June 2022

Equitable Implementation of the Good Behavior Game

Faith D. Reynolds
University of South Florida

Equitable Implementation of the Good Behavior Game

by

Faith D. Reynolds

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Educational and Psychological Studies
College of Education
University of South Florida

Major Professor: Evan Dart, Ph.D.
Nathaniel von der Embse, Ph.D.
John Ferron, Ph.D.
Donald Kincaid, Ed.D.

Date of Approval:
June 17, 2022

Keywords: school intervention, equity, classroom behavior management

Table of Contents

List of Tables

List of Figures

Abstract

Exclusionary discipline actions in schools (e.g., suspension, expulsion) are related to short- and long-term negative outcomes for students. For the past several decades, Black students have received disproportionate disciplinary actions in schools compared to their peers. Classroom behavior interventions are an alternative practice to traditional disciplinary actions and are important tools for reducing overall levels of exclusionary discipline. It is imperative to ensure equitable implementation is being used within these behavior interventions to support all students in the classroom. The Good Behavior Game (GBG) is a widely used classroom behavior intervention that has been described as a universal behavioral vaccine. Although effective at reducing traditional punitive discipline actions overall, it is unclear if the GBG is being implemented equitably. This study investigated if the GBG is being implemented equitably in classrooms and, if not, whether teachers can be trained to implement it equitably. Three teachers and their respective classrooms participated using a nonconcurrent multiple baseline design across classrooms. Findings suggest that teachers are implementing the GBG at disproportionate levels across racially diverse students. The brief training incorporating performance feedback, self-monitoring, and self-awareness was not effective for shifting these teachers to equitable implementation. This study also sought to understand if the inequitable and equitable implementation of the GBG had a differential impact on student behavior and discipline outcomes for students based on their race. The current study was unable to answer this aim. The findings of this study further the knowledge of equitable classroom behavior management practices.

Chapter I: Introduction

The U.S. Department of Education (2014, 2016) reported that children of color are disproportionately disciplined across public schools nationwide. Students of color, particularly Black students, are three times more likely than their White peers to be expelled or suspended and are more likely to receive harsher discipline actions for subjective reasons such as disrespect (American Psychological Association (APA), 2008; Bal, 2016; Girvan et al., 2017; Reno et al., 2018; Skiba et al., 2002; Smolkowski et al, 2016; U.S. Department of Education, 2014). These disproportionate uses of discipline actions perpetuate racial segregation and lead to disproportionate academic outcomes for students of color (APA, 2008; Bal, 2016). Negative outcomes such as academic failure, school dropout, entering the juvenile justice system, and identification for special education have been associated with exclusionary discipline actions (i.e., suspension, expulsion; Bal, 2016; Gregory et al., 2010). It is imperative to reduce disproportionate discipline actions to increase positive outcomes for all students.

Research and training on equitable practices for school disciplinary procedures can help reduce levels of disproportionality within discipline outcomes. Positive Behavioral Interventions and Supports (PBIS) is a framework implemented in schools that is aimed at reducing exclusionary discipline actions and increasing student social and academic outcomes. The National Technical Support Center on PBIS refers to equitable practices in schools as educational practices and policies that provide support and opportunities to students that result in similar outcomes across individual characteristics and cultural backgrounds. Factors such as implicit bias and cross-cultural misunderstandings have been found to lead to inequitable

discipline actions and produce negative outcomes for non-White students (McIntosh et al., 2014). Although PBIS is an effective method for reducing discipline actions overall, research suggests that PBIS practices do not reduce disproportionate discipline outcomes (Sandomierski, 2011). Sandomierski (2011) found no significant relationship between school PBIS implementation fidelity and levels of disproportionality for office referrals for Black students ($p = 0.1493$). There were also no significant relationships between PBIS implementation fidelity and office discipline referrals for Latino students or suspensions for Black or Latino students (Sandomierski, 2011). Barclay (2017) evaluated the critical elements of PBIS practices that are related to equitable discipline outcomes. Two of those critical elements, classroom systems, and recognition programs, were substantially related to equitable discipline outcomes. The critical element of classroom systems refers to PBIS practices being implemented in the classroom. Recognition programs refer to practices that recognize and reward students for engaging in expected behaviors. Specifically, Barclay found that office discipline referral (ODR) risk for Black students was significantly related ($b = -0.220$, $SE = 0.065$, $p = .001$) to implementation of PBIS classroom systems and higher fidelity of these classroom systems was significantly related to lower out of school suspension risk for Black ($b = -0.145$, $SE = 0.036$, $p < .001$) and Hispanic ($b = -0.057$, $SE = 0.020$, $p = .005$) students. PBIS recognition programs were found to be significantly related to lower out of school suspension (OSS) ratios for Black ($b = -2.414$, $SE = 0.816$, $p < 0.002$) and Hispanic ($b = -2.418$, $SE = 0.814$, $p < 0.003$) students. PBIS practices have been evaluated for equitable outcomes and findings suggest that two of the critical elements for implementation, classroom systems, and recognition programs, are important key factors for reducing inequities.

Classroom behavior management interventions are classroom systems that can include recognition programs and reduce the use of exclusionary discipline actions in the classroom. Fallon and colleagues (2012) identified empirically supported classroom behavior management interventions as important tools for addressing disproportionate discipline actions, particularly with the inclusion of positive practices. The Good Behavior Game (GBG; e.g., Barrish et al., 1969) is a classroom behavior management intervention that could be aligned with the PBIS critical components related to equitable outcomes, classroom systems, and reward programs. During the GBG, students are grouped into teams and the intervention is typically in place for a short but specific period of instructional time. During this time, students can receive points for their team as a consequence of engaging in problem behavior related to the classroom rules and expectations that are defined and explicitly taught before the game. After the GBG time period is over, students on teams with fewer than a prespecified point target earn a reward.

There are many variations for implementation of the GBG, some of which include positive and negative consequences (Joslyn, 2019). One of the variations involves a positive component where teachers can remove points from a student's team who is engaging in appropriate behavior. Another version, coined the Caught Being Good Game, involves points being provided to students engaging in appropriate behavior and teams earning rewards at the end of the predetermined time period if they have earned over a specified number of points (e.g., Wright & McCurdy, 2012). These variations have been directly compared to evaluate their effectiveness on student behavior and results suggest similar effectiveness between the variations (Wright & McCurdy, 2012; Wahl et al., 2016). GBG is effective across multiple settings and populations and is widely used across many types of classrooms (Embry, 2002). Long-term follow-up studies have also found that the GBG impacts the behavior of students in the future

such as reducing impulsive behaviors, substance use, and antisocial behaviors (Embry, 2002; Kellam & Anthony, 1998; Kellam et al., 1998). Due to the immediately effective and long-term outcomes associated with implementation, GBG has been referred to as a universal behavioral vaccine (Embry, 2002).

Current Study

Although the GBG is effective for reducing problem behavior in the classroom and is acceptable by teachers and students, it is unclear if teachers implement the GBG equitably or if outcomes are equitable across racially diverse students. Interventions that effectively reduce discipline rates do not necessarily reduce disproportionality in discipline actions. For example, school-wide PBIS outcomes have indicated that although their practices are highly effective in reducing overall discipline action outcomes across schools, there is still a disproportionate use of discipline actions across racial identities (Sandomierski, 2011). Currently, there is limited research on equitable use and outcomes of classroom behavior management interventions. Therefore, the purpose of this study is to understand if the GBG is implemented equitably and, if not, whether teachers can be trained to implement it equitably. Also, this study seeks to understand if the equitable implementation of the GBG has impacts on student behavior and discipline outcomes for racially diverse students. It is hypothesized that teachers who implement the GBG inequitably across racially diverse students will shift to equitable implementation following a brief training incorporating performance feedback, self-monitoring, and self-awareness. Further, it is hypothesized that inequitable and equitable implementation of the GBG will differentially impact the behavior and discipline outcomes of students based on their race. Research questions for this study include:

(a) Do teachers implement the GBG equitably across racially diverse students?

(b) Can teachers be trained to implement the GBG equitably across racially diverse students?

(c) Does equitable implementation of the GBG have an equitable impact on student behavior and discipline outcomes of racially diverse students?

Chapter II: Review of the Literature

The following will provide an overview of the literature investigating traditional school discipline practices, disproportionality within the implementation of those practices, efforts to reduce the use of traditional discipline practices and disproportionality, and the GBG.

Discipline

Exclusionary discipline involves any type of disciplinary action that removes or excludes a student from their typical educational setting. Discipline actions used in schools that are reported to Civil Rights Data Collection (CRDC) include out-of-school suspension, in-school suspension, expulsion, referral to law enforcement, school-related arrests, and corporal punishment. CRDC (2014) defines out-of-school suspension as "an instance in which a child is temporarily removed (one school day or longer) from their regular school for disciplinary purposes to another setting (e.g., home, behavior center)" (p.21). In-school suspension is defined as "instances in which a child is temporarily removed from their regular classroom(s) for at least half a day but remains under the direct supervision of school personnel" (p. 21). Expulsion is defined as "an action taken by the local educational agency removing a child from their regular school for disciplinary purposes, with the continuation of educational services, for the remainder of the school year or longer in accordance with local educational agency policy" (p.22). Referral to law enforcement is "an action by which a student is reported to any law enforcement agency or official, including a school police unit, for an incident that occurs on school grounds, during school-related events, or while taking school transportation, regardless of whether official action is taken" (p.22). A school-related arrest is "an arrest of a student for any activity conducted on

school grounds, during off-campus school activities (including while taking school transportation), or due to a referral by any school official" (p.22). The American Academy of Child and Adolescent Psychiatry (2014) defines corporal punishment as "a discipline method in which a supervising adult deliberately inflicts pain upon a child in response to a child's unacceptable behavior and/or inappropriate language."

Allman and Slate (2011) report how the use of disciplinary actions in schools to reduce student misbehavior has evolved from the 19th century until now. In the 19th century, corporal punishment was the most common form of disciplinary action and has been declining in its use and acceptability since the mid-1900s. Currently, corporal punishment is only legal, with restrictions, in public schools in 19 states (Gershoff & Font, 2018). In the 1960s, out-of-school suspension began as a school disciplinary practice and is still commonly used despite evidence of its ineffectiveness for reducing future student misbehavior. In the late 1980s, zero-tolerance policies were developed in response to school violence and shootings, where students could be expelled from school for actions such as bringing a weapon or drugs onto school property. Currently, school discipline practices are moving towards preventative measures and positive reinforcement in addition to zero-tolerance policies and the discipline actions defined above. In the 2011-2012 school year, 49 million students were enrolled in public schools. Of those students, 3.5 million received in-school suspensions, 3.45 million received out-of-school suspensions, and 130,000 were expelled, according to the U.S. Department of Education. In the 2015-2016 school year, 50.6 million students were enrolled in U.S. public schools. Of those students, about 2.7 million students received in-school suspensions, 2.55 million received one or more out-of-school suspensions, and 120,800 students received an expulsion (CRDC, 2018). Harper and colleagues (2019) reviewed the available data on school discipline actions provided

by CRDS to assess trends across the three points of available data collection in the 2011-2012, 2013-2014, and 2015-2016 academic school years. Trends show that public schools in the U.S. have decreased out-of-school suspension in most states, with eight states seeing increases in out-of-school suspensions. Students enrolled in schools who received out-of-school suspensions decreased from 5.6% to 4.7%. These decreasing trends could be due to education officials' dedication to reducing exclusionary discipline actions through various legislations requiring disciplinary data to be reported over the past decade.

The impact of exclusionary discipline actions on students in schools can be lifelong. Several short-term and long-term negative outcomes have been linked to receiving certain types of disciplinary actions in the school system. The U.S. Department of Health and Human Services and the U.S. Department of Education (2016) stated that students who receive exclusionary discipline actions such as suspension or expulsion at a young age are likely to continue to receive discipline actions throughout their academic career. There is also evidence to suggest that students who have been suspended from school are more likely to lose educational gains, graduate later than expected, drop out of school, fail a grade level, and become involved in the juvenile justice system (Fabelo et al., 2011; Skiba et al., 2014). Exclusionary discipline actions in the school system do not only impact the individual throughout their life but also have been found to have impacts on society. The use of exclusionary discipline actions in schools has repercussions on state costs in areas such as tax revenue, crime, welfare, and health (Marchbanks et al., 2014). For example, individuals who dropped out of high school due to disciplinary actions in California cost the state $2.7 billion (Rumberger & Losen, 2017). Another study found that students who dropped out of high school earned $200,000 less in their lifetime than students who did not drop out of high school (Belfield, 2014).

Various risk factors for exclusionary discipline actions have been identified through the literature at the community, family, school, classroom, and student level. At the community level, socioeconomic status (SES) is a predictor of exclusionary discipline actions (Hemphill et al., 2014). At the family level, SES, parental education level, and parental involvement are related to exclusionary discipline risk for students (Hemphill et al., 2014; McElderry & Cheng, 2014). Risk factors at the school level include variables such as school size, student racial composition, educator race, behavior support practices, and availability of mental health staff (Anyon et al., 2014; Finn & Servoss, 2014; Gilliam, 2005; Skiba et al., 2014). Exclusionary discipline actions are also related to classroom-level variables including classroom size, teacher race, and teacher practices (Bradshaw et al., 2010; Gregory et al., 2014; Martinez et al., 2016). At the student level, variables including gender, academic achievement, special education status, ethnicity/race, and types of behavior are associated with exclusionary discipline risk (Bowman-Perrott, 2013; Bradshaw et al., 2010; Gregory et al., 2010). Many of these various disciplinary action risk factors have been found to impact underrepresented children at higher rates (Nowicki, 2018; Wallace et al., 2008; Welch & Payne, 2010, 2015), leading to disproportionate use of disciplinary actions across students.

Disproportionality

The U.S. Department of Education and the U.S. Department of Justice have reported significant racial disparities in student disciplinary actions, where Black students receive exclusionary discipline actions three times greater than their White peers. Racial disparities in discipline actions within schools have been reported since the 1970s (Children's Defense Fund, 1975). In the 2015-2016 school year, Black students received 40.6% of the out-of-school suspensions and White students received 31.7% of the out-of-school suspensions. In 2013-3014,

40.3% of students who received out-of-school suspensions were Black and 32.6% were White. Harper and colleagues (2019) found that although exclusionary discipline actions have decreased overall throughout the U.S., Black students are still experiencing exclusionary discipline at rates far higher than their peers. In the 2015-2016 school year, Black students were suspended twice as much as White students. The United States Government Accountability Office (GAO) completed a study analyzing the disciplinary action data in nearly all public schools during the 2013-2014 school year from the Educations Civil Rights Data Collection. These data were examined to determine patterns of disciplinary actions in public schools. The GAO study reported that Black students, boys, and students with disabilities were found to receive disciplinary actions at disproportionate levels when compared to their peers.

The types of disciplinary action that were examined in the GAO study included out-of-school suspension, in-school suspension, expulsion, referral to law enforcement, corporal punishment, and school-related arrests. For all types of disciplinary actions examined, Black students were overrepresented. Although Black students represented 15.5% of all public-school students, they accounted for 38.7% of students suspended from schools, which is an overrepresentation of 23.2%. Black students were overrepresented 23.2% for out-of-school suspension, 16.4% for in-school suspension, 10.4% for referral to law enforcement, 14.6% for expulsion, 22.1% for corporal punishment, and 19.4% for school-related arrests. Black students were the only racial group to have disproportionate discipline actions across all types of disciplinary actions for both boys and girls. For all types of disciplinary actions examined, White students were underrepresented. Although White students represented 50.3% of all public-school students, they accounted for 32.5% of students suspended from schools, which is an underrepresentation of about -17.8%. White students were underrepresented -17.8% for out-of-

school suspension, -11.8% for in-school suspension, -12.3% for referral to law enforcement, -6.5% for expulsion, -0.2% for corporal punishment, and -16.8% for school-related arrests. The level of school poverty was also examined for patterns of discipline actions across racial groups. Black students were found to be suspended at disproportionately higher rates than their peers in all four levels of school poverty that were examined (i.e., overrepresented between 12.2% and 24.6%). Similarly, White students were found to be underrepresented in all levels of school poverty for suspension (i.e., underrepresented between -3.2% and -14.2%). The type of school setting was also examined for discipline actions across racial groups. For the traditional, magnet, charter, alternative, and special education school settings, Black students were overrepresented (i.e., between 11.4% and 31.6%) in suspensions and White students were underrepresented (i.e., between -8.3% and -19.7%).

As mentioned previously, risk factors for receiving discipline actions impact underrepresented students at higher rates and could also be contributing to disproportionate discipline actions. SES is related to discipline action use and students of color were found to have risk reduced when SES was taken into consideration (Wallace et al., 2008). School-level risk factors such as the percentage of Black students and principal attitudes towards discipline actions have been found to contribute to disproportionate discipline actions (Skiba et al., 2014). Also, schools with higher percentages of Black students were found more likely to implement harsher discipline actions (Welch & Payne, 2010). Black teachers have been found to rate externalizing behaviors more favorably and rate lower levels of problem behavior among Black students compared to White teachers (Bates & Glick, 2013; Downey & Pribesh, 2004). Additionally, Bradshaw and colleagues (2010) found that students' overall risk of being referred for disciplinary action is reduced with Black teachers. Another study found that teachers have

lower academic and social expectations of Black and Latino students compared to White and Asian students (Tenenbaum & Ruck, 2007). Further, classroom-level variables such as teachers providing emotional support to students, having classroom organization, and providing instructional support were found to reduce levels of disproportionate discipline actions (Gregory et al., 2014). Black students have also been found to receive harsher discipline actions for the same behaviors as White students (Bradshaw et al., 2010) and are four times more likely to receive referrals for subjective reasons when compared to White peers (Skiba et al., 2011). Student skin tone was also assessed for disproportionality by Hannon and colleagues (2013) and results found that students with the darkest skin tone were three times more likely to receive suspension compared to students with lighter skin tone.

School staff and cultural factors that contribute to disproportionality in school discipline actions have been identified throughout the literature. Some of these include a mismatch of culture or misunderstanding between teachers and students, implicit bias, racial stereotyping by staff, conscious or unconscious racial bias by teachers, and teachers being unprepared to manage class behavior (Bradshaw et al., 2010; Skiba et al., 2011). Although students in the U.S. are very diverse, educators in the U.S. are largely White (Goldring et al., 2013). This contributes to the mismatch of culture between students and teachers which can lead to disproportionate discipline actions. Cross-cultural misunderstandings occur when a teacher's cultural background differs from the students, and the behavior of the student is perceived to have different intentions or acceptability by the teacher. For example, an emotive interaction from the student could be perceived as being combative by the teacher, or vice versa. Implicit biases held by school staff can also contribute to who is being disciplined and why. Implicit biases are attitudes or stereotypes that impact an individual's understanding, actions, and decisions in an unconscious

manner (Staats, 2016). The implicit biases of school teachers and staff lead to differences in the evaluation of student behavior based on various identifying factors such as gender and race (Morris & Perry, 2017; Smolkowski et al., 2016). When teachers and other school staff members make decisions about whether to take disciplinary action, they make decisions about whether the student's behavior is deserving of disciplinary action, and if so, what type of disciplinary consequence is warranted. These decisions can be subjective and lead to certain types of students being disciplined at a higher rate or more harshly than other students based on the implicit biases of who is making the decision. It is imperative to provide training and resources to school staff to reduce their biases and increase their understanding of equitable practices in schools. This could reduce disproportionate discipline action outcomes so that all students can benefit from the school environment.

Reducing Discipline in School

Federal laws such as the Every Student Succeeds Act (ESSA) and the Child Care and Development Block Grant (CCDBG) Act of 2014 have provisions related to schools' use of disciplinary actions. These laws are put in place to reduce exclusionary disciplinary actions in schools as well as keep schools held accountable to reduce their disproportionate discipline outcomes. Alternative practices to addressing student behavior in schools have been developed to replace disciplinary actions that remove students from their classrooms. Some of these alternative practices include social and emotional learning (e.g., Gregory & Fergus, 2017), restorative justice practices (e.g., Wong et al., 2011), and School-Wide Positive Behavioral Interventions and Supports (SW-PBIS; e.g., Sugai & Horner, 2009).

Social-emotional learning (SEL) targeting self-awareness, self-management, social awareness, relationship skills, and responsible decision-making can enhance students' ability to

deal with challenges and can decrease schools' use of exclusionary discipline actions (Gregory & Fergus, 2017; Nowicki, 2018). Gregory and Fergus (2017) outlined the potential positive impacts of SEL programming on reducing discipline actions but stated that they currently do not show evidence for reducing disproportionate discipline actions. They provided equity considerations for SEL programming, targeting school staff and students, that could potentially aid in reducing disproportionate discipline action outcomes. Some suggestions include educators examining their own conscious and unconscious beliefs, minimizing colorblindness, adopting a sociocultural, historical orientation, and using responsible decision-making for choices about discipline policies and enforcement. Teachers are also encouraged to incorporate teaching marginalized students to recognize self-management demands as they move between cultures. They concluded that SEL needs to focus not only on the students but also on teachers who interact with students and ultimately implement the disciplinary actions.

Restorative justice is a whole school approach aimed to reduce exclusionary discipline actions by providing students with empathy skills and problem-solving to repair the harm done through relationships and by people. Restorative justice focuses on three principles: relationships and their harm, empowerment of all persons, and collaboration (Song & Swearer, 2016). All three principles of restorative justice need to be practiced in a school system to reduce discipline actions effectively (Song et al., 2020). Restorative justice practices place repairing relationships between people over assigning blame and providing disciplinary actions. The critical questions posed when using restorative practices over disciplinary actions include: who was harmed, what are the needs that gave rise to the event, and how do we make this right. Although a relatively new practice in schools, restorative justice practices are effective in reducing exclusionary

discipline actions and can be a promising method for reducing disproportionate discipline actions (Augustine et al., 2018; Gregory et al., 2016; Wong et al., 2011).

SW-PBIS is a school-wide framework aimed at improving student behavior and reducing school discipline actions through positive and preventative approaches. Key components of SW-PBIS include teaching school behavior expectations, reinforcing those expectations when they are met by students, providing consistent responses to problem behavior, monitoring student behavior in all areas of the school, and using data to make decisions on school needs (Sugai & Horner, 2006). SW-PBIS practices encourage teachers to focus on increasing students' positive behaviors instead of focusing on punishing problem behaviors. SW-PBIS is a socially acceptable framework, implemented widely in schools across the United States, and is an effective method for decreasing schools' use of disciplinary actions and increasing students' prosocial behavior (Bradshaw et al., 2010; Bradshaw et al., 2012; Childs et al., 2015). Classroom practices within SW-PBIS incorporate predictable classroom routines, prompts and active supervision, acknowledgments for expected behavior, and positive classroom expectations posted for students to see. Implementing these practices leads to improved student behavior and fewer behavior disruptions. Although SW-PBIS has years of evidence of effectiveness for decreasing school disciplinary actions, the evidence of support for reducing disproportionality in discipline actions is limited but promising (Better-Bubon et al., 2016; McIntosh et al., 2018ab; Vincent et al., 2011). The SW-PBIS literature is shifting to focusing on reducing disproportionate outcomes in schools (Leverson et al., 2019; McIntosh et al., 2018). Culturally responsive practices are recommended to be incorporated into SW-PBIS practices to reduce disproportionate discipline action outcomes in schools and can be incorporated into the current framework of SW-PBIS (Leverson et al., 2019; McIntosh et al., 2018).

Reducing Disproportionality in Discipline

The U.S. Department of Education Guiding Principles for Improving School Climate and Discipline (2014) include practices such as training school staff to implement school discipline policies fairly and equitably, using data to identify discriminatory discipline actions and reduce its use, implementing school policies that only remove students from the classroom as a last option, getting students back in the classroom as soon as possible, and implementing policies that reduce the use of referrals to law enforcement in the schools. Fallon and colleagues (2012) completed a literature review of effective culturally responsive behavior management to decrease disproportionate discipline action outcomes. They provided various recommendations to be used in the classroom setting as well as for preparing teachers to implement equitable practices. In the classroom setting, effective practices for improving equity include increasing positive interactions with all students, decreasing negative interactions with all students, setting explicit and high expectations for all students, and having a consequence system in place that is planned and delivered consistently to each student. More practices recommended for the classroom include teaching social skills, including students' culture and language, and using effective evidence-based instructional practices. Recommendations for teacher training to increase equitable discipline action outcomes include topics such as using data to evaluate outcomes, engaging families and communities, learning about students' cultures, self-assessing one's own biases and culture and how it impacts their teaching, and understanding that behavior is culturally learned. These practices have all been found to reduce disproportionate outcomes of discipline actions within the literature.

The Technical Assistance Center on PBIS has also published recommendations for enhancing equity in school disciplinary actions (McIntosh et al., 2018). The multicomponent

approach includes collecting, using, and reporting disaggregated discipline data, implementing a behavioral framework that is preventative, multi-tiered, and culturally responsive, and teaching strategies for neutralizing implicit bias in discipline action decisions. Various culturally responsive practices recommended for use within the school and classroom are outlined by Leverson and colleagues (2019). These include incorporating practices that prompt the use of data to hold staff responsible for ensuring equitable implementation and impact, having expectations that reflect the cultural values of students, explicitly teaching expected behaviors, and teaching what is expected at school without devaluing what is expected at home. More practices include emphasizing pro-social skills, educator professional development targeting cultural responsiveness and issues specific to the student population, considering student culture when designing recognition systems, faculty accepting responsibility for sustaining practices, and engaging families and communities of underserved populations. Further suggestions include disaggregating discipline data to monitor equity outcome data, using data to make decisions with a focus on equity, examining the fidelity of practices, and evaluating and reporting equity data annually.

Leverson and colleagues (2019) also recommend evaluating classroom-level vulnerable decision points to help teachers identify situations in which their decisions might be vulnerable to implicit biases. Feedback on their use of disciplinary actions in the classroom can help teachers identify where they are performing equitably or disproportionately. Within the context of a classroom behavior intervention, vulnerable decision points for providing students with rewards or punishment can also be assessed to help teachers evaluate their practices and implement procedures to reduce practices contributing to disproportionality. Lai and colleagues (2013) reported ways to reduce implicit biases. One method involves creating behavioral plans in

the classroom that avoid prejudice with the understanding that teachers need to make a conscious decision about whom they are proving consequences to and making sure consequences are distributed evenly to all students. Classroom behavior management interventions can be incorporated into the SWPBIS framework with culturally responsive practices. They can also produce data for analyzing disproportionate use and implicit biases. Incorporating equitable practices into classroom behavior management interventions is a promising method for reducing disproportionate discipline actions.

The Good Behavior Game

The Good Behavior Game (GBG) is a classroom behavior management intervention that has existed for over half a century (Barrish et al., 1969). This intervention uses an interdependent group contingency to decrease disruptive behavior and increase on-task behavior. The GBG was originally used in the classroom setting with students split into teams, a set of behavioral rules (e.g., "No talking without permission"), and students earning points for their team when they engaged in rule violations. At the end of the game, the students on the team with the fewest points earned a reward. Since the GBG's inception, many variations of the game have been examined across settings, participants, and target behaviors (Joslyn, 2019). The GBG has also been examined and found effective with variations of intervention implementation procedures (Joslyn, 2019).

The GBG has been described as a universal behavioral vaccine due to its effective outcomes across a substantial number of studies incorporating different populations, settings, and target behaviors as well as for its positive long-term outcomes (Embry, 2002). Outcomes of the intervention include effectively decreasing unwanted behaviors and increasing prosocial behaviors. The GBG has also been evaluated for long-term effects and evidence of its prevention

of aggressive behaviors and tobacco use has been found (Tankersley, 1995). Strong follow-up results also indicate long-term impacts on impulsive and disruptive behaviors. Specifically, first-grade boys with aggressive behaviors exhibited significant declines in their 6th-grade year and males were significantly less likely to initiate smoking in their early teens (Kellam et al., 1998). The GBG is the only intervention implemented by individual teachers that has long-term effects (Embry, 2002) and is a highly acceptable intervention when evaluated among teachers and students (Tingstrom et al., 2006). Efficacy has been examined with higher risk populations, young primary school children, across cultures, in non-classroom settings, with adolescents, and in other settings and results show that the GBG is a promising procedure to be used in any setting to increase appropriate behavior and decrease inappropriate behavior (Embry, 2002).

The GBG can be implemented in multiple ways depending on teacher preferences and desired outcomes. Recommendations on implementation guidelines are provided throughout the literature and manuals have been widely developed (e.g., Axelrod et al., 2020; Intervention Central, n.d., McKenna & Flower, 2014; Ohio Department of Education, n.d.). Before implementing the GBG, teachers should decide when to schedule the game and for how long it will be played. The GBG is typically scheduled during an academic period where students are expected to remain on-task and there is a need for a decrease in off-task behavior. When deciding the duration of the game, implementation is typical during a single academic instruction time ranging from 10-minutes to one hour depending on how long the instructional period is. Next, teachers should divide students into two or more teams. It is recommended that teams have equal amounts of students who engage in disruptive behavior on each team. Teachers should then clearly define the behavior rules for the game. Typical rules include staying in seat, talking only when instructed, and participating in instruction. It is important to have operational

definitions for each rule including what behavior does and does not look like for each rule. Students should all be explicitly taught the rules required of the game. Next, rewards that are determined as motivating to students need to be determined for teams that win the GBG. Examples of rewards can include tangible items (e.g., stickers, erasers), edible items (e.g., candy, chips), or intangible items and special privileges (e.g., group leaders, 5 minutes to write on the whiteboard, line leaders). Teachers should also determine the number of points that teams should start with at the beginning of the game. Once all of these factors of the game have been decided, it is important to explain all of the guidelines to students and allow them the time to have all questions answered.

When implementing the GBG, teachers should start by reviewing the rules, duration, teams, criteria for winning, and the reward for winning with the class. The teacher should then announce the start of the game and begin the timer. When a student violates a rule, the teacher should state which student was responsible for the rule violation, which rule was violated, and then remove a point from the student's team. When a student is following a rule, the teacher should state which rule is being followed along with the student who was following the rule and then provide a point to that student's team. When the timer goes off and the game is over, the teacher should announce the conclusion of the game. The teacher will then pull out a random number that decides how many points a team needs to earn the reward. Based on the number pulled and the number of points each team ended the game with, the teacher will announce the team(s) that won the game. The teacher should then provide verbal praise to the winning team(s) and provide those students the reward for winning.

The GBG is a widely used classroom behavior management intervention with long-term positive outcomes that is effective at reducing discipline actions in the classroom and increasing

prosocial behaviors. It is an intervention that is designed in a way that allows for variations and could incorporate recommended culturally responsive practices. The provision and removal of points during the GBG could allow for vulnerable decision points to be examined to assess the teacher's equitable implementation across racially diverse students. No variations of the GBG have examined incorporating practices for equity, and currently, it is not clear if the GBG is being implemented equitably by teachers. It is not known if outcomes on student behavior are equitable across racially diverse students. Evidence-based interventions should be evaluated for equity so that all students are benefiting from the practice.

`

Chapter III: Research Methods

Participants and Settings

Three teachers and their respective students were included in this study. A non-probability convenience sampling (Cooper et al., 2007), in which participants are selected due to availability was used. To be included in the study, teachers reported disruptive behavior in their classrooms during at least one 30-minute block of instructional time. Additionally, the students within their classroom were racially diverse, which was operationalized as having no more than 75% of students identifying as the same race. The United States federal categories to identify race and ethnicity were adopted (i.e., American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian or Other Pacific Islander, White). Participants were recruited from one public elementary school in the Southeastern United States. Data collection took place in the teachers' classrooms during a 30-minute instructional period where the teacher reported the most disruptive behavior and students were expected to exhibit appropriate academic behavior.

Classroom 1 contained 21 second-grade students (e.g., 9 white, 7 Latino, 4 Black, and 1 Bi-racial student) between the ages of 7 and 9. Of those students, 19 had free/reduced lunch, 10 were students with disabilities, 9 were male, and 10 were female. The teacher participant in Classroom 1 was a 23-year-old white female. She was a first-year teacher with a bachelor's degree in elementary education. The math academic period was identified as the instructional time with the most disruptive behavior.

Classroom 2 contained 21 third-grade students (e.g., 12 White, 3 Latino, 5 Black, and 1 Other) between the ages of 8 and 10. Of those students, 18 had free/reduced lunch, 2 were students with disabilities, 11 were male, and 10 were female. The teacher participant in Classroom 2 was a 23-year-old White female. She was a second-year teacher with a bachelor's degree. The language arts academic period was identified as the instructional time with the most disruptive behavior.

Classroom 3 contained 22 third-grade students (e.g., 10 White, 9 Latino, and 3 Black) between the ages of 8 and 10. Of those students, 18 had free/reduced lunch, 2 were students with disabilities, 11 were male and 10 were female. The teacher participant in Classroom 3 was a 23-year-old White female. She was a first-year teacher with a bachelor's degree. The reading academic period was identified as the instructional time with the most disruptive behavior.

Study Variables

The primary dependent variable for this study is the equitable implementation of the GBG. Equity within GBG implementation was examined by analyzing teachers' provision and removal of points within the game. Specifically, inequitable implementation is defined as a greater risk to receive fewer points or have more points removed (TRPC Ratio = 1.20 or higher; see Analysis subsection for additional details) for any non-White racial category within the classroom compared to all other students during GBG implementation. The GBG implementation for this study included an independent group contingency where students all began with five points and could have points removed or provided. During the GBG, if a student engaged in any behaviors that are against the classroom GBG rules, the teacher could remove a point from that student. If a student engaged in behaviors that were aligned with the rules of the GBG, teachers could provide a point to that student. At the end of the game, the teacher pulled a

random number, ranging from three to eight, that indicated how many points a student needed to earn a reward. Before observations took place, the racial composition of the class for the current observation was recorded to account for variance in composition due to absences. Observations were not conducted if the classroom student makeup was not diverse (i.e., 75% or more of one racial identity) due to student absences. Observations of teachers during GBG implementation collected data on the frequency of points provided and removed for each racial category. Data collection took place during the previously identified 30-minute time period, one to four times per week per classroom.

Secondary outcome data were collected to evaluate class-wide student behavior, using an individual fixed observation method (Dart et al., 2016). In this method, using 30s momentary time sampling (Cooper et al., 2007), one student was observed at the end of each 30-second interval in a predetermined order. Each student was observed individually in sequence and once all students were observed in the predetermined order, the observation began again on the first student, continuing in sequence until the 30-min observation was completed. This data indicates the percentage of intervals of disruptive behavior and appropriate behavior for each racial category. Disruptive behavior is defined as engagement in any behavior that does not match the ongoing classroom instruction and includes talking without permission (e.g., talking to self or others, yelling, whistling, or making other noises), being out of their seat without permission (e.g., standing up or walking around the room), non-compliance to teacher demands, and physical disruption to others or property (e.g., hitting, kicking, throwing objects, or destroying items). Appropriate behavior includes engaging in any behavior that matches the ongoing classroom instruction (e.g., the class is writing, the target student is writing); and talking or being out of their seat with teacher permission. These data provide information on the level of

disruptive and appropriate behavior for each racial category in any given observation. Data on student discipline outcomes were also collected including the number of discipline referrals given in the classroom before and after implementing the GBG.

The independent variable for the study is GBG equity-focused teacher training. This equity-focused training includes performance feedback, self-awareness, and self-monitoring (Fallon et al., 2018; Knochel, 2019). Teachers were provided feedback on their use of points during pre-equity-focused training implementation of the GBG. They were provided information on students who received the most and the least number of points as well as had the most and least number of points removed. Teachers were trained to monitor their use of point provision and removal during the implementation of the GBG and to use self-awareness of equitable practices when providing and removing points across racially diverse students.

Data Collection

Data Collection Training

The primary investigator trained data collectors on how to observe and collect data on teacher and student behavior before in-vivo data collection. Observers obtained an interobserver agreement (IOA) of 90% or higher using the data collection tools on a prerecorded GBG implementation video to move into data collection for study purposes.

Data Collection Procedure

Teacher and student behavior were observed and recorded during a 30-minute target academic instruction time. Teacher behavior was recorded using a frequency within 1-minute recording method (see Appendix A). Observers recorded the frequency of points given within each interval and the frequency of points removed within each interval for each racial category. At the end of the observation, data collectors recorded the total frequency of points given and the

total frequency of points removed for each racial category and the entire class. The average number of point provisions and removals per racial category and average for the class as a whole was calculated. Student behavior was recorded using momentary time sampling at the end of a 30-second interval (see Appendix B). At the end of each interval, observers recorded if the student being observed was engaging in disruptive behavior or appropriate behavior. They also indicated the race of the student observed for each interval.

Treatment and Procedural Fidelity

Fidelity measures were used to assess teachers' and researchers' implementation behavior during the study. A treatment fidelity checklist was developed and completed during each observation to assess the teacher's implementation of the GBG (see Appendix C). A procedural fidelity checklist was developed and completed to assess the researcher's implementation of the teacher GBG training before baseline (see Appendix D) and the GBG training focused on equitable practices (see Appendix E). Each checklist contains the critical components of the GBG and the equity training package, respectively. An average score of 100% procedural fidelity was recorded across each phase of the study, for each teacher participant. Table 1 depicts treatment fidelity for each teacher across phases. There was an overall average of 75% treatment fidelity for Classroom 1 across phases, 77% for Classroom 2 across phases, and 77% for Classroom 3 across phases.

Social Validity

Social validity was measured for teachers and students. The Usage Rating Profile-Intervention Revised (URP-IR; Chafouleas et al., 2011) is a 29-item, 6-point likert-type scale that measures acceptability, understanding, home school collaboration, feasibility, system climate, and system support. Subscales of the URP-IR demonstrate acceptable levels of internal

consistency reliability with alpha coefficients ranging from .72 - .95 (Briesch et al., 2013). The subscales are scored by averaging responses across like items. The average score will be used to label each teacher's level of acceptability based on the likert scale label (i.e., strongly disagree/agree, disagree/agree, slightly disagree/agree) that closest matches the average score. Teacher social validity of the GBG equity-focused training will be measured at the end of the intervention phase using the URP-IR. Teachers were also asked to verbally report their acceptability levels of the intervention at the end of the intervention phase and their perception of student acceptability of the intervention.

Interobserver Agreement

Interobserver agreement (IOA) was assessed by an independent observer between 0 - 75% of observations in each phase across each teacher. IOA was calculated for the number of points given and the number of points removed per racial identity in which the number of intervals with agreement was divided by the total number of intervals and multiplied by 100. IOA was calculated for student behavior by dividing the total amount of intervals with agreement by the total amount of intervals and multiplying by 100. If IOA fell below 70%, observers underwent retraining on data collection and behavior observation methods.

Average IOA across study phases and classrooms can be found in Table 2. IOA was collected for Classroom 1 for 0% of baseline sessions and 75% of intervention sessions for teacher behavior and treatment fidelity. For teacher behavior, IOA ranged from 98% to 100%. For treatment fidelity, IOA ranged from 73% to 100%. For student behavior, IOA data were not collected. IOA was collected for Classroom 2 for 0% of baseline and intervention sessions for teacher behavior and treatment fidelity. For student behavior, IOA data were collected for 0% of baseline and intervention sessions. IOA was collected for Classroom 3 for 11% of baseline

sessions and 29% of intervention sessions for teacher behavior, treatment fidelity, and student behavior. For teacher behavior, IOA ranged from 86% to 99%. For treatment fidelity, IOA ranged from 89% to 100%. For student behavior, IOA ranged from 70% to 85%.

Research Design

The research design used in this study was a nonconcurrent multiple baseline across classrooms (Cooper et al., 2007). The intervention was introduced across all classrooms in a nonconcurrent staggered fashion over time. Each phase consists of a minimum of four observations. Phase changes were made upon observation of a minimum of three stable observations of point removal or provision along with teacher fidelity of GBG implementation at 60% or higher for a minimum of one racial category. Phase change decisions were responsive to the presenting data, with baseline consisting of a minimum of three stable data points for a minimum of one racial category and a stagger of phase changes for each replication, as well as based on teacher availability for data collection. For design purposes, phase change decisions were also made to stagger intervention implementation and to obtain a minimum of five observations in each phase. The stability of data involves a pattern of consistent level and variability with little to no trend (Kratochwill et al., 2010). During the intervention phase, a minimum of four 30-minute observations were conducted in each classroom.

Procedures

Teacher Screening

Teachers who were interested in participating in the study were interviewed to obtain an understanding of their personal and teaching background, to understand their current class behavior management strategies, to identify school-wide behavior management systems, to identify their class's racial composition, and to determine if they have current struggles with

student disruptive behavior during instruction time (see Appendix F). Teachers were asked to provide data on student racial information for their class using non-identifiable information in list form of how many students are in each racial category. They were asked to verify the race of each student to ensure their report of student race and school identification of student race are the same. Racial categories included White, Black or African American, Hispanic or Latinx, Asian/Pacific Islander, American Indian/Alaska Native, and Other. The three racial identities with the largest *n* size of students per class were used as racial categories for data collection purposes. A fourth racial identity category of "other" was used to account for all racial identities outside of the largest three within any given classroom. One of the classrooms only had three racial identities identified by the teacher, so only three racial categories were examined in that particular classroom. One observation was conducted in the classroom of interested participants to determine the level of disruptive behavior during the time identified by the teacher. If disruptive behavior was less than 20% of intervals, teachers were excluded from the study and offered other consultation services to address any student behavior issues. Classrooms with a racially diverse student composition (no more than 75% of one racial identity) were included in this study. If the teacher met all inclusion criteria upon screening, they were trained on the implementation of the GBG to address their class's behavior concerns. Four teachers total were screened for inclusion. Three of those teachers qualified for the study and one teacher had less than 20% of intervals with disruptive behavior during the observation and did not qualify to be included in the study.

Teacher Training

The GBG training (see Appendix G) was approximately 30 – 60 minutes and consisted of instruction, modeling, rehearsal, and feedback. Teachers were provided with knowledge and

instructions for the implementation of the GBG and observed the proper implementation of the GBG modeled by the researcher. After these steps, teachers engaged in a role-play scenario of the GBG and received feedback on their implementation and answers to questions they had (see Appendix D). Teachers engaged in this role play and feedback training until they reached 90% or higher fidelity of the key components of the GBG within the training session (see Appendix C). Following the training on the implementation of the GBG, the researcher facilitated the development of the GBG rules with the teacher. The rules were positively worded and defined operationally to be taught to the students.

Baseline

The baseline phase involved an assessment of teacher implementation of the GBG for a minimum of five total observations. If the data indicated disproportionate provision and removal of points within the GBG for at least one racial category (i.e., a TRPC ratio of 1.20 or higher for any non-White racial category) across three observations, teachers moved to the intervention phase. If teachers were engaging in equitable implementation of the GBG for all racial categories (i.e., a TRPC ratio of less than 1.20 for all non-White racial categories), during these initial observations, they were not included in the intervention phase of the study and were provided ongoing consultation to assist in GBG implementation as needed. All three classrooms included in baseline moved into intervention.

Intervention

The intervention phase monitored teachers' use of equitable provision and removal of points during the GBG. The intervention consisted of a brief training incorporating performance feedback, self-monitoring, and self-awareness. This 30-minute training was provided to teachers at their convenience when no students were present. To increase their awareness of equitable

practices, the teacher was first provided an overview of what equitable practices entail within the GBG intervention. They were then provided with data on their use of point provision and removal for each racial category. Next, they were given feedback on how they can make sure that they are providing and removing points equitably across racial categories. They were trained to use a self-monitoring tool to assist them with maintaining self-awareness of their use of point provision and removal across all students (See Appendix H). Teachers were provided data on the student racial identity that received the least and most points as well as had the least and most points removed from the researcher following each intervention phase observation. Researchers reported their implementation of the training using the procedural integrity checklist for equitable GBG training (See Appendix E).

Data Analysis

The Data Accountability Center (2011) released a technical assistance guide for assessing racial disproportionality and outlined a measure called "total removals per child (TRPC)", which can be used to determine the average number of disciplinary removals per child for a specific racial group. This disproportionality measure can be used to account for data where one child can have more than one outcome (i.e., discipline removals). For purposes of this study, TRPC was calculated to determine the average number of point removal and provision per student from each racial identity in each observation as well as across all racial identities combined. A comparison of each racial identity's TRPC of point removal and addition was compared to the class TRPC of point removal and addition. TRPC ratios were calculated to determine the average number of points provision or removals per student from one racial identity compared to that for all other racial identities. TRPC ratios are calculated by dividing the TRPC of one racial group by the TRPC of a comparison group (Data Accountability Center, 2011). The comparison group

was all other students, where one racial group was compared to all students not in that racial group. TRPC ratios of 1.0 indicate no difference between the racial identity and the comparison group. A TRPC ratio of above 1.0 indicates a greater average number of point provision or removal per student from the given racial identity. A TRPC ratio of below 1.0 is indicative of a lower average number of point provision or removal per student from the given racial identity. TRPC ratios can never be below 0.00 (Data Accountability Center, 2011). A TRPC ratio of 0.00-0.25 was considered a high level of underrepresentation for point provision. Underrepresentation was not considered for point removals. A TRPC ratio of 0.00-0.25 was considered a low level of disproportionality for point removal. A TRPC ratio of 0.26- 1.20 was considered a low level of disproportionality for point provision. A TRPC ratio of 1.21- 1.99 was considered a moderate level of overrepresentation/disproportionality for point provision and removal. A TRPC ratio of 2.00 or higher was considered a high level of overrepresentation/disproportionality for point provision and removal. For observations when the TRPC for the comparison group is zero, the TRPC ratio will not be calculated. TRPC and TRPC ratios were calculated to assess disproportionality following each observation and phase.

Visual analysis of these data involved evaluating the stability of baseline data. More specifically, baseline data points were relatively stable for a minimum of one racial identity for either point provision or removal before the intervention was introduced. An evaluation of within phase effects examined level, trend, and variability. An evaluation of between phase effects occurred by evaluating the immediacy of effect, the overlap of data between baseline and treatment phases, and the consistency of the data patterns across similar phases. Finally, an evaluation of the effects across all participants occurred.

Treatment effects were evaluated through the within phase and between phase analyses in the following ways. If trend, level, and variability were consistent across similar phases and multiple participants, strong treatment effects were determined. Also, if the immediacy of the effect was large, the overlap of data between baseline and treatment was low, the consistency of data across similar phases was high, and these effects occurred across all participants, strong treatment effects were determined.

For the secondary outcome analysis, the percentage of intervals with disruptive and appropriate behavior for each racial identity was visually analyzed for changes in level and trend across phases. The number of discipline referrals given pre-study and during-study was reported.

Chapter IV: Results

The results of this study will be described in the following sequence. First, data on points provided will be reviewed for each classroom separately and then across all classrooms. Following the results for points provided, points removed will be presented for each classroom separately and then across all classrooms. For both the points provided and points removed within each classroom, TRPC data will be presented first, followed by the TRPC ratio data. Finally, the student behavior data and social validity data will be reviewed.

Points Provided

Classroom 1

TRPC. Table 3 depicts the TRPC data for points provided across classrooms and phases. TRPC data reported below provides information on the average number of point provisions per student from a specific racial identity. For baseline in Classroom 1, the average number of point provisions per White student was 0.77 (range = 0.14-1.40). The average number of point provisions per Black student was 1.18 (range = 0.25-2.00). The average number of point provisions per Latino student was 0.90 (range = 0.43-1.29). The average number of point provisions per Other student was 0.90 (range = 0.00-2.00). The average number of point provisions for the whole class was 0.81 (range = 0.33-1.37). For intervention in Classroom 1, the average number of point provisions per White student was 1.03 (range = 0.50-1.43). The average number of point provisions per Black student was 0.96 (range = 0.50-1.33). The average number of point provisions per Latino student was 0.92 (range = 0.50-1.17). The average number of

point provisions per Other student was 1.00 (range = 0.47-1.33). The average number of point provisions for the whole class was 0.97 (range = 0.47-1.33).

In Classroom 1 for baseline, Black students were provided the highest average points and White students were provided the lowest average points, with a 0.41 difference. On average, Black, Latino, and Other students were provided points at averages above the class mean and White students were provided points at an average below the class mean. In Classroom 1 for intervention, White students were provided the highest average points and Latino students were provided the lowest average points, with a 0.11 difference. On average, White, Black, and Other students were provided points at averages above the class mean and Latino students were provided points at an average below the class mean.

Figure 1 depicts the average number of points provided across racial identities and class average for all three classrooms. Classroom 1 is depicted in the top panel. During baseline, point provision was mostly undifferentiated across racial identities, however across the last two observations, Black students had higher point provisions and during the last observation in baseline, Other students were also higher. Overall, baseline showed an increasing trend across all races, however, a decreasing trend was observed for White and Latino students while black and other students continued on an increasing trend. Differentiation was observed during the last two baseline observations for at least one racial identity before moving into intervention. From the last two observations in baseline to the first two in intervention, there was a difference in the range of point provisions between racial identities moving closer to similar average point provisions, however, an immediacy of effect was not determined. In intervention, point provision was more undifferentiated across racial identities compared to baseline. Overall, there was no level change from baseline to intervention for all racial identities except a level change from the

last two baseline points for Black students compared to intervention. Overall, there was low variability for all racial identities across phases.

TRPC Ratio. Table 4 depicts the TRPC ratios for points provided across phases and classrooms. TRPC ratios reported below provide information on the average number of point provisions per student from a specific racial identity compared to that for all other students. For baseline in Classroom 1, the TRPC ratios of point provisions per White student ranged from a low level of 0.31 to a high level of 2.00 with 40% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 0.96 across observations. TRPC ratios of point provisions per Black student ranged from a low level of 0.32 to a high level of 2.50 with 40% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 1.25 across observations. TRPC ratios of point provisions per Latino student ranged from a low level of 0.91 to a moderate level of 1.57 with 60% of observations at a moderate level of disproportionality and an average TRPC ratio of 1.28 across observations. TRPC ratios of point provisions per Other student ranged from a high level of underrepresentation (0.00) to a high level of overrepresentation (4.44) with 40% of observations at a high level of disproportionality and an average TRPC ratio of 1.34 across observations.

For intervention in Classroom 1, the TRPC ratios of point provisions per White student ranged from a low level of 1.07 to a low level of 1.14 with 0% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 1.11 across observations. TRPC ratios of point provisions per Black student ranged from a low level of 0.75 to a moderate level of 1.24 with 25% of observations at a moderate level of disproportionality and an average TRPC ratio of 1.00 across observations. TRPC ratios of point provisions per Latino student ranged from a low level of 0.82 to a low level of 1.10 with 0% of observations at a moderate to high level of

disproportionality and an average TRPC ratio of 0.96 across observations. TRPC ratios of point provisions per Other student ranged from a high level of underrepresentation (0.00) to a moderate level of overrepresentation (1.55) with 50% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 0.88 across observations.

TRPC ratios in Classroom 1 during baseline indicated that Black, Latino, and Other students had moderate levels of overrepresentation for point provisions when compared to all other students. White students had low levels of disproportionality for point provisions compared to all other students. During intervention, all racial identities had low levels of disproportionality, indicating no difference between any racial identity to the comparison group. Table 5 shows the percent of observations with low, moderate, and high disproportionate point provision for Classroom 1. From baseline to intervention, observations with low levels of disproportionality increased from 55% to 81%. Observations with moderate levels of disproportionality decreased from 25% to 13% and observations with high levels of disproportionality decreased from 20% to 6% of observations.

Figure 2 depicts TRPC ratios for points provided across racial identities for all three classrooms. Classroom 1 is depicted in the top panel. In baseline, there was differentiation of TRPC ratios across racial identities in multiple observations with no single racial identity consistently differentiated. Ratios were variable across all racial identities during baseline. There was no clear pattern of one racial identity receiving more points in comparison to other races observed. However, disproportionate point provisions were observed across any given observation, but for no specific racial identity. For example, the TRPC ratio for observation 1 was 2.50 for Black students, however, it was 4.44 for Other students during the last baseline observation. Baseline observations indicated disproportionate provision of points, but not for one

racial identity over any other. An immediacy of effect was observed from baseline to intervention with TRPC ratios moving from moderate to low disproportionate levels overall. In intervention there was no clear pattern of one racial identity compared to another however, unlike baseline, there was little to no disproportionality across racial identities. For White students, level remained low from baseline to intervention. For Black, Latino, and Other students, level moved from moderate disproportionality to low disproportionality from baseline to intervention. For all racial identities, variability decreased from baseline to intervention moving closer to low levels of disproportionality.

Classroom 2

TRPC. Table 3 depicts the TRPC data for points provided across classrooms and phases. For baseline in Classroom 2, the average number of point provisions per White student was 1.27 (range = 0.00-3.17). The average number of point provisions per Black student was 0.97 (range = 0.25-2.5). The average number of point provisions per Latino student was 0.70 (range = 0.00-3.00). The average number of point provisions per Other student was 1.33 (range = 0.00-4.00). The average number of point provisions for the whole class was 1.11 (range = 0.06-2.85). For intervention in Classroom 2, the average number of point provisions per White student was 1.41 (range = 0.38-2.25). The average number of point provisions per Black student was 1.47 (range = 0.50-2.00). The average number of point provisions per Latino student was 1.38 (range = 0.00-2.00). The average number of point provisions per Other student was 1.25 (range = 0.00-2.00). The average number of point provisions for the whole class was 1.42 (range = 0.29-1.88).

In Classroom 2 for baseline, Other students were provided the highest average points and Latino students were provided the lowest average points, with a 0.63 difference. On average, White and Other students were provided points at averages above the class mean and Black and

Latino students were provided points at an average below the class mean. In Classroom 2 for intervention, Black students were provided the highest average points and Other students were provided the lowest average points, with a 0.22 difference. On average, Black students were provided points at averages above the class mean and White, Latino, and Other students were provided points at an average below the class mean.

Figure 1 depicts the average number of points provided across racial identities and class average for all three classrooms. Classroom 2 is depicted in the middle panel. During baseline, point provision was undifferentiated across racial identities. Overall, baseline showed a decreasing trend across all races, with Black and Latino student's point provision stabilizing, White student's point provision increasing during the final baseline observation, and Other student's point provision remaining variable. From the last observations in baseline to the first observation in intervention, there was a difference in the range of point provisions between racial identities moving closer to similar average point provisions, however, an immediacy of effect was not determined. In intervention, point provision was more undifferentiated across racial identities compared to baseline. No trends were observed in intervention. Variability was observed across all racial identities in intervention. Overall, there was no clear level change from baseline to intervention for all racial identities.

TRPC Ratio. Table 4 depicts the TRPC ratios for points provided across phases and classrooms. TRPC ratios reported below provide information on the average number of point provisions per student from a specific racial identity compared to that for all other students. For baseline in Classroom 2, the TRPC ratios of point provisions per White student ranged from a high level of underrepresentation (0.00) to a high level of overrepresentation (4.00) with 83% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 1.68

across observations. TRPC ratios of point provisions per Black student ranged from a low level of 0.43 to a moderate level of 1.35 with 20% of observations at a moderate level of disproportionality and an average TRPC ratio of 0.82 across observations. TRPC ratios of point provisions per Latino student ranged from a high level of underrepresentation (0.00) to a low level of 1.06 with 67% of observations at a high level of disproportionality and an average TRPC ratio of 0.35 across observations. TRPC ratios of point provisions per Other student ranged from a high level of underrepresentation (0.00) to a high level of overrepresentation (2.71) with 83% of observations at a high level of disproportionality and an average TRPC ratio of 1.00 across observations.

For intervention in Classroom 2, the TRPC ratios of point provisions per White student ranged from a low level of 0.62 to a high level of 2.25 with 25% of observations at a high level of disproportionality and an average TRPC ratio of 1.28 across observations. TRPC ratios of point provisions per Black student ranged from a low level of 0.86 to a high level of 2.00 with 25% of observations at a high level of disproportionality and an average TRPC ratio of 1.27 across observations. TRPC ratios of point provisions per Latino student ranged from a high level of underrepresentation of 0.00 to a moderate level of 1.73 with 50% of observations at a moderate and high level of disproportionality and an average TRPC ratio of 0.85 across observations. TRPC ratios of point provisions per Other student ranged from a high level of underrepresentation of 0.00 to a low level of 1.07 with 25% of observations at a high level of disproportionality and an average TRPC ratio of 0.68 across observations.

TRPC ratios in Classroom 2 during baseline indicated that Black, Latino, and Other racial identities had low levels of disproportionality for point provisions compared to all other students. White students had moderate levels of overrepresentation for point provision compared to all

other students. During intervention, Latino and Other students had low levels of disproportionality, indicating no difference between those racial identities to the comparison group. White and Black students had moderate levels of overrepresentation for point provision compared to all other students. Table 5 shows the percent of observations with low, moderate, and high disproportionate point provision for Classroom 2. From baseline to intervention, observations with low levels of disproportionality increased from 35% to 69%. Observations with moderate levels of disproportionality decreased from 13% to 6% and observations with high levels of disproportionality decreased from 52% to 25% of observations.

Figure 2 depicts TRPC ratios for points provided across racial identities for all three classrooms. Classroom 2 is depicted in the middle panel. In baseline, there was differentiation of TRPC ratios across racial identities in multiple observations with no single racial identity consistently differentiated. Ratios were variable across all racial identities during baseline, however, Latino students remained stable in the final three baseline observations. There was no clear pattern of one racial identity receiving more points in comparison to other races observed. However, disproportionate point provisions were observed across any given observation, but for no specific racial identity. For example, the TRPC ratio for observation 4 was 2.71 for Other students, however, it was 4.00 for White students during the last baseline observation. Baseline observations indicated disproportionate provision of points, but not for one racial identity over any other. An immediacy of effect was observed from baseline to intervention with TRPC ratios moving from variable levels of disproportionality to less variable and less disproportionate. In intervention there was no clear pattern of one racial identity compared to another however, compared to baseline, there were fewer disproportionate observations across racial identities. For White students, data moved from variable levels of disproportionality to a more stable level from

baseline to intervention with only one disproportionate observation during intervention. For Black students, level and variability were similar from baseline to intervention. For Latino students, level remained similar from baseline to intervention. For Other students, variability decreased from baseline to intervention with only one disproportionate observation during intervention. For all racial identities, variability decreased from baseline to intervention moving closer to low levels of disproportionality.

Classroom 3

TRPC. Table 3 depicts the TRPC data for points provided across classrooms and phases. For baseline in Classroom 3, the average number of point provisions per White student was 1.12 (range = 0.63-2.00). The average number of point provisions per Black student was 1.85 (range = 1.00-4.00). The average number of point provisions per Latino student was 1.32 (range = 0.33-2.00). The average number of point provisions for the whole class was 1.26 (range = 0.57-1.89). For intervention in Classroom 3, the average number of point provisions per White student was 0.89 (range = 0.44-1.40). The average number of point provisions per Black student was 1.17 (range = 0.50-2.00). The average number of point provisions per Latino student was 1.11 (range = 0.56-1.63). The average number of point provisions for the whole class was 1.00 (range = 0.60-1.24).

In Classroom 3 for baseline, Black students were provided the highest average points and White students were provided the lowest average points, with a 0.73 difference. On average, Black and Latino students were provided points at averages above the class mean and White students were provided points at an average below the class mean. In Classroom 3 for intervention, Black students were provided the highest average points and White students were provided the lowest average points, with a 0.28 difference. On average, Black and Latino

students were provided points at averages above the class mean and White students were provided points at an average below the class mean.

Figure 1 depicts the average number of points provided across racial identities and class average for all three classrooms. Classroom 3 is depicted in the bottom panel. During baseline, point provision was mostly undifferentiated across racial identities, however across the last four observations, Black students had higher point provisions. Baseline showed a decreasing trend across White and Latino students and no trend was observed for Black students. No trends were observed in intervention. From the last observation in baseline to the first intervention, there was a difference in the range of point provisions between racial identities moving closer to similar average point provisions, however, an immediacy of effect was not determined. In intervention, point provision was more undifferentiated across racial identities compared to baseline. Overall, there was no level change from baseline to intervention for all racial identities. Variability for all racial identities across phases was observed.

TRPC Ratio. Table 4 depicts the TRPC ratios for points provided across phases and classrooms. TRPC ratios reported below provide information on the average number of point provisions per student from a specific racial identity compared to that for all other students. For baseline in Classroom 3, the TRPC ratios of point provisions per White student ranged from a low level of 0.47 to a moderate level of 1.33 with 11% of observations at a moderate level of disproportionality and an average TRPC ratio of 0.84 across observations. TRPC ratios of point provisions per Black student ranged from a low level of 0.68 to a high level of 5.80 with 55% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 1.86 across observations. TRPC ratios of point provisions per Latino student ranged from a low level

of 0.944 to a moderate level of 1.60 with 22% of observations at a moderate level of disproportionality and an average TRPC ratio of 1.03 across observations.

For intervention in Classroom 3, the TRPC ratios of point provisions per White student ranged from a low level of 0.49 to a moderate level of 1.22 with 14% of observations at a moderate level of disproportionality and an average TRPC ratio of 0.85 across observations. TRPC ratios of point provisions per Black student ranged from a low level of 0.67 to a high level of 2.15 with 28% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 1.09 across observations. TRPC ratios of point provisions per Latino student ranged from a low level of 0.82 to a moderate level of 1.63 with 28% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 1.21 across observations.

TRPC ratios in Classroom 3 during baseline indicated that Black students had moderate levels of overrepresentation for point provisions when compared to all other students. White and Latino students had low levels of disproportionality for point provisions compared to all other students. During intervention, White and Black students had low levels of disproportionality, indicating no difference between those racial identities to the comparison group. Latino students had moderate levels of overrepresentation for point provision compared to all other students. Table 5 shows the percent of observations with low, moderate, and high disproportionate point provision for Classroom 3. From baseline to intervention, observations with low levels of disproportionality increased from 70% to 76%. Observations with moderate levels of disproportionality decreased from 19% to 14% and observations with high levels of disproportionality decreased from 11% to 10% of observations.

Figure 2 depicts TRPC ratios for points provided across racial identities for all three classrooms. Classroom 3 is depicted in the bottom panel. In baseline, there was differentiation of TRPC ratios across racial identities in multiple observations with no single racial identity consistently differentiated. TRPC ratios were variable for Black students during baseline. There was no clear pattern of one racial identity receiving more points in comparison to other racial identities. However, disproportionate point provisions were observed across the majority of observations for Black students. For White students, level remained low from baseline to intervention. For Black students, level moved from moderate disproportionality to low disproportionality from baseline to intervention. For Latino students, level moved from low disproportionality to moderate disproportionality from baseline to intervention. For Black students, variability decreased from baseline to intervention moving closer to low levels of disproportionality. White and Latino students remained similar in variability from baseline to intervention. An immediacy of effect was not observed.

All Classrooms

A nonconcurrent multiple baseline across classrooms design was utilized to assess intervention effects across multiple classrooms. Due to IOA not being collected for a minimum of 20% for each phase, the design does not meet evidence standards, therefore, across classroom data should be interpreted with caution.

TRPC. Figure 1 depicts the average number of points provided across racial identities and class average for all three classrooms. Trend, level, and variability were not consistent across similar phases and multiple participants; therefore, treatment effects were not demonstrated. Also, the immediacy of the effect was small, the overlap of data between baseline and treatment was high, and the consistency of data across similar phases was low across all participants,

therefore treatment effects were not demonstrated. Across all classrooms overall, point provision across racial identities becomes more undifferentiated from baseline to intervention.

TRPC Ratio. Figure 2 depicts TRPC ratios for points provided across racial identities for all three classrooms. Levels of variability decreased across all racial identities from baseline to intervention for Classrooms 1 and 2 while the level of variability only decreased for Black students from baseline to intervention in Classroom 3, therefore a moderate treatment effect could be determined. The immediacy of the effect was demonstrated for Classrooms 1 and 2, but not in Classroom 3. The overlap of data between baseline and treatment was moderate, and the consistency of data across similar phases was low across all participants, therefore treatment effects were not demonstrated. Across all classrooms overall, TRPC ratios across racial identities become more undifferentiated from baseline to intervention.

Points Removed

Classroom 1

TRPC. Table 6 depicts the TRPC data for points removed across classrooms and phases. TRPC data reported below provides information on the average number of points removed per student from a specific racial identity. During baseline in Classroom 1, the average number of point removals per White student was 0.02 (range = 0.00-0.10). The average number of point removals per Black student was 0.25 (range = 0.00-0.75). The average number of point removals per Latino student was 0.12 (range = 0.00-0.29). The average number of point removals per Other student was 0.10 (range = 0.00-0.50). The average number of point removals for the whole class was 0.11 (range = 0.00-0.26). For intervention in Classroom 1, the average number of point removals per White student was .04 (range = 0.00-0.14). The average number of point removals per Black student was 0.40 (range = 0.25-0.75). The average number of point removals per

Latino student was 0.16 (range = 0.00-0.33). The average number of point removals per Other student was 0.00 (range = 0.00-0.00). The average number of point removals for the whole class was 0.15 (range = 0.06-0.25).

In Classroom 1 for baseline, Black students had the highest average number of points removed and White students had the lowest average number of points removed, with a 0.23 difference. On average, Black and Latino students had points removed at averages above the class mean and White and Other students had points removed at an average below the class mean. In Classroom 1 for intervention, Black students had the highest average number of points removed and Other students had the lowest average number of points removed, with a 0.40 difference. On average, Black and Latino students had points removed at averages above the class mean and White and Other students had points removed at an average below the class mean.

Figure 3 depicts the average number of points removed across racial identities and the class average for all three classrooms. Classroom 1 is depicted in the top panel. During baseline, point removal was somewhat differentiated for Black students compared to all other racial identities. Point removal for Latino students was somewhat higher compared to other racial identities during three baseline observations. Point removal for Other and White students only occurred in one observation each during baseline. Point removal for Latino and black students was variable in baseline. An immediacy of effect was not demonstrated. During intervention, there was an increasing trend observed across point removals for Black and Latino students. No points were removed for Other and White students, except for one observation for white students. Additionally, differentiated point removal was observed for Black and Latino students when compared to Other and White students. More point removals were observed for Black

students compared to all other races and more point removals were observed for Latino students compared to White and Other students. Overall, all students had no level change from baseline to intervention. Black and Other students had a slight decrease in variability from baseline to intervention. Latino and White students had similar variability across phases.

TRPC Ratio. Table 7 depicts the TRPC ratios for points removed across phases and classrooms. TRPC ratios described below provide information on the average number of point removals per student from a specific racial identity compared to that for all other students. For baseline in Classroom 1, the TRPC ratios of point removals per White student ranged from a low level of 0.00 to a low level of 1.10 with 0% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 0.28 across observations. TRPC ratios of point removals per Black student ranged from a low level of 0.00 to a high level of 5.63 with 50% of observations at a high level of disproportionality and an average TRPC ratio of 2.28 across observations. TRPC ratios of point removals per Latino student ranged from a low level of 0.00 to a high level of 2.50 with 33% of observations at a high level of disproportionality and an average TRPC ratio of 1.21 across observations. TRPC ratios of point removals per Other student ranged from a low level of 0.00 to a high level of 3.50 with 25% of observations at a high level of disproportionality and an average TRPC ratio of 0.88 across observations.

For intervention in Classroom 1, the TRPC ratios of point removals per White student ranged from a low level of 0.00 to a low level of 0.79 with 0% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 0.20 across observations. TRPC ratios of point removals per Black student ranged from a moderate level of 1.75 to a high level of 9.00 with 100% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 5.14 across observations. TRPC ratios of point removals per Latino student ranged

from a low level of 0.00 to a moderate level of 1.44 with 0% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 0.97 across observations. TRPC ratios of point removals per Other student ranged from a low level of 0.00 to a low level of 0.00 with 0% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 0.00 across observations.

TRPC ratios in Classroom 1 during baseline indicated that Black students had high levels of overrepresentation for point removals when compared to all other students. Latino students had moderate levels of overrepresentation for point removals compared to all other students. White and Other students had low levels of disproportionality for point removals compared to all other students. During intervention, White, Latino, and Other students had low levels of disproportionality, indicating no difference between any racial identity to the comparison group. Black students had high levels of disproportionality for point removals. Table 8 shows the percent of observations with low, moderate, and high disproportionate point removal for Classroom 1. From baseline to intervention, observations with low levels of disproportionality decreased from 73% to 67%. Observations with moderate levels of disproportionality increased from 0% to 20% and observations with high levels of disproportionality decreased from 27% to 13% of observations.

Figure 4 depicts TRPC ratios for points removed across racial identities for all three classrooms. Classroom 1 is depicted in the top panel. For White and Other students, a low level was observed from baseline to intervention. For Black students, a level increase was observed from baseline to intervention in the high levels. For Latino students, level decreased from a moderate level to a low level from baseline to intervention. For Latino and Other students, variability decreased from baseline to intervention. For Black students, the variability of data

increased from baseline to intervention. For White students, variability between phases remained similar. An immediacy of effect was not demonstrated.

Classroom 2

TRPC. Table 6 depicts the TRPC data for points removed across classrooms and phases. For baseline in Classroom 2, the average number of point removals per White student was 0.83 (range = 0.17-1.30). The average number of point removals per Black student was 0.48 (range = 0.00-1.00). The average number of point removals per Latino student was 1.06 (range = 0.67-1.67). The average number of point removals per Other student was 0.17 (range = 0.00-1.00). The average number of point removals for the whole class was 0.75 (range = 0.25-1.13). For intervention in Classroom 2, the average number of point removals per White student was 0.19 (range = 0.00-0.36). The average number of point removals per Black student was 0.27 (range = X0.00-0.50). The average number of point removals per Latino student was 0.88 (range = 0.00-2.50). The average number of point removals per Other student was 0.25 (range = 0.00-1.00). The average number of point removals for the whole class 0.30 (range = 0.06-0.65).

In Classroom 2 for baseline, Latino students had the highest average number of points removed and Other students had the lowest average number of points removed, with a 0.89 difference. On average, White and Latino students had points removed at an average above the class mean and Black and Other students had points removed at averages below the class mean. In Classroom 2 for intervention, Latino students had the highest average number of points removed and White students had the lowest average number of points removed, with a 0.69 difference. On average, Latino students had points removed at averages above the class mean, and White, Black, and Other students had points removed at an average below the class mean.

Figure 3 depicts the average number of points removed across racial identities and the class average for all three classrooms. Classroom 2 is depicted in the middle panel. During baseline, point removal was differentiated for White students during two observations and Latino students during three observations compared to all other racial identities. Point removal for Other students only occurred in one observation each during baseline. No consistent trends were observed during baseline. Point removal for White and Black students was variable in baseline. An immediacy of effect was not demonstrated. During intervention, there was a decreasing trend observed across point removals for all students. Differentiated point removal was observed for Latino students initially during intervention and then no differentiation across racial identities was observed for the last two observations. White students had a slight level change from baseline to intervention. Black students had a slight decrease in variability from baseline to intervention. All other students had similar variability across phases.

TRPC Ratio. Table 7 depicts the TRPC ratios for points removed across phases and classrooms. TRPC ratios described below provide information on the average number of point removals per student from a specific racial identity compared to that for all other students. For baseline in Classroom 2, the TRPC ratios of point removals per White student ranged from a low level of 0.44 to a moderate level of 1.78 with 67% of observations at a moderate level of disproportionality and an average TRPC ratio of 1.28 across observations. TRPC ratios of point removals per Black student ranged from a low level of 0.00 to a low level of 0.87 with 0% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 0.45 across observations. TRPC ratios of point removals per Latino student ranged from a low level of 0.87 to a high level of 8.50 with 50% of observations at a high level of disproportionality and an average TRPC ratio of 2.71 across observations. TRPC ratios of point removals per Other

student ranged from a low level of 0.00 to a low level of 1.15 with 0% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 0.19 across observations.

For intervention in Classroom 2, the TRPC ratios of point removals per White student ranged from a low level of 0.00 to a low level of 0.31 with 0% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 0.17 across observations. TRPC ratios of point removals per Black student ranged from a low level of 0.00 to a moderate level of 1.50 with 33% of observations at a moderate level of disproportionality and an average TRPC ratio of 0.66 across observations. TRPC ratios of point removals per Latino student ranged from a low level of 0.00 to a high level of 6.25 with 50% of observations at a high level of disproportionality and an average TRPC ratio of 2.94 across observations. TRPC ratios of point removals per Other student ranged from a low level of 0.00 to a moderate level of 1.60 with 25% of observations at a moderate level of disproportionality and an average TRPC ratio of 0.40 across observations.

TRPC ratios in Classroom 2 during baseline indicated that Latino students had high levels of overrepresentation for point removals when compared to all other students. White students had moderate levels of overrepresentation for point removals compared to all other students. Black and Other students had low levels of disproportionality for point removals compared to all other students. During intervention, White, Black, and Other students had low levels of disproportionality, indicating no difference from the comparison group. Latino students had high levels of disproportionality for point removals. Table 8 shows the percent of observations with low, moderate, and high disproportionate point removal for Classroom 2. From baseline to intervention, observations with low levels of disproportionality increased from

71% to 72%. Observations with moderate levels of disproportionality decreased from 17% to 14%. Observations with high levels of disproportionality increased from 12% to 14% of observations.

Figure 4 depicts TRPC ratios for points removed across racial identities for all three classrooms. Classroom 2 is depicted in the middle panel. For White students, level decreased from a moderate level to a low level from baseline to intervention. For Black and Other students, level remained similar from baseline to intervention. For all students, variability remained similar from baseline to intervention. A decreasing trend was observed for Latino students during baseline and intervention. An immediacy of effect was not demonstrated.

Classroom 3

TRPC. Table 6 depicts the TRPC data for points removed across classrooms and phases. For baseline in Classroom 3, the average number of point removals per White student was 0.56 (range = 0.11-1.22). The average number of point removals per Black student was 0.22 (range = 0.00-2.00). The average number of point removals per Latino student was 0.18 (range = 0.00-0.56). The average number of point removals for the whole class was 0.35 (range = 0.14-0.65). For intervention in Classroom 3, the average number of point removals per White student was 0.41 (range = 0.11-0.67). The average number of point removals per Black student was 0.47 (range = 0.00-1.00). The average number of point removals per Latino student was 0.18 (range = 0.00-0.38). The average number of point removals for the whole class was 0.31 (range = 0.06-0.52).

In Classroom 3 for baseline, White students had the highest average number of points removed and Latino students had the lowest average number of points removed, with a 0.74 difference. On average, White students had points removed at an average above the class mean

and Black and Latino students had points removed at averages below the class mean. In Classroom 3 for intervention, Black students had the highest average number of points removed and Latino students had the lowest average number of points removed, with a 0.29 difference. On average, White and Black students had points removed at averages above the class mean and Latino students had points removed at an average below the class mean.

Figure 3 depicts the average number of points removed across racial identities and the class average for all three classrooms. Classroom 3 is depicted in the bottom panel. During baseline, point removal was somewhat undifferentiated across all racial identities. Point removal for White students was somewhat higher compared to other racial identities during five baseline observations. Point removal for Black students only occurred in one observation during baseline. Point removal for Latino and White students was variable in baseline. No trend was observed across phases. An immediacy of effect was not demonstrated. Additionally, undifferentiated point removal was observed during intervention. White and Latino students had no level change from baseline to intervention. Black students had an increase in level from baseline to intervention. White students had a slight decrease in variability from baseline to intervention. Latino and Black students had similar variability across phases.

TRPC Ratio. Table 7 depicts the TRPC ratios for points removed across phases and classrooms. TRPC ratios described below provide information on the average number of point removals per student from a specific racial identity compared to that for all other students. For baseline in Classroom 3, the TRPC ratios of point removals per White student ranged from a low level of 0.67 to a high level of 12.22 with 75% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 4.06 across observations. TRPC ratios of point removals per Black student ranged from a low level of 0.00 to a high level of 3.56 with 11% of

observations at a high level of disproportionality and an average TRPC ratio of 0.40 across observations. TRPC ratios of point removals per Latino student ranged from a low level of 0.00 to a high level of 2.67 with 22% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 0.73 across observations.

For intervention in Classroom 3, the TRPC ratios of point removals per White student ranged from a low level of 0.31 to a high level of 6.11 with 83.5% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 2.12 across observations. TRPC ratios of point removals per Black student ranged from a low level of 0.00 to a high level of 2.25 with 43% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 1.78 across observations. TRPC ratios of point removals per Latino student ranged from a low level of 0.00 to a low level of 0.81 with 0% of observations at a moderate to high level of disproportionality and an average TRPC ratio of 0.43 across observations.

TRPC ratios in Classroom 3 during baseline indicated that White students had high levels of overrepresentation for point removals when compared to all other students. Black and Latino students had low levels of disproportionality for point removals compared to all other students. During intervention Latino students had low levels of disproportionality, indicating no difference between any racial identity to the comparison group. White students had high levels of disproportionality for point removals. Black students had moderate levels of disproportionality for point removals. Table 8 shows the percent of observations with low, moderate, and high disproportionate point removal for Classroom 3. From baseline to intervention, observations with low levels of disproportionality decreased from 65% to 60%. Observations with moderate levels of disproportionality increased from 8% to 25% and observations with high levels of disproportionality decreased from 27% to 15% of observations.

Figure 4 depicts TRPC ratios for points removed across racial identities for all three classrooms. Classroom 3 is depicted in the bottom panel. For White students, a high level of disproportionality was observed from baseline to intervention. For Black students, a level increase was observed from baseline to intervention from low levels of disproportionality to higher levels. For Latino students, level remained low from baseline to intervention. For White students, variability slightly decreased from baseline to intervention and variability remained similar for Black and Latino students. An immediacy of effect was not demonstrated.

All Classrooms

A nonconcurrent multiple baseline across classrooms design was utilized to assess intervention effects across multiple classrooms. Due to IOA not being collected for a minimum of 20% for each phase, the design does not meet evidence standards, therefore, across classroom data should be interpreted with caution.

TRPC. Figure 3 depicts the average number of points removed across racial identities and the class average for all three classrooms. Trend, level, and variability were not consistent across similar phases and across multiple participants, therefore treatment effects were not demonstrated. Also, the immediacy of the effect was small, the overlap of data between baseline and intervention was high, and the consistency of data across similar phases was low across all participants, therefore treatment effects were not demonstrated. Across all classrooms overall, point removal across racial identities were similar from baseline to intervention.

TRPC Ratio. Figure 4 depicts TRPC ratios for points removed across racial identities for all three classrooms. Trend, level, and variability were not consistent across similar phases and across multiple participants, therefore treatment effects were not demonstrated. Also, the immediacy of the effect was small, the overlap of data between baseline and intervention was

high, and the consistency of data across similar phases was low across all participants, therefore treatment effects were not demonstrated. Across all classrooms overall, TRPC ratios across racial identities was similar from baseline to intervention.

Student Behavior

Classroom 1

Consent to collect student behavior data were collected for 11 out of the 21 (i.e., 52%) students in Classroom 1. Of those students, there were consents returned for 7 White, 2 Latino, 1 Black, and 1 Other student(s). Due to this classroom's setup during observations (i.e., rotating centers) making it difficult for observers to collect data and information on students with returned consents unknown initially by the data collectors, only three student behavior observations were collected total during baseline. Due to the limited number of observations and the limited number of students included, these data were deemed to be inadequate for the purpose of this study. The data that were collected reflected nondifferentiated disruptive and appropriate behavior across racial identities. Before beginning the GBG, teacher 1 reported providing four total office discipline referrals during the months of August through February. She reported providing zero office discipline referrals upon beginning the GBG in the months of March and April.

Classroom 2

Consent to collect student behavior data were collected for 4 out of the 21 (i.e., 19%) students in Classroom 2. Of those students, there were consents returned for 3 White, 1 Black, 0 Latino, and 0 Other student(s). Student behavior data were collected for one baseline observation and all intervention observations. They were not collected during initial baseline observations due to data collectors missing information on students with consent. Due to the limited number

of observations and the limited number of students included, these data were deemed to be inadequate for the purpose of this study. The data that were collected reflected nondifferentiated disruptive and appropriate behavior across racial identities. Data on student office discipline referrals for teacher 2 was not provided before intervention. She reported providing two office discipline referrals during the implementation of the GBG in her classroom.

Classroom 3

Consent to collect student behavior data were collected for 7 out of the 22 (i.e., 32%) students in Classroom 3. Of those students, there were consents returned for 4 White, 3 Latino, and 0 Black students. Student behavior data were collected for all baseline and intervention observations. Due to the limited number of students included, these data were deemed to be inadequate for the purpose of this study. The data that were collected reflected nondifferentiated disruptive and appropriate behavior across racial identities. Before beginning the GBG, teacher 3 reported providing nine total student office discipline referrals during the months of August through February. She reported providing two office discipline referrals upon beginning the GBG in the months of March and April.

Social Validity

Table 9 outlines teachers' social validity scores on the six domains measured on the URP-I. Overall, teachers reported that they slightly agree that the intervention was acceptable for use in their classrooms. They ranged from agreed to strongly agree that the intervention was understandable and ranged from disagreed to agreed that home school collaboration was necessary to implement the intervention. They reported that they slightly agreed to agree that the intervention was feasible. The teachers agreed that their school climate would support the

intervention, and they ranged from disagreed to slightly agreeing that they had the resources needed to complete the intervention.

Specifically for Classroom 1, the teacher reported that her entire class enjoyed playing the GBG and understood how to play. She stated that sometimes when students lost a point, they would make unsafe choices. She reported that four of her students had a hard time accepting a point removal. She also reported she did not like when a lower random number was pulled and students who lost points still earned rewards. She stated she did not like rewarding students after they were engaging in appropriate behavior during the GBG when they were engaging in disruptive behavior before the GBG began. For Classroom 2, the teacher reported that students in her class easily accepted the GBG. She reported that issues (e.g., arguing and getting upset) with students would arise upon the end of the intervention when a random number was pulled and they did not have enough points for a prize. The teacher in this classroom also reported that following the equitable GBG training, she was better able to focus on whom she was giving points to and remembering to give points to the students who are quieter and do not stand out to her. For Classroom 3, the teacher reported that the students in her class loved the game and were always excited and willing to play the GBG. She reported increased participation and focus every time the GBG was implemented. She also reported she did not like when a lower random number was pulled and students who lost points still earned rewards.

Chapter V: Discussion

The primary purpose of this study was to evaluate if teachers implemented the GBG equitably across racially diverse students in their classrooms. The second purpose of this study was to see if teachers could be trained to implement the GBG equitably across racially diverse students if they were not already doing so. Results indicated that teachers were initially implementing the GBG inequitably in their classrooms. The intervention for this study was a researcher-implemented training on how to incorporate equitable practices into the implementation of GBG, specifically through the use of performance feedback, self-monitoring, and self-awareness. Results of this study indicated that the intervention had limited effects on increasing equitable implementation of the GBG. Although the intervention had minimal impacts on teachers' equitable implementation of the GBG, there were various limitations to the study that possibly led the intervention to be ineffective.

Major Findings and Implications

The following section will review the major findings and implications for each research question across classrooms for both point provision and point removal.

Research Question 1

The GBG is an effective intervention for reducing problem behaviors and increasing appropriate behaviors in classrooms across multiple populations, settings, and variations (Embry, 2002). This intervention has been found to have long-term impacts on impulsive and disruptive behaviors and the prevention of aggressive behavior has been indicated (Kellam et al., 1998; Tankersley, 1995). Although this intervention is effective, limited information exists on if the

implementation of the intervention is equitable across racially diverse students. Research indicates that Black students are three times more likely to be disciplined compared to White students (U.S. Department of Education, 2014). PBIS outcomes show that although overall decreases in student discipline outcomes can be obtained, those outcomes were disproportionate for Black students when compared to White students (Sandomierski, 2011). Knochel (2019) found that teachers could be trained to implement interventions to increase student outcomes overall, but the outcomes had disparities across students of color. Similar to previous research, this current study showed that teachers implemented the GBG inequitably across racially diverse students, however disproportionate findings varied across classrooms and racial identities.

For point provision during the implementation of the GBG without the equity-focused training and self-monitoring tool, disproportionate implementation was observed in 45% of all Classroom 1 observations (i.e., 40%-60% of observations for all racial identities). Although no single racial identity was at a higher risk across all baseline observations, at least one racial identity had disproportionate observations (i.e., overrepresentation of point provision) compared to all other racial identities. For point removal during the implementation of the GBG without the equity-focused training and self-monitoring tool, disproportionate implementation was observed during 27% of all Classroom 1 observations (i.e., 25%-50% of observations for Black, Latino, and Other students). Although fewer observations of disproportionate practices were observed for point removal when compared to point provision overall during baseline, Black students were more likely than all other racial identities to have points removed and White students had zero disproportionate outcomes for point removal.

For point provision during the implementation of the GBG without the equity-focused training and self-monitoring tool, disproportionate implementation was observed during 65% of

all Classroom 2 observations (i.e., 20%-83% of observations for all racial identities). White students were disproportionately overrepresented in point provision compared to all other racial identities. Latino students were disproportionately underrepresented for point provision compared to all other racial identities. Other students were both over- and underrepresented for point provision observations during baseline. Black students had minimal risk of over- and underrepresentation of point removal compared to all other racial identities. Each baseline observation had at least one racial identity with disproportionate outcomes. For point removal during the implementation of the GBG without the equity-focused training and self-monitoring tool, disproportionate implementation was observed during 29% of all Classroom 2 observations (i.e., 50%-67% of observations for White and Latino students). Although fewer observations of disproportionate practices were observed for point removal when compared to point provision overall during baseline, White and Latino students were more likely than Black and Other students to have points removed. Each baseline observation had disproportionate practices for White or Latino students and zero observations had disproportionate practices for Black and Other students.

For point provision during the implementation of the GBG without the equity-focused training and self-monitoring tool, disproportionate implementation was observed during 30% of all Classroom 3 observations (i.e., 11%-55% of observations for all racial identities). Black students were at a higher risk (i.e., overrepresentation of point provision) compared to all other racial identities. White and Latino students had a low risk of disproportionate point provision. Disproportionate observations for at least one racial identity occurred across 78% of all observations. For point removal during the implementation of the GBG without the equity-focused training and self-monitoring tool, disproportionate implementation was observed during

35% of all Classroom 3 observations (i.e., 11-75% of observations for all racial identities). More observations of disproportionate practices were observed for point removal when compared to point provision overall during baseline. White students were at a higher risk for point removal and Black and Latino students had a minimal risk for point removal.

All three teachers implemented the GBG at varying disproportionate levels for point provision and removal. This suggests that it is important to evaluate teacher implementation of class-wide behavioral interventions for equitable practices to address concerns related to implementing interventions inequitably. These results further the notion that identifying ways to increase equity in implementation is necessary.

Research Question 2

Within the literature, practices for reducing disproportionality in schools and classrooms have been identified (Fallon et al., 2012; Leverson et al., 2019; McIntosh et al., 2018). Some of those identified practices include performance feedback, self-monitoring, and self-awareness. The current study provided a training on equitable implementation of the GBG to classrooms identified as having disproportionate point provision and removal across racially diverse students during the implementation of the GBG. This training incorporating equitable practices was created and implemented to answer the second research question: can teachers be trained to implement the GBG equitably across racially diverse students? Unlike previous research incorporating the above-mentioned components for reducing inequitable practices, this study found limited evidence that the intervention components provided produced equitable implementation of the GBG by teachers.

Points Provided. For point provision during the implementation of the GBG with the equity-focused training and self-monitoring tool, Classroom 1 had an increase of equitable

observations overall, and point provision was more undifferentiated across racial identities when compared to baseline. The difference between the racial identity provided the highest average points and racial identity provided the lowest average points decreased from a 0.41 difference to a 0.11 difference. Equitable point provision increased for White, Black, and Latino students; however equitable point provision decreased for Other students. Overall, TRPC data indicate a small change in teacher behavior upon intervention for equitable point provision and TRPC ratios indicate a small change in teacher behavior upon intervention for equitable point provision for most students, but not all, when compared to the comparison group.

For point provision during the implementation of the GBG with the equity-focused training and self-monitoring tool, Classroom 2 had an increase of equitable observations overall, and point provision was more undifferentiated across racial identities when compared to baseline. The difference between the racial identity provided the highest average points and racial identity provided the lowest average points decreased from a 0.63 difference to a 0.22 difference. The percentage of disproportionate observations for White, Latino, and Other students decreased upon intervention, however, Black students had a slight increase in disproportionate observations upon intervention. Overall, TRPC data indicate a small change in teacher behavior upon intervention for equitable point provision and TRPC ratio data indicate a small change in teacher behavior upon intervention for most, but not all, student's equitable point provision.

For point provision during the implementation of the GBG with the equity-focused training and self-monitoring tool, Classroom 3 had an increase of equitable observations overall, and point provision was more undifferentiated across racial identities when compared to baseline. The difference between the racial identity provided the highest average points and

racial identity provided the lowest average points decreased from a 0.73 difference to a 0.28 difference. Black students had an increase in the percentage of equitable observations; however, White and Latino students had a decrease in the percentage of equitable observations. Overall, TRPC data indicate a small change in teacher behavior upon intervention for equitable point provision but TRPC ratio data indicate no significant change in equitable point provision.

Overall, all teachers were able to increase their practice of providing similar amounts of points to students across racial identities and all teachers had some increases in equitable practices for some racial identities, but not for all students. A treatment effect was not indicated for the intervention's ability to increase equitable point provision during the GBG across classrooms. This suggests that the training components used to increase equitable practices alone are not effective for changing teachers' inequitable provision of points during the GBG.

Points Removed. For point removal during the implementation of the GBG with the equity-focused training and self-monitoring tool, Classroom 1 had a decrease in equitable observations overall, and point provision was differentiated for Black students when compared to all other racial identities. The difference between the racial identity with the highest average points removed and racial identity with the lowest average points removed increased from a 0.23 difference to a 0.40 difference. A slight increase in equitable observations for Other students was observed, but a decrease in equitable practices for Black and Latino students was observed. Overall, TRPC and TRPC ratio data indicate no intervention effect for equitable point removal.

For point removal during the implementation of the GBG with the equity-focused training and self-monitoring tool, Classroom 2 had similar percentages of equitable observations, and point provision was less differentiated across students. The difference between the racial identity with the highest average points removed and racial identity with the lowest average

points removed decreased from a 0.89 difference to a 0.69 difference. White students had zero observations of disproportionate removals, which was an increase in equitable practices from baseline to intervention. Black and Other students had a decrease in equitable observations of point removal upon intervention. Latino students had no change in equitable observations from baseline to intervention. Overall, TRPC data indicate a small change in teacher behavior upon intervention for equitable point removal and TRPC ratio data indicate no intervention effect for equitable point removal.

For point removal during the implementation of the GBG without the equity-focused training and self-monitoring tool, Classroom 3 had a decrease in equitable observations overall, and point provision was more undifferentiated across racial identities when compared to baseline. The difference between the racial identity with the highest average points removed and racial identity with the lowest average points removed decreased from a 0.74 difference to a 0.29 difference. White and Black students had more disproportionate observations from baseline to intervention, however, Latino students had zero disproportionate observations during intervention. Overall, TRPC data indicate a small change in behavior upon intervention for equitable point removal and TRPC ratio data indicate no intervention effect for equitable point removal.

Overall, two teachers were able to increase their practice of removing similar amounts of points for students across racial identities and one teacher had an increase in inequitable practices upon intervention. All teachers had some increases in equitable practices for at least one racial identity within each class but had decreases in equitable practices for all other students. Therefore, a treatment effect was not indicated for the intervention's ability to increase equitable point removal during the implementation of the GBG across classrooms. This suggests that the

training components used to increase equitable practices alone are not effective for changing teachers' inequitable removal of points during the GBG.

Research Question 3

The final research question aimed to evaluate if equitable implementation of the GBG had an equitable impact on student behavior and discipline outcomes of racially diverse students. Due to the limited amount of data collected on student behavior, no determination of intervention effects can be conducted.

Limitations and Future Directions

Results of this study indicated disproportionate practices during baseline GBG implementation across teachers and limited intervention effects were indicated regarding the equitable GBG training on improving equitable practices across teachers. Limitations, however, should be noted regarding this study and the results indicated. Due to a global pandemic and the limitations that come from applied research in schools, many elements of the proposed methodology had to be modified and restricted access to beginning the study in schools occurred.

First, the process of beginning the study had multiple delays, which led to limited remaining time in the academic year for data collection procedures. Approval to conduct the study in the school district was obtained in July of 2021. From there, district approval to begin recruitment and data collection procedures in schools was obtained in October of 2021. This three-month delay was due to restrictions regarding the global pandemic and district procedures regarding when research was allowed to begin in schools. The process of recruiting a school, recruiting teachers within that school, collecting informed consent, and training the recruited teachers took five months. Recruiting participants was difficult due to increased stressors and difficulties among teachers and schools associated with the global pandemic (e.g., high teacher

and student absences, larger learning gaps among students, social distancing requirements) (Pattison et al., 2021). Data collection was able to proceed in March of 2022 and observations were conducted until the end of the school year in May of 2022. Future researchers should ensure that an appropriate amount of time is available to conduct procedures and collect data to effectively answer research questions.

Difficulty with data collection persisted due to multiple observation cancellations and limited availability from data collectors and teachers. Observation cancelations occurred across all teachers due to a variety of reasons (e.g., early release day, testing, students earned game day so no academic time, spring break, field day, concert, sick day, holidays, covered multiple classes, having a rough day). Observation cancelations from data collectors occurred due to various reasons as well (e.g., sick, other obligations). The study began with five data collectors before data collection started. During data collection, two data collectors had to remove themselves from the study due to other obligations. Due to only having three data collectors completing observations, it was difficult to obtain IOA across phases and classrooms. Future researchers should be sure to obtain enough IOA observations to meet appropriate standards for evaluating research.

Limitations to data analysis procedures for this study should also be noted. Student behavior data was unable to be evaluated due to a limited number of students' parent consent returned. Due to the length of time that passed during the process of trying to collect students' parent consent, a decision was made to move forward with data collection to obtain information on the first two research questions, as time for data collection was limited. An area for future directions would include data collection on student behavior to determine if students in different racial groups are exhibiting more or less problem behavior compared to their peers. This would

be important to know because if teachers point removal from students was consistent with disruptive behavior, their point removal may not be considered inequitable and could be based on the student behavior within the classroom. Also, TRPC ratios can be difficult to interpret when based on small numbers in either the racial group or the comparison group and TRPC ratios cannot be calculated when the TRPC for the comparison group is zero. Due to the comparison group having ratios of zero, Classroom 1 is missing data for point removal specifically for Black students during one observation, and Latino students during one observation. Classroom 2 is missing data for point provision for Black students during one observation and missing data for point removal for White students during one observation and Black students during one observation. Classroom 3 is missing data for point removal for White students during two observations. A future direction could be the evaluation of equitable practices using different methods, as TRPC ratios can be hard to interpret for small N sizes.

Another limitation regarding data analysis included limitations in the design. The research design was nonconcurrent and teachers moved from baseline to intervention due to time constraints and teacher requests, so it was difficult to evaluate intervention effects through standard multiple baseline design procedures. Although historically, concurrent designs are considered more rigorous than nonconcurrent designs, recent literature argues that the skepticism of nonconcurrent designs is not well-justified (Slocum, Pinkelman, Joslyn, & Nichols, 2022). What Works Clearinghouse provides guidelines for determining if a design meets evidence standards, meets evidence standards with reservations, or does not meet evidence standards (Kratochwill et al., 2010). According to the What Works Clearinghouse, the design for this study met evidence standards in only two ways: the independent variable was systematically manipulated and the design included a minimum of three baseline conditions (Kratochwill et al.,

2010). The design met evidence standards with reservations in the following way: the design included six phases with at least three data points per phase (e.g., the design needs five data points per phase to meet evidence standards) (Kratochwill et al., 2010). Ultimately, however, the design did not meet evidence standards overall because IOA for at least 20% of each phase were not collected (Kratochwill et al., 2010). Future research should improve on the research methods used in this study to increase the evidence standards and ability to interpret the data.

Some limitations specific to teacher participants include that all three teachers were in their first or second year of teaching and had a limited history of implementing class-wide behavior interventions. Teacher 1 specifically had numerous difficulties with the implementation of the GBG and had a classroom with a high frequency of problem behavior and a high number of students with individualized education plans. She reported being afraid to remove points because she did not want students to engage in problem behavior. Future research could use a positive version of the GBG to increase buy-in for implementation from teachers. Also, this teacher was hesitant to provide information to data collectors about students' racial information, so she created a number system providing each student a number and gave racial information to data collectors associated with each assigned number. Data collectors reported that the number system was confusing and made data collection difficult. They also reported that teachers would start the GBG as soon as they entered the classroom, which led to difficulties making sure procedures were being followed before beginning the GBG and the data collector could not pause to ask for clarifying information on student data. Observations in Classroom 1 were conducted during center time, so students were moving around the class often, adding to the difficulty for data collectors to collect data. For Teacher 2, during the equity training, she stated that Latino students are the ones with problem behaviors and that is why they had so many point

removals compared to other students. Due to her perceived level of problem behaviors among Latino students in her classroom, this could have led to less effective outcomes during intervention. More frequent check-ins on how the intervention was going and any perceived difficulties in the classrooms and with data collectors could have helped reduce teacher and data collector confusion and ensure they were both on the same page.

The social validity data across all domains indicated slight agreement that the intervention was acceptable across all teachers. These data are slightly lower than what has been found previously for the GBG (Joslyn et al., 2019). However, research has indicated that teachers who did not believe in positive reinforcement were less likely to find the intervention acceptable (Tingstrom, 1994). Two of the teachers reported not wanting to reward students in the class at the end of the GBG who had any level of behavioral disruptions. Future research could incorporate the importance of positive reinforcement for behavior change to increase levels of acceptability.

Limitations should also be noted regarding treatment fidelity outcomes. Teacher 1 scored a "no" for at least half of the observations on steps 4 (e.g., tell students what they are playing for), 15 (e.g., pull mystery criterion to determine points needed to earn reward),16 (e.g., announce winning student(s) at the conclusion of the game), 17 (e.g., provide verbal praise to winning students), and 18 (e.g., immediately provide rewards). For two out of nine observations, the treatment fidelity average was under 70%. Teacher 2 scored a "no" for at least half of the observations on steps 4, 14 (e.g., announce the conclusion of the game), and 17. All observations were above 70% for treatment fidelity. Teacher 3 scored a "no" for at least half of the observations on steps 4, 17, and 18. Five out of 16 observations resulted in under 70% for treatment fidelity. Future research should employ more rigorous training on the GBG and

equitable implementation of class-wide behavior interventions to increase treatment fidelity. Poor treatment fidelity leads to difficulties interpreting data and suggests the necessity of additional training (Collier-Meek, Fallon, Sanetti, & Maggin, 2013). Researchers have identified strategies for improving treatment fidelity, such as reviewing fidelity to modify practices (Bond et al., 2009). Future research should incorporate this practice proactively to ensure treatment fidelity. More frequent and in-depth evaluation of treatment fidelity areas that are consistently being marked as "no" should be addressed to increase fidelity outcomes.

Procedural limitations should also be noted. At the beginning of data collection, it was determined that some data collectors were collecting data based on their own perceived race of students and not based on race data provided by teachers for two of the classrooms. Those observation data were removed and not represented in the data. Also, it was determined after data collection started, teachers provided their perception of students' race to the data collectors and not student racial data reported from the school. It is possible that teachers perceived race of students could have been different than the students identified race. One training was completed for both data collectors and teachers that included a lot of information regarding procedures. Following the very first day of data collection, it could have been helpful to follow up with teachers and data collectors to ensure an accurate understanding of all procedures and ensure that the procedures were done accurately. Another procedural limitation is that feedback on point provision and removal following each intervention observation was not always provided on the same day of the observation. Some of the feedback was provided multiple days after the observation and was not immediate. Also, it is unclear if teachers were reviewing their data when it was provided. Overall, teachers did not respond to emails sent with their feedback. A future recommendation would be to determine more effective ways to communicate with teachers that

are convenient for them so that they can have access to their data. Also, because the GBG employed an independent group contingency, teachers had the students' names written out for data collection purposes and could have potentially already been visually self-monitoring their point provision and removal before the intervention training. Future research could ensure data collection during baseline was not similar to data collection methods in intervention.

Threats to internal validity should also be noted due to the nature of single case design research (Kratochwill et al., 2010). It is possible that differences between classroom teachers (e.g., tolerance levels) and student characteristics (e.g., disruptive behavior levels) could account for intervention differences and could make comparisons across classrooms difficult. The composition of classrooms also changed over time due to students leaving and joining the classroom during data collection, which could compromise interpretations of an intervention effect. It is also important to note that events occurring outside of the intervention during data collection could account for any observed effect during my research design. Another threat to internal validity includes changes over time occurring naturally with participants that could account for intervention effects. An observed limitation and threat to internal validity was participant dropout. Teacher 1 dropped out after four intervention observations due to a student throwing a chair upon having a point removed during the GBG. Teacher 2 had to end observations early due to the school year ending. These participant departures led to short data series within intervention, making examination of effects difficult. Another notable threat could be classroom observations causing a change in teacher and student behavior. Finally, it is possible that reactivity, drift, bias, and complexity in recording could have influenced data and instrumentation.

Due to the limited information regarding the equitable implementation of class-wide behavior interventions, future research should continue to examine if teachers implement these interventions equitably across racially diverse students. Future research should also continue to examine effective ways for teachers to implement class-wide behavioral interventions equitably across racially diverse students. This information is important for furthering the knowledge of best practices for increasing equitable behavior outcomes in classrooms.

Conclusion

In conclusion, information on whether class-wide behavior interventions are being implemented equitably is largely unknown. This study demonstrated that teachers implemented the class-wide behavior intervention at disproportionate levels across multiple racial identities and observations. Assisting teachers in reducing their levels of disproportionality could reduce discipline disparities for students of color within the classroom setting. The results of this study indicate that it may be possible to shift teachers' implementation to a place of equity, but more rigorous procedures should be developed and incorporated to identify effective methods of change.

References

Allman, K. L., & Slate, J. R. (2011). School discipline in public education: A brief review of current practices. *International Journal of Educational Leadership Preparation*, *6*(2).

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

Augustine, C. H., Engberg, J., Grimm, G. E., Lee, E., Wang, E. L., Christianson, K., & Joseph, A. A. (2018). Can restorative practices improve school climate and curb suspensions? An evaluation of the impact of restorative practices in a mid-sized urban school district. RAND Corporation. doi:10.7249/RR2840.

Axelrod, M. I., Coolong-Chaffin, M., & Hawkins, R. O. (Eds.). (2020). *School-based Behavioral Intervention Case Studies: Effective Problem Solving for School Psychologists*. Routledge.

Bal, A. (2016). From intervention to innovation: A cultural-historical approach to the racialization of school discipline. *Interchange, 47*, 409-427.

Barclay, C. (2017). Benchmarks of equality? School-wide positive behavior interventions and supports and school discipline risk and disparities for Black and Hispanic students. Graduate Theses and Dissertations.

Barrish, H.H., Saunders, M, & Wold, M.M. (1969). Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*, *2*, 119-124.

Belfield, C. R. (2014). The economic burden of high school dropouts and school suspensions in Florida.

Betters-Bubon, J., Brunner, T., & Kansteiner, A. (2016). Success for all? The role of the school counselor in creating and sustaining culturally responsive positive behavior interventions and supports programs. *The Professional Counselor*, *6*, 263-277.

Bond, G., Drake, R. E., McHugo, G. J., Rapp, C. A., & Whitley, R. (2009). Strategies for improving fidelity in the national evidence-based practices project. *Research on Social Work Practice*, *19*(*5*), 569–581. https://doi.org/10.1177/1049731509335531

Bowman-Perrott, L., Benz, M. R., Hsu, H.-Y., Kwok, O.-M., Eisterhold, L. A., & Zhang, D. (2013). Patterns and predictors of disciplinary exclusion over time: An analysis of the seels national data set. *Journal of Emotional and Behavioral Disorders, 2*, 83.

Bradshaw, C. P., Mitchell, M. M., O Brennan, L. M., & Leaf, P. J. (2010). Multilevel exploration of factors contributing to the overrepresentation of black students in office disciplinary referrals. *Journal of Educational Psychology*, *2*, 508.

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

Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: Revision of the usage rating profile-intervention. *Journal of School Psychology*, *51*(1), 81–96. https://doi-org.ezproxy.lib.usf.edu/10.1016/j.jsp.2012.08.006

Chafouleas, S. M., Briesch, A. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2011). *Usage Rating Profile – Intervention (Revised)*. Storrs, CT: University of Connecticut.

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.

Children's Defense Fund. (1975). School suspensions: Are they helping children? Washington, DC: Author.

Collier-Meek, M. A., Fallon, L. M., Sanetti, L. M., & Maggin, D. M. (2013). Focus on implementation: Assessing and promoting treatment fidelity. *Teaching Exceptional Children*, *45*(5), 52-59.

Cooper, J., Heron, T. and Heward, W. (2007). *Applied Behavior Analysis*. 2nd ed. Upper Saddle River, Nj: Pearson.

Danov, S. E., & Symons, F. J. (2008). A survey evaluation of the reliability of visual inspection and functional analysis graphs. *Behavior Modification*, *32*(6), 828–839. https://doi.org/10.1177/0145445508318606

Data Accountability Center (October, 2011). Methods for Assessing Racial/Ethnic Disproportionality in Special Education: A Technical Assistance Guide (Revised), Westat, Rockville, MD, Julie Bollmer, Jim Bethel, Tom Munk, and Amy Bitterman.

Dart, E. H., Radley, K. C., Briesch, A. M., Furlow, C. M., & Cavell, H. J. (2016). Assessing the accuracy of classwide direct observation methods: Two analyses using simulated and naturalistic data. *Behavioral Disorders*, *41*(3), 148–160. https://doi.org/10.17988/BD-15-49.1

Embry, D. D. (2002). The good behavior game: A best practice candidate as a universal behavioral vaccine. *Clinical Child and Family Psychology Review, 4*, 273.

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

Fallon, L. M., Cathcart, S. C., DeFouw, E. R., O'Keeffe, B. V., & Sugai, G. (2018). Promoting teachers' implementation of culturally and contextually relevant class-wide behavior plans. *Psychology in the Schools*, *55*(3), 278–294. https://doi-org.ezproxy.lib.usf.edu/10.1002/pits.22107

Fallon, L. M., O'Keeffe, B. V., & Sugai, G. (2012). Consideration of culture and context in school-wide positive behavior support: A review of current literature. *Journal of Positive Behavior Interventions*, *14*(4), 209-219.

Gershoff, E. T., & Font, S. A. (2016). Corporal punishment in U.S. public schools: Prevalence, disparities in use, and status in state and federal policy. *Social Policy Report*, *30*, 1.

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

Gregory, A., & Fergus, E. (2017). Social and emotional learning and equity in school discipline. *The Future of Children 27*(1), 117-136. doi:10.1353/foc.2017.0006.

Gregory, A., Skiba, R., & Noguera, P. (2010). The achievement gap and the discipline gap: Two sides of the same coin? *Educational Researcher, 39,* 59–68.

Hemphill, S. A., Plenty, S. M., Herrenkohl, T. I., Toumbourou, J. W., & Catalano, R. F. (2014). Student and school factors associated with school suspension: A multilevel analysis of students in Victoria, Australia and Washington State, United States. *Children and Youth Services Review*, *36*, 187–194. https://doi-org.ezproxy.lib.usf.edu/10.1016/j.childyouth.2013.11.022

Intervention Central. (n.d.). *Good behavior game.* https://www.interventioncentral.org/behavioral-interventions/schoolwide-classroommgmt/good-behavior-game

Joslyn, P. R., Donaldson, J. M., Austin, J. L., & Vollmer, T. R. (2019). The good behavior game: A brief review. *Journal of Applied Behavior Analysis*, *3*, 811.

Kellam, S. G., & Anthony, J. C. (1998). Targeting early antecedents to prevent tobacco smoking: Findings from an epidemiologically based randomized field trial. *American Journal of Public Health, 88*, 1490–1495.

Kellam, S.G., Mayer, L. S., Rebok, G. W., & Hawkins, W. E. (1998). Effects of improving achievement on aggressive behavior and of improving aggressive behavior on achievement through two preventive interventions: An investigation of causal paths. In B. P. Dohrenwend (Ed.), Adversity, stress, and psychopathology (pp. 486–505). New York: Oxford University Press.

Knochel, A. (2019). Bridging the achievement gap: Promoting equity in teachers' use of behavior-specific praise with self-monitoring and performance feedback. Graduate Theses and Dissertations. https://scholarcommons.usf.edu/etd/7831

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M & Shadish, W. R. (2010). Single-case designs technical documentation. Retrieved from What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Lai, C. K., Hoffman, K. M., Nosek, B. A., & Greenwald, A. G. (2013). Reducing implicit prejudice. *Social and Personality Psychology Compass*, *7*, 315-330.

Leverson, M., Smith, K., McIntosh, K., Rose, J., & Pinkelman, S. (2019). PBIS Cultural Responsiveness Field Guide: Resources for trainers and coaches. OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports. www.pbis.org.

Marchbanks III, M. P., Blake, J. J., Smith, D., Seibert, A. L., & Carmichael, D. (2014). More than a drop in the bucket: The social and economic costs of dropouts and grade retentions associated with exclusionary discipline. *Journal of Applied Research on Children*, *5*(2), 17.

McIntosh, K., Ellwood, K., McCall, L.,, & Girvan, E. J. (2018). Using discipline data within a PBIS framework to enhance equity in school discipline. *Intervention in School and Clinic*, *53*, 146-152.

McIntosh, K., Girvan, E. J., Horner, R. H., Smolkowski, K., & Sugai, G. (2014). Recommendations for addressing discipline disproportionality in education. OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports. www.pbis.org.

McIntosh, K., Girvan, E. J., Horner, R. H., Smolkowski, K., & Sugai, G. (2018). A 5-point intervention approach for enhancing equity in school discipline. OSEP Technical Assistance Center on Positive Behavioral Interventions and Supports.

Miltenberger, R. (2004). *Behaviour modification*: principals and procedure (3rd ed.) Belmont, CA. Wadsworth Publishing.

Morris, E. W., & Perry, B. L. (2017). Girls behaving badly? Race, gender, and subjective evaluation in the discipline of African American girls. *Sociology of education*, *90*(2), 127-148.

Nowicki, J. M., & US Government Accountability Office. (2018). K-12 education: Discipline disparities for Black students, boys, and students with disabilities. Report to Congressional Requesters. GAO-18-258. *US Government Accountability Office*.

Ohio Department of Education. (n.d.). *The good behavior game manual.* https://education.ohio.gov/getattachment/Topics/Other-Resources/School-Safety/Building-Better-Learning-Environments/Promote-Pro-Social-Behavior/Good-Behavior-Game-Manual.pdf.aspx

Pattison, K., Hoke, A. M., Schaefer, E. W., Alter, J., & Sekhar, D. L. (2021). National survey of school employees: COVID-19, school reopening, and student wellness. *The Journal of School Health*, *91*(5), 376–383. https://doi.org/10.1111/josh.13010

Reno, G. D., Friend, J., Caruthers, L. & Smith, D. (2018). Who's getting targeted for behavioral interventions? Exploring the connection between school culture, positive behavior support, and elementary student achievement. *The Journal of Negro Education, 86,* 423-438.

Rumberger, R. W., & Losen, D. J. (2017). The Hidden Costs of California's Harsh School Discipline: And the Localized Economic Benefits from Suspending Fewer High School Students. *Civil Rights Project-Proyecto Derechos Civiles*.

Sandomierski, T. (2011). Disciplinary outcomes by race and gender in schools implementing positive behavior support: Does fidelity of implementation reduce disproportionality? (Doctoral dissertation, University of South Florida).

Skiba, R. J., Horner, R. H., Chung, C.-G., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race is not neutral: A national investigation of african american and latino disproportionality in school discipline. *School Psychology Review*, *1*, 85.

Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review, 34(*4*),* 317-342.

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.

Slocum, T. A., Pinkelman, S. E., Joslyn, P. R., Nichols, B., (2022). Threats to internal validity in multiple-baseline design variations. *Perspectives on Behavior Science*. https://doi.org/10.1007/s40614-022-00326-1

Smolkowski, K., Girvan, E. J., McIntosh, K., Nese, R. N. T., & Horner, R. H. (2016). Vulnerable decision points for disproportionate office discipline referrals: comparison of discipline for African American and white elementary school students. *Behavioral Disorders, 41,* 178-195.

Song, S. Y., Eddy, J. M., Thompson, H. M., Adams, B., & Beskow, J. (2020). Restorative consultation in schools: A systematic review and call for restorative justice science to promote anti-racism and social justice. *Journal of Educational & Psychological Consultation*, *30*(4), 462–476. https://doi-org.ezproxy.lib.usf.edu/10.1080/10474412.2020.1819298

Samuel Y. Song & Susan M. Swearer (2016). The cart before the horse: The challenge and promise of restorative justice consultation in schools. *Journal of Educational and Psychological Consultation*, *26*(4), 313-324, DOI: 10.1080/10474412.2016.1246972

Staats, C. (2016). Understanding implicit bias: What educators should know. *American Educator*, *39*(4), 29–33.

Sugai, G., & Horner, R. (2009). Responsiveness-to-intervention and school-wide positive behavior supports: Integration of multi-tiered system approaches. *Exceptionality*, *17*(4), 223-237. doi: 10.1080/09362830903235375

Tingstrom, D. (1994). The good behavior game: An investigation of teachers' acceptance. *Psychology in the Schools*, *31*(1), 57–65. https://doi.org/10.1002/1520-6807(199401)31:1<57::AID-PITS2310310108>3.0.CO;2-K

U.S. Department of Education Office for Civil Rights. (2014). Data snapshot: School discipline. Washington, DC: U.S. Department of Education

U.S. Department of Education. (2016). The 38th annual report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2016. Washington, DC: Office for Special Education and Rehabilitative Services, U.S. Department of Education.

U.S. Department of Health and Human Services and U.S. Department of Education: Policy Statement on Expulsion and Suspension Policies in Early Childhood Settings (November 2016).

Vincent, C. G., Swain-Bradway, J., Tobin, T. J., & May, S. (2011). Disciplinary referrals for culturally and linguistically diverse students with and without disabilities: Patterns resulting from school-wide positive behavior support. *Exceptionality*, *19*, 175-190.

Wahl, E., Hawkins, R. O., Haydon, T., Marsicano, R., & Morrison, J. Q. (2016). Comparing versions of the good behavior game: Can a positive spin enhance effectiveness? *Behavior Modification, 40,* 493-517. https://doi-org.ezproxy.lib.usf.edu/10.1177/0145445516644220

Wong, D. W., Cheng, C. K., Ngan, R. H., & Ma, S. K. (2011). Program effectiveness of a restorative whole-school approach for tackling school bullying in Hong Kong. *International Journal of Offender Therapy and Comparative Criminology*, *55*(6), 846–862. doi:10.1177/0306624X10374638

Wright, R. A., & McCurdy, B. L. (2012). Class-wide positive behavior support and group contingencies: Examining a positive variation of the good behavior game. *Journal of Positive Behavior Interventions, 14,* 173-180.

Appendix A

Data Collection Tool: Teacher Behavior

Date: _____ Classroom: _____ Observer: _____ Phase: _____

Directions: Collect frequency within 1-minute data for teacher behavior. For each racial category, record the number of students present for the observation for "N". Record frequency of points given within each interval by placing a tally in the "+" column. Record frequency of points removed within each interval by placing a tally in the "-" column.

	1 (N=)		**2 (N=)**		**3 (N=)**		**4 (N=)**	
	+	-	+	-	+	-	+	-
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								
26								
27								
28								
29								
30								
Total								
Rate								

Points Given: total frequency = _____; Rate per student = _____
Points Removed: total frequency = _____; Rate per student = _____
IOA: Points given: # of intervals with agreement ___ / total # of intervals ___ x100 = ____%
Points removed: # of intervals with agreement ___ / total # of intervals ___ x100 = ____%

Appendix B

Data Collection Tool: Student Behavior

Date: _____ Classroom: _____ Observer: _____ Phase: _____

Directions: Use momentary time sampling to collect data at the end of a 30-second interval for student behavior. Before observing, determine the order you will observe each student. Each student will be observed individually and once all have been observed in the given order, the students will be observed again in the same order. For each racial category, record the number of students present for the observation for "N". Record a "D" if the student is engaging in disruptive behavior at the end of the interval. Record an "A" if the student is engaging in appropriate behavior at the end of the interval. Record the observed students' race using "W" for White, "B" for Black, "L" for Latino, and "O" for Other for each interval.

Time	Inter	D/A	Race	Time	Inter	D/A	Race	Time	Inter	D/A	Race
0:30	**1**			10:30	**21**			20:30	**41**		
1:00	**2**			11:00	**22**			21:00	**42**		
1:30	**3**			11:30	**23**			21:30	**43**		
2:00	**4**			12:00	**24**			22:00	**44**		
2:30	**5**			12:30	**25**			22:30	**45**		
3:00	**6**			13:00	**26**			23:00	**46**		
3:30	**7**			13:30	**27**			23:30	**47**		
4:00	**8**			14:00	**28**			24:00	**48**		
4:30	**9**			14:30	**29**			24:30	**49**		
5:00	**10**			15:00	**30**			25:00	**50**		
5:30	**11**			15:30	**31**			25:30	**51**		
6:00	**12**			16:00	**32**			26:00	**52**		
6:30	**13**			16:30	**33**			26:30	**53**		
7:00	**14**			17:00	**34**			27:00	**54**		
7:30	**15**			17:30	**35**			27:30	**55**		
8:00	**16**			18:00	**36**			28:00	**56**		
8:30	**17**			18:30	**37**			28:30	**57**		
9:00	**18**			19:00	**38**			29:00	**58**		
9:30	**19**			19:30	**39**			29:30	**59**		
10:00	**20**			20:00	**40**			30:00	**60**		

1 (N=____): Total D = ____ Total A = ____
2 (N=____): Total D = ____ Total A = ____
3 (N=____): Total D = ____ Total A = ____
4 (N=____): Total D = ____ Total A = ____

IOA: # of intervals with agreement ___ / total # of intervals ___ x100 = ____%

Disruptive Behavior: Talking without permission such a talking to self or others, yelling, whistling, or making other noises; being out of their seat without permission including standing up or walking around the room; non-compliance to teacher demands; and physical disruption to others or property including hitting, kicking, throwing objects, or destroying items.

Appropriate Behavior: Engagement in any behavior that matches the ongoing classroom instruction (e.g., the class is writing, the target student is writing); and talking or being out of their seat with teacher permission.

Appendix C

Treatment Fidelity

Steps	Implemented?
1. Announce the game	
2. State rules of the game	
3. Announce the duration of the game	
4. Tell students what they are playing for (reward)	
5. Announce the start of the game	
6. Start timer	
7. Verbally indicate which rule was violated	
8. State which student was responsible for the rule violation	
9. Remove a mark when a student violates a rule	
10. Verbally indicate which rule was followed	
11. State which student was responsible for the rule-following	
12. Place a mark when a student follows a rule	
13. Ensure that the conclusion of the game is accompanied by an audible indicator (e.g., the alarm from the timer is audible to students)	
14. Announce the conclusion of the game	
15. Pull mystery criterion to determine points needed to earn reward	
16. Announce winning student(s) at the conclusion of the game	
17. Provide verbal praise to winning student(s)	
18. Immediately provide rewards	
Number of Steps Performed	_____

Appendix D

Procedural Integrity: GBG Training

Steps	Implemented?
1. Provide all trainees with a personal copy of the supplemental GBG Training Tool	
2. Introduce the training and PowerPoint	
3. Spoke clearly to ensure trainees could hear throughout the presentation	
4. Powerpoint presentation was visible to all trainees	
5. Trainer outlined each step of the GBG	
6. Trainer ensured that all questions were addressed throughout the presentation	
7. Trainer ensured that trainees were provided opportunities to discuss hypothetical situations and troubleshoot barriers	
8. Trainer ensured that trainees were provided opportunities to practice each step of implementing the GBG	
9. The trainer provided individual feedback for teachers rehearing the step of the GBG	
10. Trainer addressed all questions or concerns before concluding the training	
Number of Steps Performed	_____

Appendix E

Procedural Integrity: Equitable GBG Training

Steps	Implemented?
1. Provide teacher with PowerPoint on overview of equitable implementation of the GBG	
2. Provide the teacher with data on the student racial identities that received the least and most points during baseline	
3. Provide the teacher with data on the student racial identities that had the least and most points removed during baseline	
4. Provide feedback on how they can use equitable classroom practices during implementation	
5. Provide tool and training on self-monitoring equitable implementation	
6. Answer any questions teachers may have regarding study procedures	
Number of Steps Performed	_____

Appendix F

Teacher Screening

1. How many years have you been teaching?
2. What certifications do you have?
3. What is your highest educational level attained?
4. What is your race, ethnicity, gender, and age?
5. Are there any specific academic periods that your class has particular difficulty focusing?

   If so, how many students would you say are off-task during that time?
6. What classroom management procedures do you use?
7. How many students have received referrals in your classroom? How many per student?
8. How do you reward your students for good behavior?
9. How do you handle your student's inappropriate behavior?
10. What type of off-task behaviors do your students engage in?
11. What kinds of on-task behavior would you like to see your students engage in?
12. What is the demographic make-up of your classroom? Race, gender, age…

Appendix G

Task Analysis of Behavior Skills Training for Teacher Participants

Didactic

1. The primary investigator will provide concise details of the Good Behavior Game (GBG) and its effective history.
2. Each step of the GBG will be outlined with opportunities for teacher participants to have questions addressed.

Modeling

1. The primary investigator will provide video examples of teachers implementing the GBG in classroom settings.
2. The primary investigator will provide brief demonstrations of implementing each step of the GBG.
3. Demonstrations will include example statements which teachers can utilize when implementing with their students.

Rehearsal

1. Each teacher participant will have an opportunity to practice each step of implementing the GBG.

Feedback

1. The primary investigator will provide individual feedback based on the rehearsal implementation of the GBG.
2. Final questions or concerns posed by teacher participants will be addressed before concluding training.

Appendix H

Self-Monitoring Tool for Equitable GBG Implementation

Student	Race	Starting Points	Point Given	Point Removed	Total (starting points + points given - points removed)
Ex: John Doe	Ex: White	Ex: 11111	Ex: 11	Ex: 1	
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			
		11111			

Appendix I

IRB Approval

UNIVERSITY of SOUTH FLORIDA

APPROVAL

August 17, 2021

Faith Reynolds
8708 Busch Oaks St
Tampa, FL 33617

Dear Ms. Faith Reynolds:

On 8/16/2021, the IRB reviewed and approved the following protocol:

Application Type:	Initial Study
IRB ID:	STUDY002617
Review Type:	Expedited 7
Title:	Equitable Implementation of the Good Behavior Game
Funding:	Society for the Study of School Psychology
Approved Protocol and Consent(s)/Assent(s):	• Equitable Implementation of the Good Behavior Game Protocol_FReynolds.docx; • Social Behavioral Assent_FR_GBG.pdf; • Social Behavioral Parental Permission_FR_GBG.pdf; • Social-Behavioral Adult Consent_FReynolds.pdf; Approved study documents can be found under the 'Documents' tab in the main study workspace. Use the stamped consent found under the 'Last Finalized' column under the 'Documents' tab.

This study involving child participants falls under the minimal risk category 45 CFR 46.404: Research not involving greater than minimal risk.

Requirements for Assent and/or Permission by Parents or Guardians: 45 CFR 46.408 Permission of one parent is sufficient. Verbal Assent will be obtained as outlined in the IRB application.

Within 30 days of the anniversary date of study approval, confirm your research is ongoing by clicking Confirm Ongoing Research in BullsIRB, or if your research is complete, submit a study closure request in BullsIRB by clicking Create Modification/CR.

UNIVERSITY of SOUTH FLORIDA

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

Sincerely,

Jennifer Walker
IRB Research Compliance Administrator

Appendix J

Tables and Figures

Table 1

Average Treatment Fidelity across Study Phases and Classrooms

Treatment Fidelity			
	Baseline	Intervention	Overall
Classroom 1	79% (56%-94%)	69% (56%-78%)	75%
Classroom 2	77% (72%-83%)	78% (72%-83%)	77%
Classroom 3	77% (61%-89%)	78% (33%-95%)	77%

Note. Range of scores are noted in parentheses.

Table 2

Average Interobserver Agreement across Study Phases and Classrooms

Study Phase: Baseline			
	Teacher Fidelity	Teacher Behavior	Student Behavior
Classroom 1	N/A	N/A	N/A
Classroom 2	N/A	N/A	N/A
Classroom 3	100%	86%	85%
Study Phase: Intervention			
	Teacher Fidelity	Teacher Behavior	Student Behavior
Classroom 1	88%	99%	N/A
Classroom 2	N/A	N/A	N/A
Classroom 3	89%	98%	75%

Note. N/A = Not Applicable.

Table 3

Average TRPC for Points Provided across Phases and Classrooms

Study Phase: Baseline					
	White	Black	Latino	Other	Overall
Classroom 1	0.77 (0.14-1.40)	1.18 (0.25-2.00)	0.90 (0.43-1.29)	0.90 (0.00-2.00)	0.81 (0.33-1.37)
Classroom 2	1.27 (0.00-3.17)	0.97 (0.25-2.5)	0.70 (0.00-3.00)	1.33 (0.00-4.00)	1.11 (0.06-2.85)
Classroom 3	1.12 (0.63-2.00)	1.85 (1.00-4.00)	1.32 (0.33-2.00)	N/A	1.26 (0.57-1.89)
Study Phase: Intervention					
	White	Black	Latino	Other	Overall
Classroom 1	1.03 (0.50-1.43)	0.96 (0.50-1.33)	0.92 (0.50-1.17)	1.00 (0.47-1.33)	0.97 (0.47-1.33)
Classroom 2	1.41 (0.38-2.25)	1.47 (0.50-2.00)	1.38 (0.00-2.00)	1.25 (0.00-2.00)	1.42 (0.29-1.88)
Classroom 3	0.89 (0.44-1.40)	1.17 (0.50-2.00)	1.11 (0.56-1.63)	N/A	1.00 (0.60-1.24)

Note. Range of scores are noted in parentheses. N/A = Not Applicable.

Table 4

Average TRPC Ratios for Points Provided across Phases and Classrooms

Study Phase: Baseline				
	White	Black	Latino	Other
Classroom 1	0.96 (0.31-2.00)	1.25 (0.32-2.50)	1.28 (0.91-1.57)	1.34 (0.00-4.44)
Classroom 2	1.68 (0.00-4.00)	0.82 (0.43-1.35)	0.35 (0.00-1.06)	1.00 (0.00-2.71)
Classroom 3	0.84 (0.47-1.33)	1.86 (0.68-5.80)	1.03 (0.44-1.60)	N/A
Study Phase: Intervention				
	White	Black	Latino	Other
Classroom 1	1.11 (1.07-1.14)	1.00 (0.75-1.24)	0.96 (0.82-1.10)	0.88 (0.00-1.55)
Classroom 2	1.28 (0.62-2.25)	1.27 (0.86-2.00)	0.85 (0.00-1.73)	0.68 (0.00-1.07)
Classroom 3	0.85 (0.49-1.22)	1.09 (0.67-2.15)	1.21 (0.82-1.63)	N/A

Note. Range of scores are noted in parentheses. N/A = Not Applicable.

Table 5

Percent of Low, Moderate, and High Disproportionate Observations for Point Provision

Classroom 1 – Point Provision Study Phase: Baseline					
	White	Black	Latino	Other	Overall
Low	60%	60%	40%	60%	55%
Moderate	20%	20%	60%	0%	25%
High	20%	20%	0%	40%	20%
Study Phase: Intervention					
	White	Black	Latino	Other	Overall
Low	100%	75%	100%	50%	81%
Moderate	0%	25%	0%	25%	13%
High	0%	0%	0%	25%	6%

Classroom 2 – Point Provision Study Phase: Baseline					
	White	Black	Latino	Other	Overall
Low	17%	80%	33%	17%	35%
Moderate	33%	20%	0%	0%	13%
High	50%	0%	67%	83%	52%
Study Phase: Intervention					
	White	Black	Latino	Other	Overall
Low	75%	75%	50%	75%	69%
Moderate	0%	0%	25%	0%	6%
High	25%	25%	25%	25%	25%

Classroom 3 – Point Provision Study Phase: Baseline				
	White	Black	Latino	Overall
Low	89%	45%	78%	70%
Moderate	11%	22%	22%	19%
High	0%	33%	0%	11%
Study Phase: Intervention				
	White	Black	Latino	Overall
Low	86%	72%	72%	76%
Moderate	14%	14%	14%	14%
High	0%	14%	14%	10%

Table 6

Average TRPC for Points Removed across Phases and Classrooms

Study Phase: Baseline					
	White	Black	Latino	Other	Overall
Classroom 1	0.02 (0.00-0.10)	0.25 (0.00-0.75)	0.12 (0.00-0.29)	0.10 (0.00-0.50)	0.11 (0.00-0.26)
Classroom 2	0.83 (0.17-1.30)	0.48 (0.00-1.00)	1.06 (0.67-1.67)	0.17 (0.00-1.00)	0.75 (0.25-1.13)
Classroom 3	0.56 (0.11-1.22)	0.22 (0.00-2.00)	0.18 (0.00-0.56)	N/A	0.35 (0.14-0.65)
Study Phase: Intervention					
	White	Black	Latino	Other	Overall
Classroom 1	0.04 (0.00-0.14)	0.40 (0.25-0.75)	0.16 (0.00-0.33)	0.00 (0.00-0.00)	0.15 (0.06-0.25)
Classroom 2	0.19 (0.00-0.36)	0.27 (0.00-0.50)	0.88 (0.00-2.50)	0.25 (0.00-1.00)	0.30 (0.06-0.65)
Classroom 3	0.41 (0.11-0.67)	0.47 (0.00-1.00)	0.18 (0.00-0.38)	N/A	0.31 (0.06-0.52)

Note. Range of scores are noted in parentheses. N/A = Not Applicable.

Table 7

Average TRPC Ratios for Points Removed across Phases and Classrooms

Study Phase: Baseline				
	White	Black	Latino	Other
Classroom 1	0.28 (0.00-1.10)	2.28 (0.00-5.63)	1.21 (0.00-2.50)	0.88 (0.00-3.50)
Classroom 2	1.28 (0.44-1.78)	0.45 (0.00-0.87)	2.71 (0.87-8.50)	0.19 (0.00-1.15)
Classroom 3	4.06 (0.67-12.22)	0.40 (0.00-3.56)	0.73 (0.00-2.67)	N/A
Study Phase: Intervention				
	White	Black	Latino	Other
Classroom 1	0.20 (0.00-0.79)	5.14 (1.75-9.00)	0.97 (0.00-1.44)	0.00 (0.00-0.00)
Classroom 2	0.17 (0.00-0.31)	0.66 (0.00-1.50)	2.94 (0.00-6.25)	0.40 (0.00-1.60)
Classroom 3	2.12 (0.31-6.11)	1.78 (0.00-2.25)	0.43 (0.00-1.81)	N/A

Note. Range of scores are noted in parentheses. N/A = Not Applicable.

Table 8

Percent of Low, Moderate, and High Disproportionate Observations for Point Removal across Phases and Classrooms

Classroom 1 – Point Removal Study Phase: Baseline					
	White	Black	Latino	Other	Overall
Low	100%	50%	67%	75%	73%
Moderate	0%	0%	0%	0%	0%
High	0%	50%	33%	25%	27%
Study Phase: Intervention					
	White	Black	Latino	Other	Overall
Low	100%	0%	50%	100%	67%
Moderate	0%	33%	50%	0%	20%
High	0%	67%	0%	0%	13%

Classroom 2 – Point Removal Study Phase: Baseline					
	White	Black	Latino	Other	Overall
Low	33%	100%	50%	100%	71%
Moderate	67%	0%	0%	0%	17%
High	0%	0%	50%	0%	12%
Study Phase: Intervention					
	White	Black	Latino	Other	Overall
Low	100%	67%	50%	75%	72%
Moderate	0%	33%	0%	25%	14%
High	0%	0%	50%	0%	14%

Classroom 3 – Point Removal Study Phase: Baseline				
	White	Black	Latino	Overall
Low	25%	89%	78%	65%
Moderate	12.5%	0%	11%	8%
High	62.5%	11%	11%	27%
Study Phase: Intervention				
	White	Black	Latino	Overall
Low	16.5%	57%	100%	60%
Moderate	67%	14%	0%	25%
High	16.5%	29%	0%	15%

Table 9

Average Social Validity Scores across Classrooms

Usage Rating Profile-Intervention			
	Classroom 1	Classroom 2	Classroom 3
Acceptability	3.67	4.22	4.44
Understanding	5.67	5.33	5.00
Home School Collaboration	5.33	2.33	3.67
Feasibility	4.33	4.67	4.67
System Climate	4.60	4.80	4.60
System Support	2.00	4.00	4.00
Overall	4.27	4.06	4.40

Figure 1

TRPC for Points Provided for all Racial Identities and Class Average across Phases and Classrooms

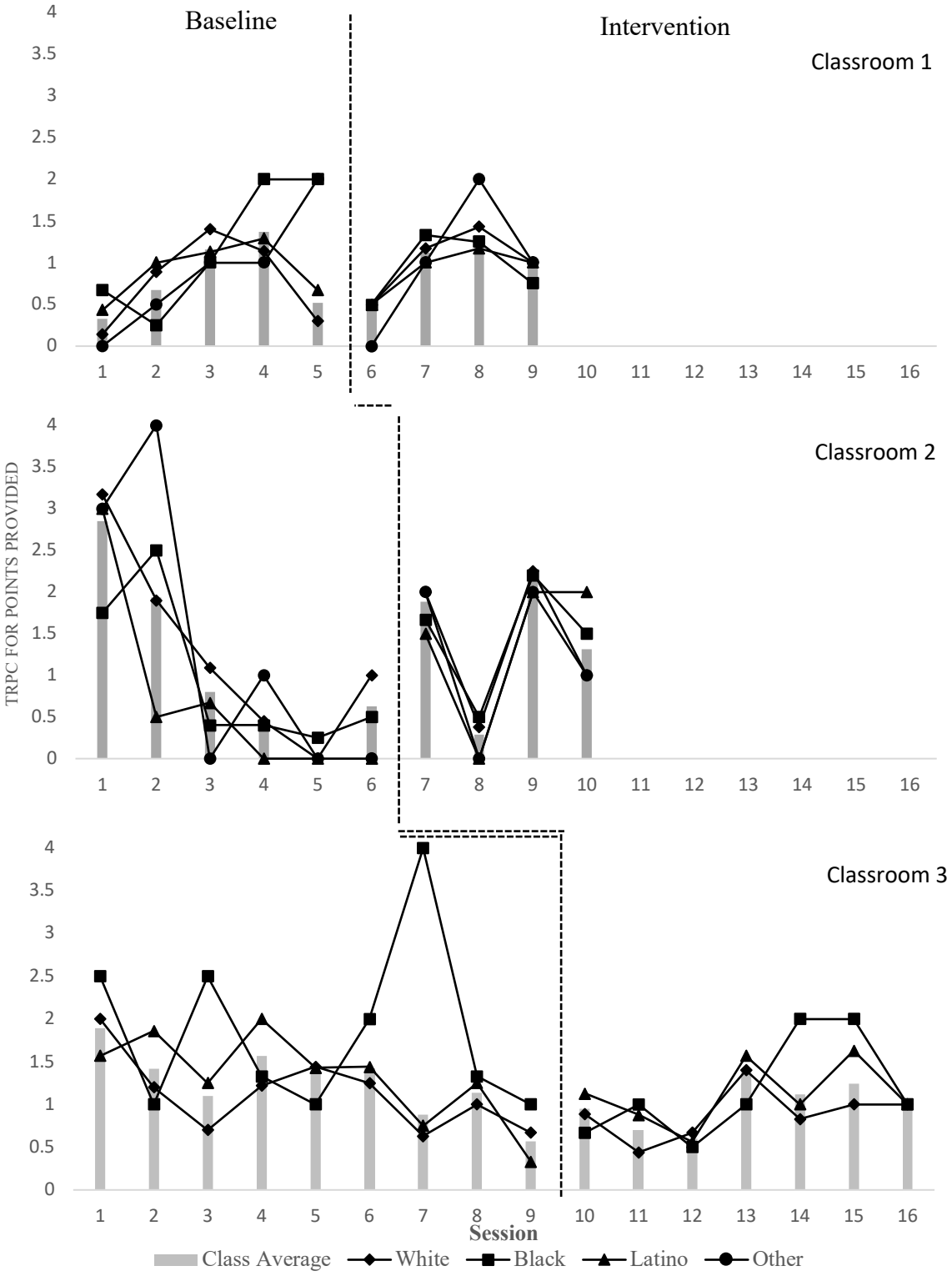

Figure 2

TRPC Ratio for Points Provided for all Racial Identities across Phases and Classrooms

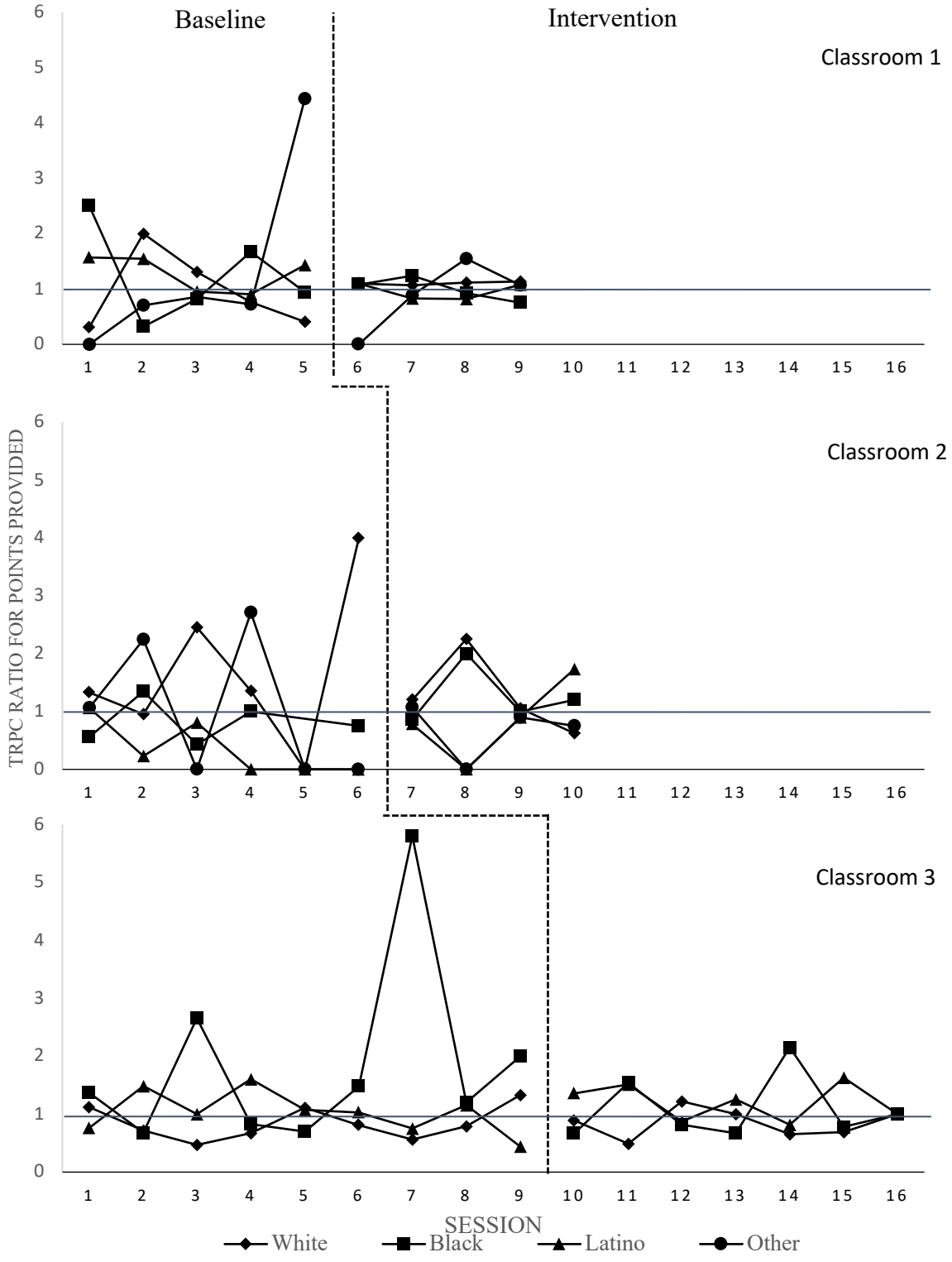

Figure 3

TRPC for Points Removed for all Racial Identities and Class Average across Phases and Classrooms

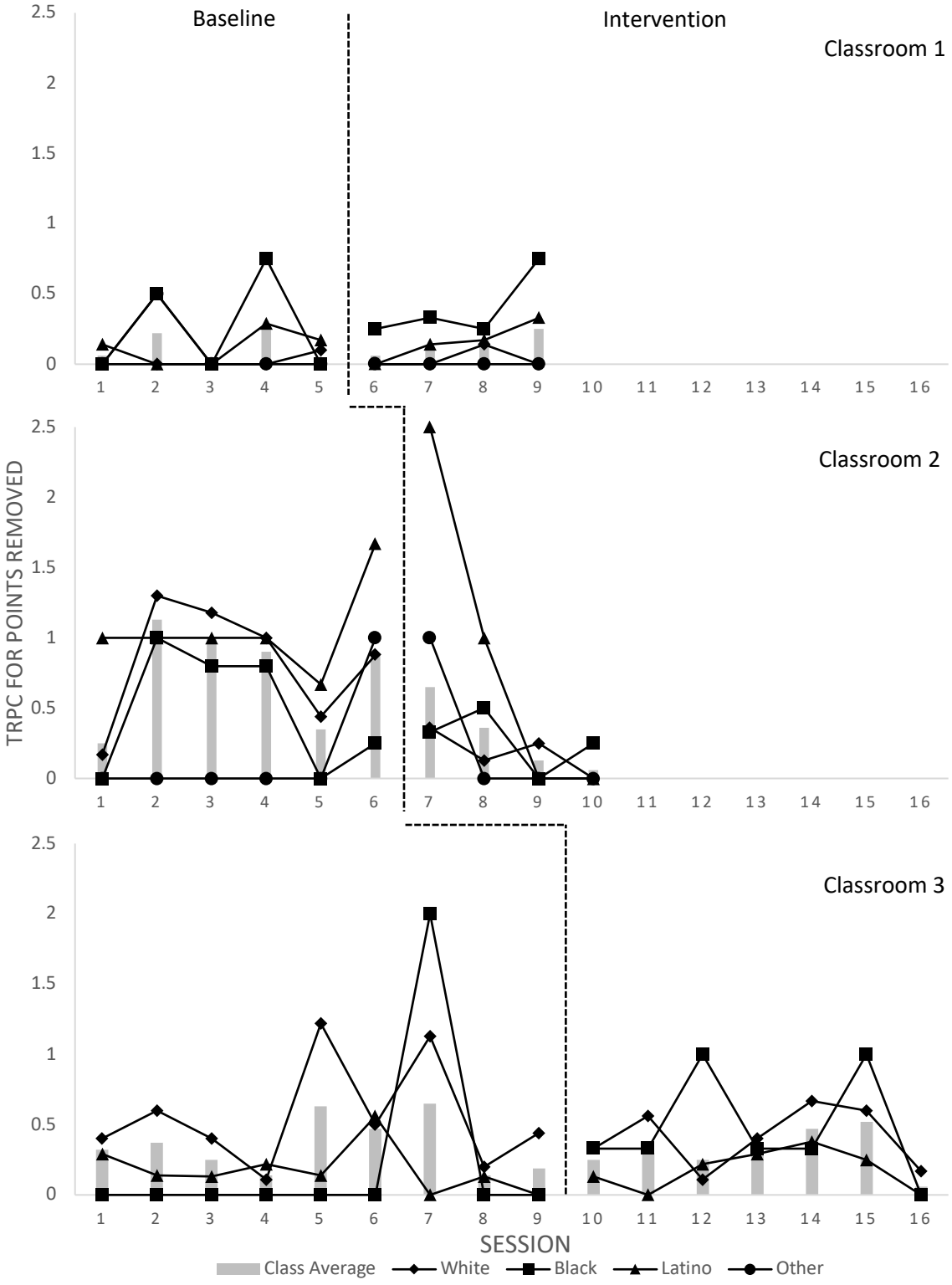

Figure 4

TRPC Ratio for Points Removed for all Racial Identities across Phases and Classrooms