy recommendations for reducing disparities—Discipline disparities: A research-to-practice collaborative.* Center for Evaluation and Education Policy, The Equity Project at Indiana University.

Losen, D. J., Keith, M. A., II, Hodson, C. L., & Martinez, T. E. (2016). *Charter schools, civil rights and school discipline: A comprehensive review.* The Center for Civil Rights Remedies at The Civil Rights Project – Proyecto Derechos Civiles.

Losen, D. J., Martinez, T., & Gillespie, J. (2012). *Suspended education in California.* The Center for Civil Rights Remedies at The Civil Rights Project – Proyecto Derechos Civiles.

Love, E., Morency, M., Miller, K., Onyeka, O., & Richards, M. (2018, May 18). How to involve young people in afterschool programs. *Scholars Strategy Network*. https://scholars.org/brief/how-involve-young-people-afterschool-programs

Luiselli, J. K., Putnam, R. F., Handler, M. W. & Feinberg, A. B. (2005). Whole-school positive behaviour support: effects on student discipline problems and academic performance. *Educational Psychology: An International Journal of Experimental Educational Psychology,* 25(2-3), 183-198. http://dx.doi.org/10.1080/0144341042000301265

Lustick, H. (2020a). Culturally responsive restorative discipline. *Educational Studies, 56*(6), 555-583. https://doi.org/10.1080/00131946.2020.1837830

Lustick, H. (2020b). Going restorative, staying tough: Urban principals' perceptions of restorative practices in collocated small schools. *Education and Urban Society*. https://doi.org/10.1177/0013124520974335

Lustick, H., Norton, C., Lopez, S. R., & Greene-Rooks, J. H. (2020c). Restorative practices for empowerment: A social work lens. *Children & Schools, 42*(2), 89-97. https://doi.org/10.1093/cs/cdaa006

Maeng, J. L., Cornell, D., & Huang, F. (2020). Student threat assessment as an alternative to exclusionary discipline. *Journal of School Violence, 19*(3), 377-388. https://doi.org/10.1080/15388220.2019.1707682

Mallett, C. A. (2016). The school-to-prison pipeline: A critical review of the punitive paradigm shift. *Child and Adolescent Social Work Journal, 33*(1), 15-24.

Mallett, C. A. (2017). The school-to-prison pipeline: Disproportionate impact on vulnerable children and adolescents. *Education and Urban society*, *49*(6), 563-592.

Mansfield, K. C., Fowler, B., & Rainbolt, S. (2018). The potential of restorative practices to ameliorate discipline gaps: The story of one high school's leadership team. *Educational Administration Quarterly, 54*(2), 303-323. https://doi.org/10.1177/0013161X17751178

Marchbanks III, M. P., Peguero, A. A., Varela, K. S., Blake, J. J., & Eason, J. M. (2018). School strictness and disproportionate minority contact: Investigating racial and ethnic disparities with the "School-to-Prison Pipeline". *Youth Violence and Juvenile Justice, 16(*2), 241-259.

Marsh, S. (2014). School pathways to juvenile justice system project. Reno, NV: *National Council of Juvenile and Family Court Judges*.

Martinez, S. (2009). A system gone berserk: How are zero-tolerance policies really affecting schools? *Preventing School Failure, 53*(3), 153–157.

McCluskey, G., Lloyd, G., Kane, J., Riddell, S., Stead, J., & Weedon, E. (2008). Can restorative practices in schools make a difference?. *Educational Review*, *60*(4), 405-417.

McDaniel, S. C., Bruhn, A. L., & George, H. P. (2020). Establishing the literature base of SWPBIS research in schools with large populations of underrepresented students. *Preventing School Failure, 64*(4), 290-300. https://doi.org/10.1080/1045988X.2020.1762158

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30

McIntosh, K., Girvan, E. J., Horner, R. H., Smolkowski, K., & Sugai, G. (2014). Recommendations for addressing discipline disproportionality in education. Eugene, OR: *OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports*.

McNeal, L., & Dunbar Jr, C. (2010). In the eyes of the beholder: Urban student perceptions of zero tolerance policy. *Urban Education, 45*(3), 293-311.

Merrell, K. W., & Gueldner, B. A. (2010). *Social and emotional learning in the classroom: Promoting mental health and academic success*. The Guilford Press.

Mielke, M., Farrington, D. P. (2021). School-based interventions to reduce suspension and arrest: A meta-analysis. *Aggression and Violent Behavior, 56*. https://doi.org/10.1016/j.avb.2020.101518

Mirsky, L. (2007). SaferSanerSchools: Transforming school cultures with restorative practices. *Reclaiming Children and Youth*, *16*(2), 5.

Mirsky, L., & Wachtel, T. (2008). The worst school I've ever been to: Empirical evaluations of a restorative school and treatment milieu. *Reclaiming Children & Youth*, *16*(2).

Mitchell, M. M., & Bradshaw, C. P. (2013). Examining classroom influences on student perceptions of school climate: The role of classroom management and exclusionary discipline strategies. *Journal of School Psychology, 51*(5), 599-610. https://doi.org/10.1016/j.jsp.2013.05.005

Morrison, B. E., & Vaandering, D. (2012). Restorative justice: Pedagogy, praxis, and discipline. *Journal of School Violence*, *11*(2), 138-155.

Morrissey, K. L., Bohanon, H., & Fenning, P. (2010). Positive Behavior Support: Teaching and acknowledging expected behaviors in an urban high school. *TEACHING Exceptional Children, 42*(5), 26-35. https://doi.org/10.1177/004005991004200503

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

Na, C., & Gottfredson, D. C. (2013). Police officers in schools: Effects on school crime and the processing of offending behaviors. *Justice Quarterly*, *30*(4), 619-650.

Nance, J. P. (2015). Students, police, and the school-to-prison pipeline. *Wash. UL Rev., 93*, 919.

National Women's Law Center and the NAACP Legal Defense and Education Fund. (2014). *Unlocking opportunity for African American girls: A call to action for educational equity.* https://www.nwlc.org/sites/default/files/pdfs/unlocking_opportunity_for_african_american_girls_final.pdf

National Child Traumatic Stress Network Schools Committee. (October 2008). *Child trauma toolkit for educators.* National Center for Child Trauma Stress.

Nekvasil, E. K., & Cornell, D. G. (2015). Student threat assessment associated with safety in middle schools. *Journal of Threat Assessment and Management, 2*(2), 98-113. https://doi.org/10.1037/tam0000038

Nese, R. N. T., Bastable, E., Gion, C., Massar, M., Nese, J. F. T., & McCroskey, C. (2020). Preliminary analysis of an instructional alternative to exclusionary discipline. *Journal of At-Risk Issues, 23*(1), 1–14.

Nocera, E. J., Whitbread, K. M., & Nocera, G. P. (2014). Impact of School-Wide Positive Behavior Supports on student behavior in the middle grades. *RMLE Online: Research in Middle Level Education, 37*(8). https://doi.org/10.1080/19404476.2014.11462111

Noyes, J., Booth, A., Cargo, M., Flemming, K., Harden, A., Harris, J., Garside, R., Hannes, K., Pantoja, T., & Thomas, J. (2021). Chapter 21: Qualitative Evidence. In Higgins, J.P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., & Welch, V.A. (Eds). *Cochrane Handbook for Systematic Reviews of Interventions version 6.2*. John Wiley & Sons.

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

Okonofua, J. A., Paunesku, D., & Walton, G. M. (2016). Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proceedings of the National Academy of Sciences*, *113*(19), 5221-5226. https://doi.org/10.1073/pnas.1523698113

Ortega, L., Lyubansky, M., Nettles, S., & Espelage, D. L. (2016). Outcomes of a restorative circles program in a high school setting. *Psychology of Violence, 6*(3), 459-468. https://doi.org/10.1037/vio0000048

OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports (2021). Positive Behavioral Interventions & Supports. www.pbis.org.

Osher, D. M., Poirier, J. M., Jarjoura, R. G., Brown, R., & Kendziora, K. (2014). Avoid simple solutions and quick fixes: Lessons learned from a comprehensive districtwide approach to improving student behavior and school safety. *Journal of Applied Research on Children, 5*(2).

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., …, Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*(71). https://doi.org/10.1136/bmj.n71

Payne, A. A., & Welch, K. (2018). The effect of school conditions on the use of restorative justice in schools. *Youth Violence and Juvenile Justice*, *16*(2), 224-240.

Perinetti, G. (2018). StaTips Part IV: Selection, interpretation and reporting of the intraclass correlation coefficient. *South European Journal of Orthodontics and Dentofacial Research, 5*(1), 3-5. https://doi.org/10.5937/sejodr5-17434

Petras, H., Masyn, K. E., Buckley, J. A., Ialongo, N. S., & Kellam, S. (2011). Who is most at risk for school removal? A multilevel discrete-time survival analysis of individual-and context-level influences. *Journal of Educational Psychology*, *103*(1), 223- 237.

Petrosino, A., Guckenburg, S., DeVoe, J., & Hanson, T. (2010). What Characteristics of Bullying, Bullying Victims, and Schools Are Associated with Increased Reporting of Bullying to School Officials? Issues & Answers. REL 2010-No. 092. *Regional Educational Laboratory Northeast & Islands*.

Piquero, A. R. (2008). Disproportionate minority contact. *The Future of Children*, 59-79.

Pranis, K. (2005). *The little book of circle processes*. Old Approach to Peacemaking.

Rainbolt, S., Fowler, E. S., & Mansfield, K. C. (2019). High school teachers' perceptions of restorative discipline practices. *NASSP Bulletin, 103*(2), 158-182. https://doi.org/10.1177/0192636519853018

Reardon, S. F., Grewal, E. T., Kalogrides, D., & Greenberg, E. (2012). B rown Fades: The End of Court-Ordered School Desegregation and the Resegregation of American Public Schools. *Journal of Policy Analysis and Management, 31*(4), 876-904.

Reeves, D. B. (2010). Transforming professional development into student results. ASCD.

Riddle, T., & Sinclair, S. (2019). Racial disparities in school-based disciplinary actions are associated with county-level rates of racial bias. *Proceedings of the National Academy of Sciences, 116*(17), 8255-8260. https://doi.org/10.1073/pnas.1808307116

Ripski, M. B., & Gregory, A. (2009). Unfair, unsafe, and unwelcome: Do high school students' perceptions of unfairness, hostility, and victimization in school predict engagement and achievement? *Journal of School Violence, 8*(4), 355-375.

Rudasill, K. M., Snyder, K. E., Levinson, H., & Adelson, J. L. (2018). Systems view of school climate: a theoretical framework for research. *Educational Psychology Review, 30*(1), 35-60.

Rumberger, R. W., & Losen, D. J. (2016). *The high cost of harsh discipline and its disparate impact,* The Center for Civil Rights Remedies at The Civil Rights Project – Proyecto Derechos Civiles.

Safer, D. J., & Malever, M. (2000). Stimulant treatment in Maryland public schools. *Pediatrics, 106*(3), 533-539.

Sandwick, T., Hahn, J. W., & Ayoub, L. H. (2019). Fostering community, sharing power: Lessons for building restorative justice school cultures. *Education Policy Analysis Archives, 27*(145), 1-32. https://doi.org/10.14507/epaa.27.4296

Scott, T. M., & Barrett, S. B. (2004). Using staff and student time engaged in disciplinary procedures to evaluate the impact of school-wide PBS. *Journal of Positive Behavior Interventions, 6*(1), 21-27. https://doi.org/10.1177/10983007040060010401

Sharkey, J. D., & Fenning, P. A. (2012). Rationale for designing school contexts in support of proactive discipline. *Journal of School Violence, 11*(2), 95-104.

Simonsen, B., Eber, L., Black, A. C., Sugai, G., Lewandowski, H., Sims, B., & Myers, D. (2012). Illinois statewide Positive Behavioral Interventions and Supports: Evolution and impact on student outcomes across years. *Journal of Positive Behavior Interventions, 14*(1), 5-16. https://doi.org/10.1177/2332858417711428

Sinclair, M. F., Christenson, S. L., & Thurlow, M. L. (2005). Promoting school completion of urban secondary youth with emotional or behavioral disabilities. *Exceptional Children, 71*(4), 465–482.

Sirriyeh, R., Lawton, R., Gardner, P., & Armitage, G. (2012). Reviewing studies with diverse designs: the development and evaluation of a new tool. *Journal of Evaluation in Clinical Practice, 18*(4), 746-752. https://doi.org/10.1111/j.1365-2753.2011.01662.x

Skiba, R. J. (2000). *Zero tolerance, zero evidence: An analysis of school disciplinary practice* (Policy Research Report #SRS2). Bloomington, IN: Indiana Education Policy Center.

Skiba, R. J., Arrendonda, M. I., & Rausch, M. K. (2014). *New and developing research on disparities in discipline.* The Equity Project at Indiana University.

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

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

Smolkowski, K., Girvan, E. J., McIntosh, K., Nese, R. N., & Horner, R. H. (2016b). Vulnerable decision points for disproportionate office discipline referrals: Comparisons of discipline

for African American and White elementary school students. *Behavioral Disorders*, *41*(4), 178-195.

Solomon, B. G., Klein, S. A., Hintze, J. M., Cressley, J. M., & Peller, S. L. (2012). A meta-analysis of school-wide positive behavior support: An exploratory study using single-case synthesis. *Psychology in the Schools, 49*(2), 105-121. https://doi.org/10.1002/pits.20625

Sprick, R. S., Garrison, M., & Howard, L. (2002). Foundations: Establishing positive discipline policies (2nd ed.). Pacific Northwest Publishing.

Staats, C. (2014). Implicit racial bias and school discipline disparities. http://kirwaninstitute.osu.edu/wp-content/uploads/2014/05/ki-ib-argument-piece03.pdf.

Steinberg, M. P., & Lacoe, J. (2017). What do we know about school discipline reform? Assessing the alternatives to suspensions and expulsions. *Education Next, 17*(1), 44-53.

Stinchcomb, J. B., Bazemore, G., & Riestenberg, N. (2006). Beyond zero tolerance: Restoring justice in secondary schools. *Youth Violence and Juvenile Justice, 4*(2), 123-147. https://doi.org/10.1177/1541204006286287

Stoiber, K., & Gettinger, M. (2016). Multi-tiered systems of support and evidence-based practices. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of Response to Intervention* (pp. 121-141). Springer.

Styles, B. & Torgerson, C. (2018). Randomised controlled trials (RCTs) in education research – methodological debates, questions, challenges. *Educational Research, 60*(3), 255-264. https://doi.org/10.1080/00131881.2018.1500194

Sugai, G., & Horner, R. H. (2009). Defining and describing schoolwide positive behavior support. In W. Sailor, Dunlap, G., Sugai, G., & R. Horner (Eds.), *Handbook of positive behavior support* (pp. 307-326). Springer, Boston, MA.

Sugai, G., Simonsen, B., Freeman, J., & La Salle, T. (2016). Capacity development and multi-tiered systems of support: Guiding principles. *Australasian Journal of Special and Inclusive Education, 40*(2), 80-98. https://doi.org/10.1017/jse.2016.11

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development, 88*(4), 1156-1171.

Teske, S. C., Huff, B., & Graves, C. (2013). Collaborative role of courts in promoting outcomes for students: The relationship between arrests, graduation rates, and school safety. *Family Court Review, 51*(3), 418-426. https://doi.org/10.1111/fcre.12038

Thomas, J., & Harden, A. (2008). Methods for the thematic synthesis of qualitative research in systematic reviews. *BMC Medical Research Methodology, 8*(45). https://doi.org/10.1186/1471-2288-8-45

Thorsborne, M., & Blood, P. (2013). *Implementing restorative practice in schools: A practical guide to transforming school communities*. Jessica Kingsley Publishers.

Torres, M. S., Jr., & Stefkovich, J. A. (2009). Demographics and police involvement: Implications for student civil liberties and just leadership. *Educational Administration Quarterly, 45*(3), 450-473.

Turpin, T., & Hardin, D. T. (1997). A case study of an in-school suspension program in a rural high school setting. *Research in the Schools, 4*(2), 57-63.

United States Department of Education. (2014). *Guiding Principles: A Resource Guide for Improving School Climate and Discipline.*

Office for Civil Rights (2018). *2015-2016 Civil rights data collection (CRDC)*[Data file]. https://www2.ed.gov/about/offices/list/ocr/docs/crdc-2015-16.html

Ujifusa, A. (2018, December 4). House lawmakers seek federal money for more school police officers. *Education Week.*

Valdebenito, S., Eisner, M., Farrington, D. P., Ttofi, M. M., & Sutherland, A. (2018). School-based interventions for reducing disciplinary school exclusion: A systematic review. *Campbell Systematic Reviews, 14*(1). https://doi.org/10.4073/csr.2018.1

Vara-Orta, F. (2018, December 21). It's official: DeVos has axed Obama discipline guidelines meant to reduce suspensions of students of color. *Chalkbeat.*

von der Embse, N., Rutherford, L., Mankin, A., & Jenkins, A. (2019). Demonstration of a trauma-informed assessment to intervention model in a large urban school district. *School Mental Health, 11*(2), 276-289. https://doi.org/10.1007/s12310-018-9294-z

Wadhwa, R. (2016). New phase of internationalization of higher education and institutional change. *Higher Education for the Future*, *3*(2), 227-246.

Wald, J., & Losen, D. J. (2003). Defining and redirecting a school-to-prison pipeline. *New Directions for Youth Development, 99,* 9-15.

Ward, B., & Gersten, R. (2013). A randomized evaluation of the Safe and Civil Schools model for Positive Behavioral Interventions and Supports at elementary schools in a large urban school district. *School Psychology Review, 42*(3), 317-333. https://doi.org/10.1080/02796015.2013.12087476

Warren, J. S., Bohanon-Edmonson, H. M., Turnbull, A. P., Sailor, W., Wickham, D., Griggs, P., & Beech, S. E. (2006). School-Wide Positive Behavior Support: Addressing behavior problems that impede student learning. *Educational Psychology Review, 18*(2), 187-198.

Welch, K., & Payne, A. A. (2010). Racial threat and punitive school discipline. *Social Problems, 57*(1), 25-48.

Welsh, R. O., & Little, S. (2018). The school discipline dilemma: A comprehensive review of disparities and alternative approaches. *Review of Educational Research, 88*(5), 752-794.

Ysseldyke, J. E., & Christenson, S. (2002). *Functional assessment of academic behavior: Creating successful learning environments*. Sopris West.

Zehr, H. (2015). The little book of restorative justice. Intercourse, PA: Good Books.

Zellerer, E. (2013). Realizing the potential of restorative justice. *Reconstructing Restorative Justice Philosophy*, 269-290.

Zhang, Y., Cook, C. R., & Lyon, A. R. (2021). A simple matter of time? School-level analysis of the relationship between time allocation, treatment integrity, and student outcome. *School Mental Health.* https://doi.org/10.1007/s12310-021-09412-2

Zuckerman, J. T. (2007). Classroom management in secondary schools: A study of student teachers' successful strategies. *American Secondary Education, 35*, 4-16.

VITA

Dr. DiClemente was born in Houston, Texas and raised in Ellicott City, Maryland. Before attending Loyola University Chicago, she attended Franklin and Marshall College in Lancaster, Pennsylvania where she earned a Bachelor of Arts in Psychology, with great distinction, in 2013.

While at Loyola, Dr. DiClemente has served as Vice President and President of the Clinical Student Association, and on several committees, including the Social Committee and the Diversity Committee. Dr. DiClemente also competed in the Three Minute Thesis competition at the graduate school level in 2017. She received the American Psychological Foundation/COGDOP Graduate Student Research Scholarship in 2019 to conduct her dissertation research.

Currently, Dr. DiClemente is completing her clinical psychology doctoral internship at Lucile Packard Children's Hospital at Stanford / Children's Health Council in Palo Alto, CA. She lives in Mountain View, CA.