

USING RISK RATIOS AS A METHOD OF CALCULATING SUBSTANTIAL RACIAL AND
ETHNIC DISPROPORTIONALITY RATES IN SCHOOL DISCIPLINE

by

ADAM GHEMRAOUI

B.Sc., York University, 2012

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR

THE DEGREE OF

MASTER OF ARTS

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(School Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

September 2017

© Adam Ghemraoui, 2017

Abstract

The discipline gap – a phenomenon by which students of colour (e.g., Black students) are disproportionately affected by school exclusionary discipline compared to their White peers – has been reliably documented for more than 50 years. Researchers have relied on different metrics, analysis methods, and data sources to measure the discipline gap. Regulators have proposed the standard use of risk ratios as a metric to measure disproportionality. Risk ratios require that the target group (e.g., Black students) be compared to another group (e.g., White students), however, there is a paucity of studies on the differential impact of using White students versus all other students as comparison groups. I analyzed data from 5,422 schools from the 2012 – 2014 academic school years across the United States by fitting two series of mixed models to account for the nested structure of the data. I evaluated the effect of using different comparison groups on risk ratio values as well as school disproportionate status. Results indicate that the use of all other students as a comparison group yields significantly higher mean risk ratio values over three years for Black students to receive at least one out-of-school suspension (OSS) by a factor of 2.621. The predicted odds of a school’s risk ratio value being significantly disproportionate (i.e., compared to a threshold value) increases by a factor of 1.790 when using all other students as a comparison group. The mean risk ratio values for Black students to receive at least one OSS were significantly higher in 2014 – 2015 than in 2012 – 2013, regardless of which comparison group is used. Implications for both policy makers and researchers are discussed in light of the findings and proposed legislation.

Lay Summary

In the United States, Black students are suspended at higher rates than their White counterparts, a phenomenon that is called the “discipline gap”. While we know there is a gap, we don’t have a standard way of measuring it. One major issue is who should we compare Black students’ suspension rates with – the rates of suspension for White students or all other students? The findings of the study tell us that the gap is much higher when we compare the suspension rates of Black students with those of all other students. Also, Black students have been suspended at much higher rates than their White peers in 2014 – 2015 compared to in 2012 – 2013. This is important to know since there are proposed legislative changes that will make state Departments of Education compare Black students with all other students.

Preface

This thesis is the original, unpublished work of Adam Ghemraoui, completed under the supervision of Dr. Sterett Mercer. The data used in this thesis, in part, were obtained from the Educational and Community Supports research unit at the University of Oregon, in collaboration with Dr. Kent McIntosh. This study was conducted with approval from the University of British Columbia's Behavioural Research Ethics Board (H12-02388).

Table of Contents

Abstract.....	ii
Lay Summary.....	iii
Preface.....	iv
Table of Contents.....	v
List of Tables.....	viii
List of Figures.....	ix
Acknowledgements.....	x
Dedication.....	xi
Chapter 1: Introduction.....	1
Chapter 2: Literature Review.....	4
The Discipline Gap.....	4
Long Term Effects of Exclusionary Discipline.....	5
School-to-prison Pipeline.....	8
Factors Contributing to Discipline Gap.....	11
Methodological Issues in Identifying the Discipline Gap.....	13
Using Risk Ratios to Identify Significant Disproportionalities.....	17
Proposed Study.....	19
Chapter 3: Methods.....	21
Participants and Settings.....	21

Measures	21
Procedures	22
Risk indices for Black and White students.	23
Risk index for all other students.	23
Risk ratios.	23
Determining significant disproportionalities.	24
Data Analysis Plan	24
Chapter 4: Results	26
Descriptive Statistics.....	26
Model Fitting: Risk Ratios as Outcome Variable.....	27
Model Fitting: Disproportionate Status as Outcome Variable.....	29
Chapter 5: Discussion	31
Study Strength and Limitations	33
Implications for Practice and Policy	35
Implications for Research	37
Chapter 6: Summary	39
Figures.....	40
Tables	42
References.....	52
Appendix A.....	60

List of possible infractions and outcomes that can be recorded in SWIS..... 60

List of Tables

Table 1 Proportion of juvenile arrests by Black youth between 2003 and 2012	42
Table 2 Summary of schools included in analysis	42
Table 3 Percentage of schools in observed data that have disproportionate risk ratios.....	43
Table 4 ICC values based on empty model	43
Table 5 Mean, SD, skewness, and kurtosis of risk ratios pre- and post- log transformation.....	43
Table 6 Mean risk ratios and log risk ratios for Black students to receive at least one OSS, based on observed data	44
Table 7 Summary of model estimates with systematic addition of random intercepts	44
Table 8 Summary of results of log likelihood tests between models with random intercepts.....	45
Table 9 Summary of explanatory variables in models with addition of Comparison Group, Time, and Comparison Group*Time.....	46
Table 10 Summary of results of log likelihood tests between models with fixed effects	47
Table 11 Summary of results of log likelihood tests between models with addition of random slope of time.....	47
Table 12 Summary of explanatory variables for final model (i.e., Model 1.7) selected from first series of models.....	48
Table 13 Model predicted log risk ratios	49
Table 14 Summary of results of chi-square tests between models in second series of models	50
Table 15 Summary of explanatory variables for models in second series of model fitting.....	51
Table 16 Model predicted probability of disproportionate status	51

List of Figures

Figure 1 Quantile-Quantile Plot of risk ratios and log risk ratios variables	40
Figure 2 Risk ratios for Black students to receive at least one OSS, based on observed data	41
Figure 3 Log risk ratios for Black students to receive at least one OSS, based on observed data	41

Acknowledgements

I am grateful to, and have utmost respect for, my supervisor, Dr. Mercer, for his mentorship and patience in helping me develop my research and writing skills. Thank you to my cohort for their support throughout the last three years. I would like to acknowledge my parents, who have believed in me when I didn't believe in myself and have supported me when I couldn't support myself., and have always set expectations that I become the best possible version of myself. Finally, I would like to acknowledge my wife, Yasmine, whose sacrifices and love have allowed me to pursue my degree. Through my faith, I am able to recognize the enumerable blessings from my Creator that are in my life.

Dedication

My thesis is dedicated to students who have been discriminated against due to their skin colour. I hope that my thesis can have an impact on reducing the discipline gap so that future students may perhaps be treated in a more equitable manner.

Chapter 1: Introduction

According to the latest data released by the Office of Civil Rights (Department of Education, 2016) in the 2013 – 2014 school year, 2.8 million K-12 students in the United States received one or more out-of-school suspensions (OSS). This statistic translates to approximately six out of every 100 students receiving at least one OSS, a 20% decrease compared to the 2011-2013 school year (Department of Education, 2016). A review of the data snapshot released by the Office of Civil Rights demonstrates that on a national level, serious gaps exist in the rates at which students of colour (e.g., Black and Latino students) are disciplined compared to the majority group (i.e., White students). For example: (a) Black preschool students are 3.6 times more likely to receive one or more OSS compared to White students and (b) the percentage of Black boys that received an OSS was 18%, while 6% of students from the entire K – 12 student population received at least one OSS (Department of Education Office for Civil Rights, 2016). These findings describe what is known as the *discipline gap*, in which students from diverse backgrounds (e.g., Black, Latino, and American Indian) are subjected to differential and disproportionate rates of exclusionary discipline (Gregory, Skiba, & Noguera, 2010). The discipline gap is like the achievement gap, which describes the disparities along racial and ethnic lines found in academic outcomes. For example, in 2007, the rates for both discipline and grade retention were higher for Black students than White students. Twenty-one percent of Black students had been retained versus 9% of White students, and 13% of Black students had been expelled versus 1% of White students (KewelRamani, 2010). Given the potential link between the two, it is important to evaluate the less researched discipline gap to address the achievement gap (Gregory et al., 2010). Morris and Perry (2016) provide empirical evidence linking racial discrepancies in OSS (i.e., the discipline gap) to racial discrepancies in academic outcomes (i.e.,

the achievement gap). Using hierarchical modeling, Morris and Perry analyzed a longitudinal dataset to quantify the impact of disproportionate exclusionary practices on achievement levels. Consistent with the literature, they found that Black students were at greater risk for receiving an OSS compared to White students, even when controlling for other factors such as gender and socioeconomic status (SES). Mediation analyses revealed that inequalities in exclusionary discipline experiences explain 20% of the achievement gap in reading between Black and White students, providing support that the achievement gap may be a by-product of the discipline gap.

It should be noted that unlike in Canada, there are federal regulations in the United States which compel states to publish data on student academic outcomes, and in some instances, on student discipline, too. A recent report released by Ontario Ministry of Education includes a recommendation to mandate the collection of student demographic data, such as students' race and religion (Quan, 2017), however, at this time, a national dataset on Canadian student outcomes in education is yet to be made available. While most of the studies cited in this study rely on U.S. data, equity issues in education do exist in Canada. For example, in Canada's largest school board, students from minority backgrounds made up the most number of students who did not graduate (Toronto District School Board, 2010). Also, according to the Auditor General of British Columbia, students from Aboriginal backgrounds have lower scores on standardized testing, lower graduation rates, and higher rates of special education designations, when compared to their non-Aboriginal peers (Bellringer, 2015). Specifically, in the 2015 – 2016 school year, only 49% of students who identified as being Aboriginal received a typical high school graduation certificate (i.e., Dogwood Diploma), while 73% of non-Aboriginals graduated with a Dogwood (B.C. Ministry of Education, 2016). Therefore, this study relies on U.S. data solely due to its availability. Despite that these data are specific to the U.S., the results from this

study may be of benefit to researches and practitioners both within and outside of the U.S. as the method of calculating racial disproportionality may be done in any context.

In the following section, I outline the history of the discipline gap, negative effects of exclusionary school discipline, and disproportionate rates of minority groups in the juvenile justice system. I provide a critical perspective on some of the methods that are used to measure disproportionate rates of discipline. I then outline my study to evaluate the metrics used to measure the discipline gap.

Chapter 2: Literature Review

The Discipline Gap

Exclusionary discipline causes a student to be removed from their classroom and miss instructional time, and is used with the intention to deter behaviours that violate a school's code of conduct (Noltemeyer & McLoughlin, 2010). In general, expulsion is when a student is permanently removed from the school and suspension is when a student is barred from attendance for a period of 10 days or less, sometimes with students still responsible for missed work without any special provisions for accommodations (Arcia, 2006; Noltemeyer & McLoughlin, 2010). Some exclusionary discipline of students with disabilities that lasts longer than 10 consecutive days, or more than 10 days in a year, triggers specific consequences (e.g., procedural reviews, manifestation determination) for both students, educators, and administrators in the U.S. (U. S. Department of Education, n.d.). Out-of-school suspensions are used as punishment for the offending students in order to reduce future misbehaviour, despite the lack of evidence to support this reduction (Noltemeyer, Marie, McLoughlin, & Vanderwood, 2015).

The discipline gap is not a new phenomenon. For more than 50 years, there has been documentation of over-representation of culturally diverse students (e.g., Black students) in exclusionary school disciplinary actions. Although a 1954 landmark Supreme Court case mandated that all students be provided a fair education that is free from discrimination (*Brown v. Board of Education of Topeka*, 1954), the public education system has not caught up to the intent of this decree. Almost 20 years later, the Children's Defense Fund (1975) analyzed data released by the Office for Civil Rights. They analyzed suspension data for 2,862 school districts which included more than 24 million students. A specific section entitled "Racial Discrimination in School Suspension" (Children's Defense Fund, 1975, p. 12) contains early empirical support

for the existence of the discipline gap. Specifically, the researchers found that Black students accounted for only 27.1% of the students registered in school districts, however, they accounted for 42.3% of students suspended. Suspension rates were two and three times higher for Black students than White students. Furthermore, more than two-thirds of the school districts included in the sample were found to have higher rates of suspension for Black students compared to White students (Children's Defense Fund, 1975). Since then, the number of studies examining the discipline gap has increased and researchers have repeatedly found that Black students are disciplined at significantly higher rates than some of their peers (Noltemeyer & McLoughlin, 2010; Reschly, 1997; Skiba et al., 2011; Wallace, Goodkind, Wallace, & Bachman, 2008).

Long Term Effects of Exclusionary Discipline

The discipline gap, as one of the most documented racial inequities that exists, places students of colour at risk for a myriad of negative outcomes (Skiba, Shure, & Williams, 2012). Although there is widespread use of exclusionary disciplinary practices in schools, there is limited data to support the effectiveness of such practices (Arcia, 2006; Christle, Nelson, & Jolivette, 2004). For example, Townsend (2000) posits that barring Black students access to an education by means of exclusionary discipline can create a domino effect. The end school-related result of this domino effect is that that Black students are at higher risk for drop out. An indirect effect of receiving an OSS is the poor socializing that may occur when students spend time on the streets, which may allow these students to engage in illegal activities. In other words, when Black students are not afforded an opportunity to learn because of an OSS, the missed learning opportunity may lead to lower achievement levels compared to White students. Additionally, students with tendencies of receiving OSSs may be subject to lower academic expectations from their teachers, remedial programming, and be retained at higher rates due to low achievement.

Townsend also theorizes that the disproportionate rate at which Black students are disciplined may lead Black students to perceive themselves as being unable to abide by rules, and therefore, may instill in them overarching negative attitudes towards abiding to rules. Other consequences of OSSs include students feeling lower levels of school connectedness and a higher likelihood of engaging in illegal activities (Gregory et al., 2010).

According to DeRidder (1991), receiving an OSS may hasten the dropout process for a student, especially considering that receiving an OSS is rated as one of the top three school-specific reasons for dropping out. This issue is further exacerbated when students choose to dropout even when given the option to choose between withdrawal and receiving an OSS (DeRidder, 1991). Even if students return to school after experiencing an OSS, they may no longer have a positive relationship with school personnel as a result of feeling mistreated and unfairly punished (Wolf & Kupchik, 2016). In relation to the student-school relationship, Noltemeyer et al. (2015) conducted a meta-analysis of 34 studies. They found that there was a statistically significant and inverse relation between school exclusionary practices, such as OSSs, and academic achievement. They found a positive relation between exclusionary practices and dropout. Overall, they conclude that the use of suspension hinders student engagement, which can lead to dropout and lower achievement levels.

Arcia (2006) analyzed longitudinal data to study the correlation between OSS and achievement scores and dropout rates. Arcia found that the number of students suspended increased as grade levels increased. These students also had lower achievement scores compared to non-suspended students at baseline, as well as lower gains in achievement scores in later years. Finally, Arcia found that as the number of OSS received increased, the percentage of students who had later dropped out also increased. Specifically, 21% of students who received an

OSS that lasted between one and 10 days had dropped out, whereas, 43% of students who received an OSS that lasted for 21 or more days dropped out. This study shows that the risk of for a negative outcome (e.g., dropping out) increases as the duration of the exclusionary discipline increases.

While studies have mainly addressed the outcomes of suspension at the student level, Lee, Cornell, Gregory, and Fan (2011) conducted hierarchical regression analyses at the school level to understand if there are differential impacts of schoolwide suspension rates on Black and White students' dropout rates. In a sample of 289 Virginia public high schools, they found a positive correlation between high suspension rates and dropout rates. Specifically, schools that suspended 22% of their student body had a drop out rate that was 56% greater than schools that suspended only 9% of its students. Furthermore, they found that student body composition and dropout rate were inversely related for Black students, independent of suspension rates.

Consistent with the findings of other studies (e.g., Gregory et al., 2010), Lee et al. (2011) suggest that Black students may feel less connected to their schools if there is a lower percentage of Black students that make up the student body composition. This study provides another example of an adverse outcome (e.g., dropping out) that is correlated with the use of OSSs.

Christle et al. (2004) examined the difference between high and low suspending schools and their characteristics, as well as student outcomes (e.g., achievement), in 161 middle schools in Kentucky. The following school characteristics were positively correlated with suspension rate: student violation of school and/or district policies, SES, law violation, retention rate, and dropout rate. School attendance, academic achievement, and White student body composition were negatively associated with suspension rates (Christle et al., 2004). These finding imply that schools with higher White-student body composition, attendance rates, and scores on

standardized testing, have lower suspension rates. In contrast, schools with a higher percentage of students from low SES, a higher number of students committing school and/or district policy violations, and a higher number of students dropping out, have higher suspension rates. The generalizability of this study is limited due to the small sample size.

While most studies have examined academic outcomes (e.g., likelihood of dropping out of school), Wolf and Kupchik (2016) studied non-academic outcomes associated with exclusionary discipline. They examined how school suspension impacted the likelihood of adverse events occurring in adulthood. They hypothesized that since exclusionary discipline has an immediate consequence on students, such as a student being labeled deviant, consequently, there would be negative outcomes experienced in adulthood. Based on their analysis of a nationally representative longitudinal dataset, they found that experiencing a suspension increases the likelihood of the individual becoming a crime victim, engaging in illegal activities, and becoming incarcerated, even when the researchers controlled for student- and school-level variables. Additionally, Black adults who have received an OSS as students were significantly more likely to experience victimization, criminal activity, and incarceration.

School-to-prison Pipeline

As outlined earlier, an outcome of school exclusionary practices that has been researched is the later involvement of disciplined students in illegal activities. Kim, Losen, and Hewitt (2010) provide a nuanced overview of the school-to-prison pipeline. Overall, the school-to-prison pipeline describes the convergence between the public education and the juvenile justice systems. It is a phenomenon by which students, through subpar educational services, are set up for failure, and thus, these children become disengaged and are more likely to drop out, and ultimately get involved with the juvenile justice system. Consistent with the discipline gap, racial

and ethnically diverse students may also be disproportionately impacted by the school-to-prison pipeline phenomenon. At the end of the pipeline (i.e., contact with the justice system), Black individuals are represented at disproportionate rates relative to the population composition (Nicholson-Crotty, Birchmeier, & Valentine, 2009).

In 2012, the number of violent crimes committed by juveniles (i.e., under the age of 18) were “at historic lows” (Puzzanchera, 2014, p. 4), however, upon further evaluation, the potential effects of the school-to-prison pipeline become more evident. Comparing the proportion of Black youth arrested from 2003 to 2012 reveals that for almost every single crime category in which data are reported for both years, the proportion has increased or stayed the same, except for drug abuse violations which saw a decrease from 26% in 2003 to 23% in 2012. Additionally, in 2012, the racial composition of the national juvenile population of children aged 10-12 was 76% White and 17% Black. For every single offence type, the Black proportion of juvenile arrests far exceeded the national composition of non-incarcerated Black juveniles, except for violations related to liquor laws (Puzzanchera, 2014; Snyder, 2005). See Table 1 for a detailed overview of the Black proportion of juvenile arrests for each offence type.

According to Kim et al. (2010), given the culture of high-accountability and high-stakes testing that exists because of the No Child Left Behind Act (NCLB), some schools may inadvertently compel students to drop out (i.e., be more inclined to suspend/expel offending students) to better the schools’ testing results. The NCLB Act will be replaced by the Every Student Succeed Act (ESSA) for the 2017 – 18 school year. Under the new Act, states must publish data on at least one state-selected indicator of school quality (e.g., suspension rates; Department of Education, n.d.). This new flexibility may lessen the burden on states that required them to ensure that all students were proficient across academic areas (as was required

in NCLB), and instead start using other data points (e.g., suspension rates) to better inform practice and policy. Also, as schools previously adopted zero-tolerance policies and consequently increased their exclusionary disciplinary practices, they also began to integrate the juvenile justice system within their buildings. Schools relied on police officers to respond to student misbehaviour, which meant that students not only may have faced an academic consequence to their behaviour—being suspended or expelled—but also an encounter with the juvenile justice system and faced legal consequences such as arrests (Wolf & Kupchik, 2016). These policies were reinforced by federal funding. For example, the Violent Crime Control and Enforcement Act of 1994 contained provisions to increase police presence in schools through community policing (Curry, 1997).

Direct evidence in support of the school-to-prison pipeline is scarce, however, given the similitude of the disproportionalities that exist within both systems, the relation between them is assumed without explicit empirical support (Nicholson-Crotty et al., 2009). Although there are theoretical mechanisms (e.g., labelling theory, self-fulfilling prophecy; Monroe, 2005; Nicholson-Crotty et al., 2009) that may potentially provide an explanation for the phenomenon, these do not appear to have been tested. For instance, Black students who are disciplined at disproportionate rates compared to their White peers, may cause others to perceive Black students as being more delinquent to the extent that they would be labelled as delinquent. In turn, this may cause these students to self-internalize this delinquent label and therefore actually become delinquent at disproportionately higher rates both while within the school system and even after they are no longer in the system. Another hypothesis proposed by Nicholson-Crotty et al. (2009) is a logical deduction that, since Black students experience an OSS at disproportionate rates, they may likely be spending more time on the streets and thus be more susceptible to

coming into contact with the police. To study the empirical link between school discipline and contact with the juvenile justice system, Nicholson-Crotty et al. (2009) conducted a multivariate analysis on the influence of several factors on the referral rates for Black and White youth, using a cross-sectional time-series design. They evaluated whether disproportionate discipline in schools predicted disproportionate contact with the juvenile justice system, within the same counties. Their results indicate that, even after controlling for environmental factors, in counties in which there is a discipline gap, there is also a higher rate of contact for Black youth with the juvenile justice system. Specifically, in a multivariate analysis, 39% of the variance seen in the differential referral rates for Black youth to the juvenile justice system could be explained by the following factors: relative risk for OSS, population density, Black/White poverty differential, and percentage of Black citizens employed.

Factors Contributing to Discipline Gap

Since the use of zero-tolerance approaches in the 90s, there has been an increase in exclusionary discipline (Wald & Losen, 2003). According to the American Psychological Association, zero-tolerance policies are “the application of predetermined consequences, most often severe and punitive in nature, that are intended to be applied regardless of the gravity of behavior, mitigating circumstances, or situational context” (2008, p. 852). The purpose of such policies is intended to reduce bias from disciplinary decisions and thus decrease the disciplinary gap, however, schools are actually less safe (Skiba & Peterson, 2000), and as a result, the desired outcomes are not achieved (American Psychological Association, 2008). Furthermore, zero-tolerance policies may be most affecting a group of students who are in most need of improvement as a consequence of increased rates of OSSs, which may further widen the achievement gap (Gregory et al., 2010).

While factors such as SES, racial and ethnic background, and locality are correlated, it is difficult to delineate each factor to understand its unique contribution to the discipline gap. More research is needed to get a better understanding of how other factors, such as stress levels affect student behaviour, and in turn contribute to the discipline gap (Gregory et al., 2010).

Nonetheless, studies have attempted to evaluate the influence of some of these factors on the discipline gap. Gregory et al. (2010) argue that low SES and locality (e.g., residing in high-crime, high-poverty neighbourhood) are sometimes wrongly used to justify the existence of the discipline gap (Gregory et al., 2010). Wallace et al. (2008) studied the effects of SES on the discipline gap and found that controlling for these factors only slightly reduces their effect on the discipline gap, however, these factors remain significant predictors nonetheless.

The existence of the discipline gap begs an obvious question – are Black students indeed misbehaving and breaking school rules at higher rates compared to their White peers, or are there biases (e.g., systematic discrimination) affecting teachers’ and administrators’ abilities to administer school punishment in an equitable manner? A recent study by Smolkowski, Girvan, McIntosh, Nese, and Horner (2016) evaluated a framework by which disproportionality is more likely to occur at specific times in specific places. They found that throughout the course of a school day, there are specific times and settings when teachers are more likely to be implicitly biased and discipline students disproportionately. For example, Black students were at a greater risk for being disciplined in the classroom when behaviours were deemed to be subjective (i.e., there is room for judgement).

Monroe (2005) believes that since schools reflect current societal norms, the discipline gap may be because of factors that are external to schools (e.g., the negative portrayal of Blacks as violent in the media). Relating these societal factors to the discipline gap, Monroe explains

that although teachers may not make conscious efforts to connect their disciplinary practices to negative perceptions of Black students, they may still be guided by racist stereotypes that cause them to see Black students as troublesome and threatening.

Methodological Issues in Identifying the Discipline Gap

As evidenced from the studies reviewed above, the existence of the discipline gap has been replicated many times. Despite being able to reliably detect this phenomenon in multiple studies, there are still methodological issues in identifying the discipline gap. Specifically, there is a lack of consistency within the literature, as well on a policy level, on how the discipline gap is measured and reported. The following will provide an overview of such differences in methodologies across areas of practice and research.

First, researchers have relied on a variety of data sources that have been used in studies of the discipline gap. For example, some have relied on self-report questionnaires from 8th, 10th, and 12th graders about school discipline (e.g., Wallace et al., 2008), while others have used school disciplinary records such as office discipline referrals (ODRs; Skiba et al., 2011; Skiba, Michael, Nardo, & Peterson, 2002) and publicly available data of state discipline counts (Noltemeyer & McLoughlin, 2010).

Second, the statistical analyses used to evaluate the presence of the discipline gap have also differed between researchers. While logistic regression has been used by some (e.g., Skiba et al., 2011; Wallace et al., 2008), the validity of such analyses should be cautioned, given the nature of the data sources. Since logistic regression analyses assume there is an independence of cases (i.e., no clustering), this assumption is not met with the data sources typically used to evaluate the discipline gap. Therefore, the use mixed modelling may be a more appropriate analysis instead (Peugh, 2010). It appears that only recently have researchers considered the

nested (i.e., clustered) nature of the data (e.g., Greflund, McIntosh, Mercer, & May, 2014; Smolkowski et al., 2016). When measuring changes of the discipline gap over time, some have studied the changes of marginal means of OSSs (Noltemeyer & McLoughlin, 2010) and the changes of prevalence rates for OSSs (Wallace et. al., 2008). This issue has also been studied by the non-scientific community as seen in a recent, non-peer reviewed entry by the Seattle Times (Seattle Times, n.d.).

Third, and perhaps of most significance, the metrics used by researchers to measure the discipline gap are wide ranging. Some commonly used metrics include composition indices, risk ratios, and odds ratios. A composition index (Gregory et al., 2010) compares the percentage of students that belong to a certain group that meet a certain criterion (e.g., Black students that have received at least one OSS) with the proportion that group represents in the total population (e.g., total Black student enrollment as a percentage of all students enrolled). For example, by using a composition index, Smith and Harper (2015) found that while Black students were 34% of the entire student body across Alabama, Black students accounted for 64% of school suspensions. Although the composition index is intuitive to understand, critiques in both research and practice for using a composition index for measuring the discipline gap exists. For example, a composition index is unable to detect if a discrepancy is significant. Also, a composition index does not allow for comparisons to be made, such as across school districts, since the composition of the populations may not be the same (Dyches & Prater, 2010; Gibb & Skiba, 2008). The issue of determining if differences in exclusionary rates between groups are significant is pertinent when using metrics such as composition indices. While these metrics do not address the precision of the disproportionalities (i.e., do not provide a measure of the confidence interval around the values calculated), some do provide a measure of how substantial the

disproportionalities are. For this thesis, the terms significant and substantial are used interchangeably to refer to a difference in discipline outcomes between groups that are meaningful and inequitable. Although some (e.g., Chinn & Hughes, 1987) have set parameters for determining if discrepancies between composition indices are significant, these seem to be arbitrary, seldom used, and lacking empirical support.

Another commonly used metric is the risk index. A risk index “is the proportion of a group that is at risk of a particular outcome” (Boneshefski & Runge, 2014, p. 151). To calculate the risk index, the proportion of students from a group that has met a certain condition (e.g., received one OSS) is divided by the total number of students of the same group. For example, if 20 Black students received an OSS in a school that has 200 Black students, then the risk index for Black students to receive one OSS at that school would be 20 divided by 200, yielding 10%. This information alone is limited for interpretational purposes. As with the composition index, a risk index does not detect the presence of a substantial disproportionality, nor allow for the comparison between groups. To address the latter, a risk ratio is calculated, which allows the comparison of the risk between target and comparison groups. Therefore, if at that same school in the previous example we were interested in knowing the relative risk between Black and White students to receive an OSS, then the risk index for White students would be calculated. If in that school there were a total of 250 White students, 10 of whom have received an OSS, then the risk index for White students would be 10 divided by 250, yielding 4%. Therefore, the risk ratio would be 10% divided by 4%, yielding 2.5. A risk ratio of less than 1 indicates an underrepresentation and greater than 1 indicates an overrepresentation (Boneshefski & Runge, 2014), however, this broad classification of risk ratio values does not provide a measure of statistical significance. The Department of Education has provided example threshold risk ratio

values for determining significant disproportionality. The threshold values are based on the two median absolute deviations above the national median of the local educational agencies. For total removals (i.e., including both in-school and out-of-school suspension), the example threshold value is 1.873 (Office of Special Education and Rehabilitation Services, 2016).

Some researchers have calculated odds ratios from logistic regression analyses on discipline data. For example, Wallace et al. (2008) used logistic regression to compute odd ratios. They found that Black boys were 3.3 more times more likely to be suspended or expelled than White boys. Davies, Crombie, and Tavakoli (1998) present a series of arguments against the use of odd ratios, specifically, that they may be misleading if interpreted as risk ratios. Some (e.g., Skiba et al., 2011) have argued for the use of odds ratios, since odds ratios may be a more accurate measure of disproportionality since both occurrences and non-occurrences are considered.

In practice, there are differences in the way states calculate the discipline gap. For example, the State of Washington uses composition indices (i.e., the percentage of students from a specific racial/ethnic group who have received an OSS; Bollmer, Bethel, Munk, & Bitterman, 2014) and does not use risk ratios to focus more on within-group characteristics than to make comparisons between-groups (Office of Superintendent of Public Instruction, n.d.). To the contrary of that state's position on using risk ratios, there appear to be many more states that use a form of a risk ratio (e.g., weighted, alternative, difference) to measure disproportionality.

The lack of consistency across states in how the discipline gap is measured makes reporting on nationwide trends difficult. Researchers have used different data sources to calculate different metrics using different analyses when studying the discipline gap. Finally, the ambiguity in the academic literature on exactly what constitutes significant disproportionality is

also manifest at the policy level. States self-define the threshold for what would be deemed substantial disproportionality, and consequently decide how to report on disproportionalities. Another issue with having states self-define thresholds for substantial disproportionality is the research indicating that the reported rates of local education agencies (LEA; e.g., school districts) identified as having substantial disproportionality are lower than what would be expected given rates of disciplinary actions against racial and ethnic groups (Office of Special Education and Rehabilitative Services & Department of Education, 2016). Current federal law has consequences for states found to have LEAs with significant disproportionalities. One consequence is states being forced to redirect funding to provide additional academic and/or behavioural supports to serve those children that are disproportionately affected within the LEA (U. S. Department of Education, 2008). However, as the laws currently stand, there is no direction on how the discipline gap is to be measured.

The U.S. federal government has proposed some changes to address the ambiguity on how the discipline gap should be measured. The proposed changes would create a standard method that states must use to determine if significant disproportionality based on race and ethnicity is occurring across multiple metrics, including suspension and expulsions. Such methodology would involve using forms of risk ratios.

Using Risk Ratios to Identify Significant Disproportionalities

As discussed, there are methodological differences in the way that researchers have studied and evaluated the discipline gap (e.g., using different metrics), which is also seen at a policy level. The proposed changes introduced by the federal government may be a step in the right direction to help create some consistencies for more accurate comparisons of the discipline

gap between states, and on a national level. The following section will elaborate on these proposed changes in more detail.

A risk ratio requires the comparison of the relative risk of a target group (e.g., Black students who have received at least one OSS) to a comparison group (e.g., all students who have received at least one OSS). Although there is a lack of consensus within the literature about which comparison group should be used, Boneshefski and Runge (2014) argue that the contextual factors (e.g., overall representation of Black students in the study body composition) need to be considered, in order to yield risk ratios that better depict the over- and/or under-representation in disproportionality. The proposed federal legislation would use all other students as a comparison group, instead of White students or all students. All other students would mean that the rate for the target group is removed from the comparison group, unlike what is done when *all students* are used as the comparison group.

The proposed legislation would allow states to calculate alternate risk ratios, which would be used if an LEA has less than 10 students in the comparison group (i.e., in the all other students group) or if the comparison group has a risk index – as defined earlier – of zero (Office of Special Education and Rehabilitative Services & Department of Education, 2016). If either of those conditions were met, then states would calculate an alternate risk ratio using all other racial and ethnic groups within the state, as opposed to the LEA.

Overall, the inconsistency in the ways in which the discipline gap has been studied is of concern. Primarily, the lack of consistency may make it difficult to accurately model changes to the discipline gap, on all levels (i.e., federal, state, or district). The differences in the methodologies used to study the discipline gap may be of detriment on studies evaluating the effectiveness of policies or interventions for narrowing the discipline gap. Also, the lack of

research on the use of different comparison groups (i.e., comparing the risk for Black students compared to their White peers versus compared to all their peers at the school, district, and/or state levels) has major implications for practice. Effectiveness studies of efforts (e.g., policies, interventions) put in place to narrow the gap may lead to different outcomes, depending on which group is used. Additionally, since current legislation mandates that states reallocate their budget should LEAs be found to be significantly disproportionate, then depending on which metric is used, such consequences may be falsely triggered. Considering that some studies have shown that the discipline gap is narrowing, albeit remaining substantial, perhaps using different comparison groups may yield different results.

Proposed Study

Because the phenomenon of the discipline gap has been reliably found, the issue is no longer *is there a discipline gap?* Current research has focused on ways of intervening (e.g., Smolkowski et al., 2016), however, there does not appear to be sufficient research on the methodologies used to measure the discipline gap, and therefore, intervention studies may not yield accurate results since the underlying metrics lack sufficient validation. Given the limited research on how the discipline gap has been measured, this study moves beyond conceptualizing it to addressing methodological issues in measuring it. The use of risk ratios as a metric is evaluated, and therefore, a better understanding of the consequences of using such a metric is gained. This study can be a precursor to future research that can look at evaluating the validity of using risk ratios for evaluating the discipline gap. Potentially, future efficacy and intervention studies that address the discipline gap can rely on better metrics to reach more accurate conclusions about the interventions in question. Finally, this study adds to the literature a recent study of the change of the discipline gap across three years (i.e., is school discipline becoming

more equitable?), as well as an understanding of the implications of using different comparison groups in the likelihood that a school is found with significant risk ratio values. Specifically, the research questions that are answered are as follows:

RQ 1: Are there differences in risk ratios for Black students receiving one or more OSS, based on the comparison group (Group 0 = White students or Group 1= all other students) used?

RQ 2: Are there differences in classifications (i.e., percentage of schools identified as being disproportionate) based on the comparison group used?

RQ 3: From 2012 – 2014, are there statistically significant differences between the yearly mean risk ratio values, and if so, are these differences dependent on the comparison group used to calculate risk ratios?

Chapter 3: Methods

Participants and Settings

Extant data for 5,422 schools in 50 states across three years (i.e., 2012 – 2013, 2013 – 2014, and 2014 – 2015) were in the initial dataset. The average enrollment per school, per year, was 547 students from kindergarten to grade 12. Inclusionary criteria were: (a) office discipline referrals (ODRs) must have been available for the 2012-2013, 2013-2014, or 2014-2015 academic years; (b) enrollment data must have been available from the National Center for Education Statistics (NCES); (c) at least 80% of the ODRs for each school must have had recorded racial or ethnicity group coded; and (d) a minimum of 10 students must have been identified in each racial/ethnic category as having had at least one OSS. The last rule was included to increase the reliability of the risk ratios that are calculated. (Bollmer et al., 2014). Overall, 594 unique schools met the inclusionary criteria within 288 unique districts and 36 unique states. Data regarding the sample size that was used in the analysis after removing the schools that did not meet the inclusionary criteria are presented in Table 2. A school was defined to be missing if data for that school were available for at least one school year but not all three years.

Measures

An office discipline referral (ODR) event is captured when a student violates his or her school's code of conduct, in which the misbehaviour was seen by some school personnel, with the end result is the creation of a record describing the incident (Sugai, Sprague, Horner, & Walker, 2000). An ODR can be captured and entered for a wide array of pre-defined infractions. Similarly, the possible outcomes (i.e., disciplinary action) taken is also limited to a list of pre-determined consequences (e.g., expulsion). See appendix A for a list of both possible infractions

and outcomes. For this study, ODRs with out-of-school suspension (OSS) as the outcome were included. As per Bonesheski and Runge (2014), a count of unique numbers of students in each racial and ethnic group (i.e., White and Black) receiving at least one OSS was included. Student race and ethnicity were operationally defined as what was entered in the data management system for each ODR. Consistent with federal regulations (U.S. Department of Education, 2008), the data management system that was used to collect ODRs requires that student ethnicity to be captured through self-identification—observer identification may be used if a student fails to self-identify—in only one of the following two categories for ethnicity: (a) Not Hispanic/Latino, or (b) Hispanic/Latino. Students must also identify in one or more of the following categories for race: (a) American Indian/Alaskan Native, (b) Asian, (c) Black, (d) Pacific Islander/Native Hawaiian, and (e) White. For the purposes of this study, ODRs were categorized as originating for a Black and White student if the student had been identified as only Black or only White (i.e., no other racial groups were selected), respectively. In both cases, the student must be identified as Not Hispanic/Latino in the ethnic category. This categorization method has some major limitations which are discussed later.

Procedures

One data management system widely used by schools recording ODRs is the School-Wide Information System (SWIS), which is a web-based application that allows school personnel to enter ODRs (May et. al., 2013). Based on an analysis of the schools included in the SWIS dataset between 2005 and 2006, Spaulding et al. (2010) found that the schools in the SWIS dataset was generally representative of schools on a national level (i.e., schools that may not be using SWIS).

To determine a school's enrollment by ethnic and racial groups, data from the National Center for Education Statistics (NCES; www.nces.ed.gov) were used. The NCES collects data on all public schools in the U.S. The NCES provides enrollment for six possible racial/ethnic groups (American Indian/Alaskan, Asian/Pacific Islander, Black, Hispanic, White, and Two or More Races). Although there are also five possible racial classifications of students in SWIS, they do differ from that of the groups in NCES. In SWIS, Asian and Pacific Islander are unique categories. This difference in categorization of students' racial and/or ethnic backgrounds will not affect the outcome of this study since only Black and White students' data will be analyzed. Schools' ODR data and enrollment by ethnic and racial groups were merged using each school's unique NCES number.

Risk indices for Black and White students. To calculate these indices, the number of unique students in each racial group who have received at least one OSS was summed per academic year, and divided by the total number of students enrolled in each group, separately, which yielded two risk indices per year. These indices were calculated using observed data.

Risk index for all other students. An additional risk index was calculated for all other students (excluding Black students) using the observed data. This risk index was calculated by using the number of unique ODRs excluding Black students, divided by the total number of students (excluding Black students) enrolled. This was done for each academic year.

Risk ratios. Risk ratios were calculated for the following: (a) the risk for Black students to receive one OSS compared to the risk for all other students (excluding Black students); and (b)

the risk for Black students to receive one OSS compared to the risk for White students to receive an OSS, for each year.

Determining significant disproportionalities. Risk ratios calculated were compared with the U.S. Department of Education’s proposed risk ratio threshold for determining disproportionality, which is equal to 1.873 for total removals (Office of Special Education and Rehabilitation Services, 2016). This threshold risk ratio value was calculated by taking two median absolute deviations for total removals (i.e., in-school and out-of-school suspensions for any duration, as well as expulsions) above the national median of local educational agencies.

Data Analysis Plan

Since the data are nested (i.e., students in schools, schools in districts, and districts in states), mixed modeling may be appropriate (Peugh, 2010) to take into consideration potential non-independence of the residuals due to clustering. Intraclass correlation (ICC) values were calculated to confirm that the data structure violated the independence assumption (Peugh, 2010).

R 3.4.0 (R Core Team, 2017) was used to fit two random intercept models with the following factors: (a) Comparison Group¹, (b) Time, and (c) Comparison Group*Time (an interaction effect). Comparison Group was dummy coded (0 = White students and 1 = all other students) depending on the comparison groups used in the risk ratio calculations. Time was represented as a categorical variable with three levels (0 = 2012 – 2013, 1 = 2013 – 2014, and 2 = 2014 – 2015). Since this study did not evaluate all possible levels of time (e.g., years before 2012), Time was analyzed as a random factor. This implies that the interpretation of the results would be limited to only the three years sampled in the data, and not across time in general. Comparison Group

¹ Note that Comparison Group (i.e., capitalized) refers to the name of the factor in the models that were fit.

was analyzed as a fixed factor given that both possible comparison groups that could be used were analyzed (i.e., all possible comparison groups that are present in the population are sampled in the data). The interaction term was analyzed as a random factor, due to the inclusion of the random factor of Comparison Group. The nlme package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017) was used to fit the first series of models (i.e., Models 1.x), with risk ratios as the criterion variable. The lme4 package (Bates, Maechler, Bolker, & Walker, 2015) was used to fit the second series of models (i.e., Models 2.x) with disproportionate status as the criterion variable. A logit link was used in Models 2.x as the criterion variable (i.e., disproportionate status) was dichotomously coded (0 = Not disproportionate, 1 = Disproportionate). Two risk ratios were calculated per school, per year, each with a different comparison group. Results from the first series of models that were fit were used answer RQ 1. The statistical significance of the beta for Comparison Group was evaluated. The second series of models was used to answer RQ 2. The statistical significance of the beta for Comparison Group was evaluated. This helped determine if the calculation method (i.e., using all other students or White students as the comparison group) had a significant impact on the number of schools that were identified as disproportionate. Descriptive statistics (e.g., percentage of schools that are identified depending on the comparison group used) were also provided. Results from both series of models addressed RQ 3. Specifically, the random effect of Time and the interaction terms were evaluated in the final model selected in each series to answer if yearly mean risk ratio values are statistically different from 2012 – 2014, and if such a difference would be dependent on the use of a specific comparison group when calculating risk ratios.

Chapter 4: Results

Descriptive Statistics

Table 3 contains the percentage of schools, based on the observed data, that had risk ratios for Black students to receive at least one OSS, that exceed the example threshold of 1.873, by comparison group. When using White students as the comparison group, the percentage of schools that were identified as exceeding the threshold is 93%, 94%, and 95%, for 2012, 2013, and 2014, respectively. When using all other students as the comparison group, the percentage of schools that were identified as exceeding the threshold is 96%, 95%, and 96%, for 2012, 2013, and 2014, respectively.

Intraclass correlation (ICC) values were calculated by computing the variance at each level (i.e., school, district, and state) and comparing it to the total variance in an empty model (i.e., model with only random intercepts at each level) to confirm that the use of mixed models would be appropriate.. The ICC values at the school, district, and state levels are 0.388, 0.197, and 0.186, respectively. These values indicate that some of the variance between-levels is shared with the variance within-levels, and therefore, the use of mixed models are warranted. The ICC values are presented in Table 4.

Because the outcome variable in the first model is a continuous variable, its normality was assessed to check the assumption of normality when using general linear models. The kurtosis and skewness of the variable (i.e., risk ratios) indicate that variable is asymmetrically distributed and heavy-tailed. A log transformation was conducted to correct the skewness and kurtosis of the variable. See Table 5 for information about the skewness and kurtosis pre- and post- transformation. See Figure 1 for a distribution plot of the variable pre- and post-transformation.

As illustrated in Figures 2 and 3, based on the observed data, the change of the risk ratios and log risk ratios for Black students to receive at least one out-of-school suspension from 2012 – 2013 to 2014 – 2015 increased, regardless of the comparison group used. See Table 6 for the mean risk ratios and log risk ratios for Black students to receive at least one out-of-school suspension.

Model Fitting: Risk Ratios as Outcome Variable

Models were fit using maximum likelihood estimation. First, an intercept-only model was fit with log risk ratios as the outcome variable. This model (i.e., Model 1.0) was used a baseline model to compare goodness-of-fit with other models. Subsequent models (e.g., Models 1.1 – 1.3) were fit by systematically adding estimates and comparing model goodness-of-fit by conducting a log-likelihood ratio test. See Table 7 for a summary of the estimates for the intercept-only models. Table 8 contains the results of the log-likelihood tests between subsequent intercept-only models. By comparing the log-likelihood ratios as well as AIC and BIC values of subsequent models created, Model 1.3 was found to have the best fit (AIC = 1321.98 and BIC = 1349.85) compared to the baseline model (AIC = 7418.38 and BIC = 7429.53). Model 1.3 was statistically significantly different ($p < .001$) than the previous model (i.e., Model 1.2).

After establishing that the best-fitting random intercept model is a model with random intercepts at all three levels (i.e., school, district, and state), explanatory variables were systematically added and model goodness-of-fit was evaluated. First, the fixed effect of Comparison Group was added (i.e., Model 1.4). then the random effect Time (i.e., Model 1.5), and finally an interaction, Comparison Group*Time (i.e., Model 1.6). A summary of the model results with the systematic addition of the explanatory variables and the results of comparing model goodness-of-fit and statistical significant difference testing are presented in Tables 9 and

10 respectively. Given that the models include a log-transformed outcome variable, the model estimates are therefore also on the log scale. The exponentiated values are presented to ease interpretability of the values. Adding the random effect of Comparison Group*Time (i.e., Model 1.6) did not yield a model that was significantly different ($p = .620$) than the model without the interaction term (i.e., Model 1.5). Although Model 1.5 is more parsimonious, Model 1.6 was selected as model to be used as a baseline when adding new estimates, since it allows for the evaluation of the interaction between Comparison Group and Time, which is required to answer one of the research questions of this study.

Random slopes were systematically added to the previously selected best-fitting model (i.e., Model 1.6) to allow for the rate of change of the random effect of Time to vary at each level. See Table 11 for a summary of the goodness-of-fit for the subsequent models that were fit. Adding a random slope of Time at the school level (i.e., Model 1.7) yielded a model that was significantly different than the model without a random slope (i.e., Model 1.6). A model did not converge when random slopes were subsequently added at the district level alone (i.e., Model 1.8), and then district and state levels (i.e., Model 1.9). Convergence was achieved when models were fit with a random slope of Time only at the district level (Model 1.10; AIC = 1282.568, BIC = 1366.201), and then only at the state level (Model 1.11; AIC = 1169.987, BIC = 1253.621), however, the AIC and BIC values for Model 1.7 were still the lowest.

The summary of the estimates for Model 1.7 are presented in Table 12. Given that the model includes a log-transformed outcome variable, the model estimates are therefore also on the log scale. The exponentiated estimate values are presented in parentheses to ease interpretability of the log-scale values. The fixed effect of Comparison Group ($b = 0.044, p < .001$) was statistically significant in the model, indicating that log risk ratios were significantly higher when

using all other students as a comparison group. The random effect of Year 2 versus Year 1 ($b = 0.011, p = .668$) was not statistically significant while the random effect of Year 3 versus Year 1 ($b = 0.065, p = .020$) was, indicating that when all else is equal in the model, the mean log risk ratio values did not differ significantly between 2012 – 2013 and 2013 – 2014, but the mean log risk ratio values were significantly higher in 2014 – 2015 compared to in 2012 – 2013. Neither level of the interaction term was found to be statistically significant in the model. Overall, these findings answer RQ 1 and part of RQ 3. Specifically, mean log risk ratios across three years are significantly higher when using all other students as a comparison group and that log risk ratio values have remained stable between 2012 – 2013 to 2013 – 2014, but have significantly increased from 2012 – 2013 to 2014 – 2015. Based on Model 1.7, the predicted risk ratios were calculated and are presented in Table 13. Since the model coefficients are on a log scale, exponentiated values are also presented for ease of interpretability.

Model Fitting: Disproportionate Status as Outcome Variable

A second series of random-intercept models were fit, this time with disproportionate status as the outcome variable. This variable was dummy coded (0 = Not disproportionate, 1 = Disproportionate). Unlike the first series of models that were fit, this series requires the use of a logit link function given that the outcome variable is a dichotomous variable. Models were systematically fit with added components and compared to subsequent models. Statistical significant difference was determined by calculating a chi-square statistic. The first model (i.e., Model 2.0) was fit with the fixed effect of Comparison Group with a random intercept at the school, district, and state levels. Then the random effect of Time was added (i.e., Model 2.1). The variance components in Model 2.1 indicate that all the random effect variance in the model was accounted for by the first two levels (i.e., school and district levels), and no variance was

accounted for in the third level (i.e. state level; $SD = 0$). Therefore, the random intercept at the state level was dropped in the subsequent model. The random effect of Comparison Group*Time, the interaction term, was added next (i.e., Model 2.2). The results of the chi-square test between the two models are presented in Table 14. The chi-square statistic ($\chi^2 = 1.485$, $df = 1$, $p = .223$) indicates that fit for Model 2.2 is not better than Model 2.1. The AIC and BIC values across all three models are also comparable. Model 2.2 was selected, although it is not the most parsimonious model given that it includes the non-significant interaction term. The interaction term is needed to answer RQ 3. The fixed effect of Comparison Group All other students ($b = 0.582$, $p = .020$) was statistically significant in the model, indicating that the use of all other students predicted a significantly higher likelihood of a school having a significantly disproportionate risk ratio value. By exponentiating the beta of Comparison Group, we find that predicted odds of a school having disproportionate status increase by a factor of 1.790. Neither Time coefficient was significant. Overall, these findings provide answers to RQ 2 and part of RQ 3. Specifically, using all other students as a comparison group significantly increases the likelihood that the school's risk ratio value will be significantly disproportionate. Also, there is not a significant difference between the use of different comparison groups on the likelihood that a school's risk ratio value will be significantly disproportionate. See Table 15 for a summary of model estimates. Table 16 contains the predicted probabilities based on Model 2.2.

Chapter 5: Discussion

The purpose of this study was to analyse the use of different comparison groups when calculating risk ratios as a metric to evaluate the discipline gap. What is the differential impact of using White and all other students as comparison groups when calculating the risk ratio for Black students to receive at least one out-of-school suspension? The lack of consistency in the literature and throughout state policies on how the discipline gap is researched, measured, and reported on makes it difficult to hypothesize what the current state of the discipline gap may be. Nonetheless, given the lack of evidence of targeted intervention on a system level, it was hypothesized that the yearly mean risk ratio values would be non-statistically different from the 2012 – 2014. Given that the impetus behind this study is the insufficient research to support proposed changes in legislation that would mandate the use of all other students as comparison groups when calculating risk ratios, this study of the differential impact of different comparison groups was exploratory in nature.

Two series of mixed models on a sample of 594 schools in 288 unique districts across 36 states. Models were fit to answer the follow research questions:

RQ 1: Are there differences in risk ratios for Black students receiving one or more OSS, based on the comparison group (i.e., White or all other students) used?

RQ 2: Are there differences in classifications (i.e., percentage of schools identified as being disproportionate) based on the comparison group used?

RQ 3: From 2012-2014, are there statistically significant differences between the yearly risk ratio values, and if so, are these differences dependent on the comparison group used to calculate risk ratios?

Results from the first series of model fitting (i.e., Models 1.0 – 1.11) were used to answer RQ1. The final model (i.e., Model 1.7) that was selected to answer RQ1 indicates that, when all else is equal, using all other students as a comparison group yields significantly higher risk ratio values for Black students to receive one or more OSS. Specifically, the predicted risk ratio across three years for Black students to receive at least one out-of-school suspension are statistically significantly different depending on which comparison group is used. Predicted risk ratios range between 2.479 – 2.646 per year when using all other students and 2.553 – 2.719 per year when using White students as comparison groups. Although the difference is statistically different, it may not be a practical difference.

Results from the second series of model fitting (i.e., Models 2.0 – 2.2) were used to answer RQ 2. The results from Model 2.2 indicate that the likelihood of disproportionate status differed by comparison group. Specifically, using all other students as a comparison group increases the odds of a school having disproportionate status versus not by a factor of 1.79.

Based on Model 1.7 estimates, the mean differences of log risk ratio values were not statistically significantly different between the first two years (i.e., 2012 – 2013 and 2013 – 2014) and but were statistically significantly higher between the second and third years (i.e., 2013 – 2014 and 2014 – 2015). In other words, risk ratio values were found to be significantly higher in 2014 – 2015 compared to 2012 – 2013. With regards to RQ 3, the effect of Comparison Group on Time was evaluated by the interaction terms in both final models selected (i.e., Models 1.7 and 2.2). None of the interaction terms were found to be significant, indicating that neither risk ratio values nor likelihood of school disproportionate status over Time differ significantly by Comparison Group. Based on the sample, this would indicate that the discipline gap between Black and other students, when using risk ratios as a metric, has significantly increased (i.e.,

widened), regardless of which comparison group is used (i.e., White and all other students). This is different than what would be expected based on the population demographics where risk ratios calculated using White or all other students as a comparison group should yield narrower values over time (Girvan, McIntosh, & Smolkowski, 2017), as well as the declining trend of suspensions for Black students that was found previously found (Noltemeyer & McLoughlin, 2010). Noltemeyer and McLoughlin posited that their finding of a declining trend could have potentially been due to a decreased reliance of the schools on exclusionary discipline. While no empirical evidence is provided to support such a position, the opposite finding in this study of increasing risk ratio values (i.e., the mean risk ratio values for 2014 were statistically significantly higher than in 2012) may be due to the use of a wider sample. This study included data from schools across 36 states, which may have attenuated the effects of a supposed change (i.e., less reliance on exclusionary discipline) in one state since all the schools in the Noltemeyer study were in Ohio. The issue of limited consistency in how the discipline gap is measured is manifest here. I had hypothesized that due to the lack of evidence of any specific intervention or policy change to address the discipline gap on a *national* level, results would reveal non-statistically different yearly mean risk ratio values, however, it appears that under business as usual, racial inequities, specifically those that disadvantage Black students, are increasing.

Study Strength and Limitations

The primary strength of this study is the use of mixed modeling to consider the nested nature of educational data. Historically, researchers of the discipline gap have relied on methodologies that fail to consider the shared variance that may exist within- and between-levels. Another strength of the study is the use of recent school discipline data. The discipline

gap has been widely documented and evidenced for more than 50 years, and it is important from a social justice perspective that the spotlight continues to be placed on inequities that exist.

Finally, the use of a sample that includes schools across 36 states provides a wider view of the discipline gap. While many previous studies have looked at the discipline gap at a more local level (e.g., Noltemeyer & McLoughlin, 2010), few have analyzed nationwide data. Given that educational policies generally fall under federal government purview, having access to overarching trends about the discipline gap on a national level, research may influence national-level policy more accurately.

Although the categorization of students in predefined racial and ethnic groups is consistent with U.S. Census categorization methodology, it is a significant limitation of this study. For this study, students needed to have been identified with only one racial group (i.e., Black or White) and could not have any additional racial group associations. This leads to serious limitations given that students who identify with other racial groupings may have similar experiences as those who are categorized in only one category. Perhaps by excluding students with multiple racial affiliations, but were still Black, the risk ratios that were calculated in this study may be an underestimate. Although outside the scope of this study, the mere categorization of people as being Black, for example, is sometimes highly debated – see Ladson-Billings and Tate (1995) for a historical overview of this issue.

Any risk ratio greater than 1 is an over-representation; however, significant disproportionality in this study was defined as risk ratios that exceeded a predefined threshold (i.e., 1.873). This cut-off point is an example threshold that states may choose to use when evaluating disproportionality between racial groups. However, this cut-off point was derived from a statistical analysis by calculating two median absolute deviations above the national

median of local educational agencies for total removals includes in-school and out-of-school suspensions, and expulsions. The risk ratios calculated in this study only included out-of-school suspensions. The use of total removals (i.e., 1.873) resulted in more schools being found disproportionate than if the two median absolute deviations above the national mean for local educational agencies for just out-of-school suspension/expulsions (i.e., 2.008 – 3.000) has been used. The example thresholds as provided by the Office of Special Education and Rehabilitation Services (2016) do not separate out-of-school suspensions and expulsion rates. Furthermore, the cut-off point as suggested may not be a socially valid threshold of determining inequality. Though, perhaps any threshold chosen will be subject to this critique. Furthermore, the use of an absolute cut-off point does not adequately address any precision error. For example, a cut-off point without a confidence interval to be used to be as a measure of precision may lead to incorrect identification of schools being disproportionate. Any final threshold chosen should address measurement error.

Implications for Practice and Policy

The findings of this study are important to policy makers for several reasons. The finding that mean risk ratio values have increased may be an indication that the discipline gap is widening. Policies should not be based on the results of one metric (Girvan et al., 2017) or findings from a limited number of years. The allocation of resources to address the inequality that exists may be warranted. While policies need to address the discipline gap, policy makers are not left with many recommendations that specifically address racial inequities in discipline. While some interventions seem promising, such as My Teaching Partner Secondary (MTP-S; see Gregory et al., 2016) until more research-based interventions or strategies can be recommended, policy makers should look at reducing overall exclusionary discipline rates. Given the myriad of

negative outcomes associated with receiving an OSS, efforts to reduce absolute OSS rates may have far-reaching positive consequences on a student who avoids receiving an OSS, regardless if it is a White or Black student who benefits.

The use of all other students as a comparison group yielded significantly higher risk ratio values. However, although the difference is statistically significant, is not necessarily practically different. Therefore, if the intention behind the proposed legislation is to ensure that more offending schools are identified and are held accountable for being significantly disproportionate, then the move towards using all other students as a comparison group may be supported. Considering the proposed legislation to use all other students, it is conceivable then that more schools will be found as being significantly disproportionate if some of the states were previously reporting risk ratios with White students as a comparison group. While some may argue that it is better to over-identify schools – which may happen by using all other students as a comparison group – policy makers should carefully evaluate the consequences that are triggered when schools are found to be disproportionate. A balanced approach between imposing the use of resources in a certain way to reduce risk ratios and ensuring that schools are not unnecessarily taking resources from other worthy initiatives is needed. Regardless of how the proposed legislation will impact identification rates, the use of a common comparison group will allow federal policy-makers to make decisions based on data from states that are comparable.

Another important point to consider is the threshold that is used to determine if significant disproportionality. This study was based on the use of a threshold for determining significant disproportionality, however, the proposed legislation would allow states to self-define thresholds. Since reported rates of disproportionality for LEAs are lower than what would be expected when states are able to self-define thresholds (Office of Special Education and

Rehabilitative Services & Department of Education, 2016), it is not known how states will define thresholds. While the impetus behind the proposed legislation is to standardize the way that disproportionality is measured and reported, having an unstandardized way of determining threshold values across states may lead to too much variability in the way states measure and report on disproportionality. Close monitoring will be needed to gauge how this freedom impacts reported rates of disproportionality.

Implications for Research

Since the study by the Children's Defence Fund in 1975, researchers have continued to collect evidence that indicated that students from racial and ethnic minority backgrounds received exclusionary discipline at higher rates than their White counterparts. This study provides one of the most recent updates about the issue of the discipline gap, and the findings of this study underscore the need for more research to be done in this field. The findings tell us that Black students may be disproportionately disciplined at an increasing rate, however, there is much more research to be done about identifying and addressing causal factors of the discipline gap. While previous studies have looked at different factors such as SES (Gregory et al., 2010) results have been mixed. More recent research (e.g., Smolkowski et al., 2016) that is addressing the implicit bias that exists amongst educators appears to be very promising.

This study addressed the statistical difference between different comparison groups in risk ratios and it did not address the validity of this metric. While risk ratios are suggested to be most the most sensitive of metrics to the discipline gap (Girvan et al., 2017), further research is needed to establish validity of risk ratios as a metric to be used when evaluating the discipline gap.

Finally, this research addressed the use of different comparison groups when using Black students as a target group. More research is needed to confirm that the findings of this study would be found for other minority groups as well, such as Hispanics and American Indian. This is especially important given that the proposed legislation calls for the use all other students for all minority groups, however, there does not appear to be any research to support the use of one metric for all groups. This research did not look at the difference in risk ratio values between different types of schools (i.e., elementary, middle, and high schools). Research should look at how risk ratios are impacted given that exclusionary disciplinary practices vary between school type. For example, higher suspension rates were found in middle schools than elementary schools by Arcia (2007).

Chapter 6: Summary

In conclusion, the use of all other students as a comparison group when calculating risk ratios for Black students yields significantly higher risk ratios when compared to the use of White students as a comparison group. The likelihood that a school's risk ratio value is significantly disproportionate increases when using all other students as a comparison group. Yearly mean risk ratios for Black students to receive at least one out-of-school suspensions have increased from 2012 to 2014, regardless of which metric is used to evaluate the discipline gap.

Figures

Figure 1 Quantile-Quantile Plot of risk ratios and log risk ratios variables

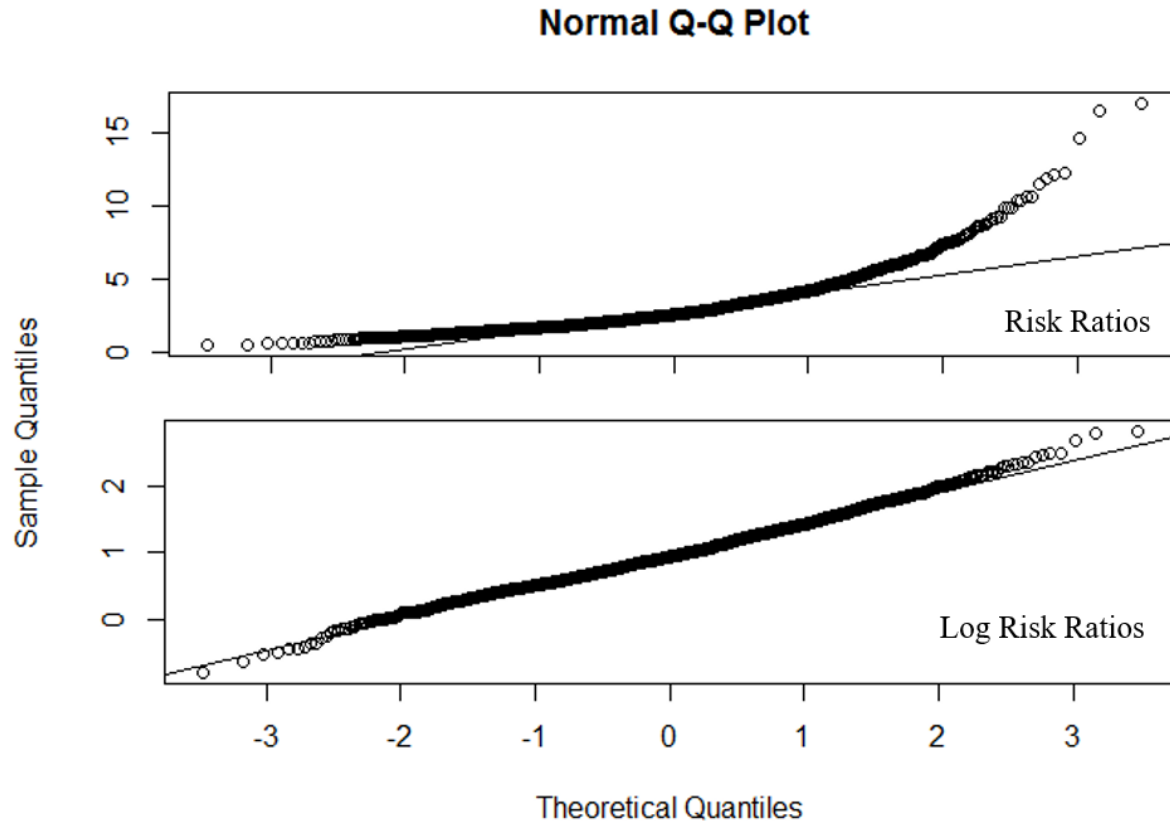


Figure 2 Risk ratios for Black students to receive at least one OSS, based on observed data

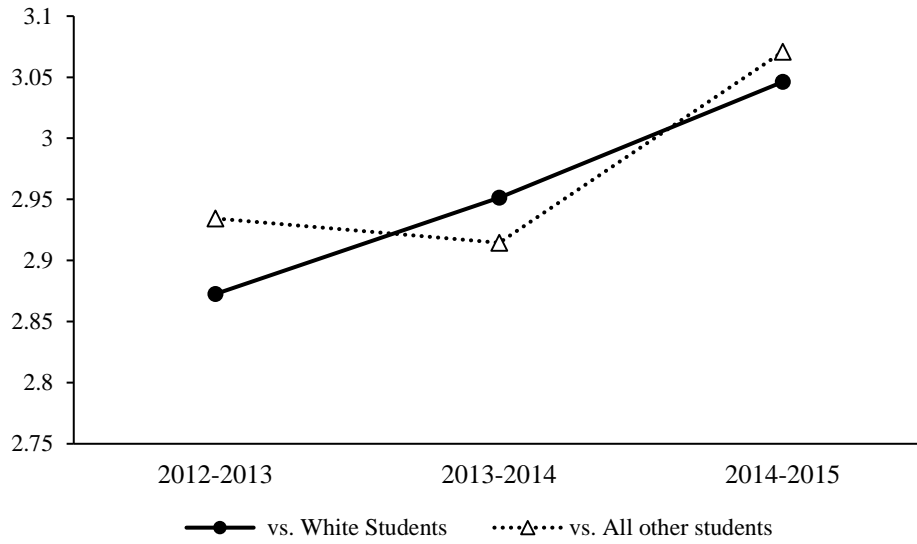
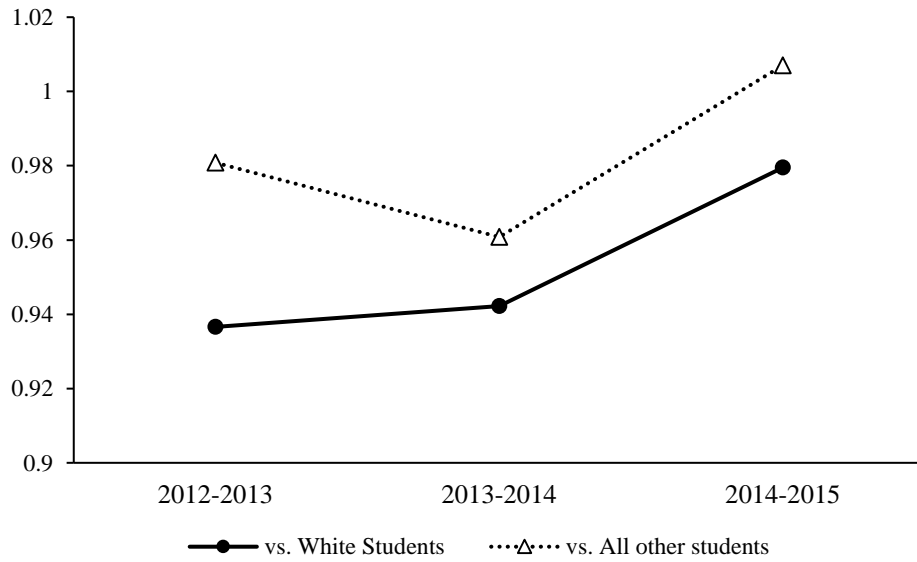


Figure 3 Log risk ratios for Black students to receive at least one OSS, based on observed data



Tables

Table 1 Proportion of juvenile arrests by Black youth between 2003 and 2012

Most Serious Offence	2003	2012
Murder	48%	52%
Forcible rape	33%	33%
Robbery	63%	69%
Aggravated assault	38%	43%
Simple Assault	NR	39%
Burglary	26%	39%
Larceny-theft	27%	35%
Motor vehicle theft	40%	40%
Weapons	32%	37%
Drug abuse violations	26%	23%
Runaways	20%	NR
Vandalism	18%	23%
Liquor laws	4%	7%

Note: NR implies that data for that specific offence category was not reported for that year.

Table 2 Summary of schools included in analysis

Year	<i>n</i> unique schools	<i>n</i> unique districts	<i>n</i> unique states	N missing schools
		Mean count of unique schools per district (range)	Mean count of districts per state (range)	
2012-2013	355	189 1.87 (1 – 22)	33 10.76 (1 – 53)	239
2013-2014	308	170 1.81 (1 – 19)	34 9.06 (1 – 49)	286
2014-2015	312	172 1.81 (1 – 16)	30 10.4 (1 – 42)	282
Total	594	288 -	36 -	-

Table 3 Percentage of schools in observed data that have disproportionate risk ratios

Year	Comparison Group	
	White students	All other students
2012-2013	93%	96%
2013-2014	94%	95%
2014-2015	95%	96%

Note: Values represent percentage of schools in observed data that are found to have disproportionate (i.e., the risk ratios for Black students to receive one or more OSS) is greater than the example threshold (i.e., 1.873)

Table 4 ICC values based on empty model

Level	SD (Variance)	ICC Value
L1 - School	0.318 (0.101)	0.388
L2 - District	0.227 (0.051)	0.197
L3 - State	0.22 (0.049)	0.186

Table 5 Mean, SD, skewness, and kurtosis of risk ratios pre- and post- log transformation

Variable	Mean (SD)	Skewness	Kurtosis
Risk Ratios	2.962 (1.620)	2.341	13.202
Log Risk Ratios	0.967 (0.477)	0.300	3.392

Table 6 Mean risk ratios and log risk ratios for Black students to receive at least one OSS, based on observed data

Year	Versus White students		Versus All other students	
	Mean (SD)	Range	Mean (SD)	Range
2012-2013	2.873 (1.558)	0.528 - 10.335	2.934 (1.433)	1.004 - 11.53
2013-2014	2.952 (1.743)	0.586 - 12.243	2.915 (1.566)	0.858 - 14.674
2014-2015	3.046 (1.779)	0.445 - 17.06	3.071 (1.653)	0.82 - 16.458
Overall	2.953 (1.69)	0.445 - 17.06	2.972 (1.548)	0.82 - 16.458

Year	Log Risk Ratio		Log Risk Ratio	
	Mean (SD)	Range	Mean (SD)	Range
2012-2013	0.937 (0.474)	-0.639 - 2.336	0.981 (0.424)	0.004 - 2.445
2013-2014	0.942 (0.526)	-0.534 - 2.505	0.961 (0.453)	-0.154 - 2.686
2014-2015	0.98 (0.513)	-0.809 - 2.837	1.007 (0.471)	-0.198 - 2.801
Overall	0.952 (0.503)	-0.809 - 2.837	0.983 (0.449)	-0.198 - 2.801

Table 7 Summary of model estimates with systematic addition of random intercepts

Model	Description	SD at		
		School Level	District Level	State Level
1.0	Baseline (no intercept)	-	-	-
1.1	+ random intercept at school level	0.426	-	-
1.2	+ random intercept at district level	0.315	0.303	-
1.3	+ random intercept at state level	0.318	0.227	0.220

Table 8 Summary of results of log likelihood tests between models with random intercepts

Model	Description	<i>df</i>	AIC	BIC	Log Likelihood	Models Tested	χ^2	<i>p</i>
1.0	Baseline	2	2648.574	2659.726	-1322.29			
1.1	+ random intercept at school level	3	1411.435	1428.161	-702.717	1.0 vs. 1.1	1239.14	<.001
1.2	+ random intercept at district level	4	1349.301	1371.603	-670.65	1.1 vs. 1.2	64.1342	<.001
1.3	+ random intercept at state level	5	1321.975	1349.853	-655.988	1.2 vs. 1.3	29.3255	<.001

Table 9 Summary of explanatory variables in models with addition of Comparison Group, Time, and Comparison Group*Time

Explanatory Variables	Model 1.3 (Empty)	Model 1.4 – add Comparison Group	Model 1.5 – add Time	Model 1.6 – add Comparison Group*Time
Intercept	0.9477*** (2.5798)	0.9323*** (2.5403)	0.9097*** (2.4836)	0.9030*** (2.467)
Comparison Group All other students vs. White students		0.0308** (1.0313)	0.0308** (1.0313)	0.0441* (1.0451)
Time				
Year 2 vs. Year 1			0.0053 (1.0053)	0.0180 (1.0182)
Year 3 vs. Year 11			0.0658*** (1.068)	0.07413*** (1.0769)
Comparison Group *Time				
All other students vs. White students*Year 2 vs. Year 1				-0.0255 (0.9748)
All other students vs. White students*Year 3 vs. Year 1				-0.0167 (0.9834)

Note: A variable intercept was used, which varied at each level (i.e., school, district, and state). Comparison Group is a fixed effect, while Time is a random effect. Exponentiated values are presented in parentheses for ease of interpretation. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 10 Summary of results of log likelihood tests between models with fixed effects

Model	Description	df	AIC	BIC	Log Likelihood	Models Test	χ^2	p
1.3	Baseline (random intercepts at all levels)	5	1321.975	1349.853	-655.9875			
1.4	+ fixed effect of Comparison Group	6	1316.212	1349.666	-652.1061	1.3 vs. 1.4	7.762807	<.001
1.5	+ random effect of time	8	1302.125	1346.730	-643.0625	1.4 vs. 1.5	18.087225	0.001
1.6	+ random effect of Comparison Group*Time	10	1305.168	1360.924	-642.584	1.5 vs. 1.6	0.956887	0.620

Table 11 Summary of results of log likelihood tests between models with addition of random slope of time

Model	Description	df	AIC	BIC	Log Likelihood	Models Test	χ^2	p
1.6	Baseline (random intercepts at all levels, fixed effects of Comparison Group, Time, and Comparison Group*Time)	10	1305.1681	1360.924	-642.584			
1.7	Baseline + random slope of Time at the school level	15	946.8565	1030.49	-458.4283	1.6 vs. 1.7	368.3115	<.001
1.8	Model 1.7 + random slope of Time at district level				<i>Did not converge</i>			
1.9	Model 1.8 + random slope of Time at the state level				<i>Did not converge</i>			
1.10	Baseline + random slope of Time at district level	15	1282.568	1366.201	-626.2838			
1.11	Baseline + random slope of Time at state level	15	1169.987	1253.621	-569.9934			

Table 12 Summary of explanatory variables for final model (i.e., Model 1.7) selected from first series of models

Explanatory Variables	Estimate (Exponentiated)
Intercept	0.9078*** (2.4789)
Comparison Group All other students vs. White students	0.0441*** (1.0451)
Time	
Year 2 vs. 1	0.0107 (1.0108)
Year 3 vs. 1	0.0652* (1.0674)
Comparison Group * Time	
All other students vs. White students*Year 2 vs. Year 1	-0.0255 (0.9748)
All other students vs. White students*Year 3 vs. Year 1	-0.0167 (0.9834)
Variance Components	Standard Deviation
Random Intercepts	
School Level (with random slope)	0.3386
District Level	0.2262
State Level	0.2184
Random Slope for Effects of Time at School Level	
Intercept	0.3386
Year 2 vs. Year 1	0.3064
Year 3 vs. Year 1	0.3670
Residual	0.1714
Note: A variable intercept was used, which varied at each level (i.e., school, district, and state). Comparison Group is a fixed effect, while Time is a random effect. Exponentiated values are presented in parentheses for ease of interpretation * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$	

Table 13 Model predicted log risk ratios

Year	Comparison Group	
	White students	All other students
2012-2013	0.908 (2.479)	0.952 (2.591)
2013-2014	0.919 (2.505)	0.937 (2.553)
2014-2015	0.973 (2.646)	1 (2.719)
Overall	0.933 (2.543)	0.963 (2.621)

Note: Regression coefficients are based on Model 1.7. Exponentiated values are presented in parentheses for ease of interpretation

Table 14 Summary of results of chi-square tests between models in second series of models

Model	Description	<i>df</i>	AIC	BIC	Log Likelihood	Deviance	Models Test	χ^2	<i>p</i>
2.0	Baseline (random intercepts at all levels and fixed effect of Comparison Group)	5	1764.5	1792.3	-877.23	1754.5			
2.1	+ random effect of Time	7	1768.1	1807.2	-877.07	1754.1	2.0 vs. 2.1	0.327	0.8492
2.2	2.1 + random effect of Comparison Group*Time; - random intercept at state level	8	1768.7	1813.3	-876.33	1752.7	2.1 vs. 2.2	1.4853	0.2229

Table 15 Summary of explanatory variables for models in second series of model fitting

Explanatory Variables	Model 2.0 (Baseline)	Model 2.1 – add Comparison Group	Model 2.2 – add Time, remove random intercept at state level
Intercept	2.7448*** (15.562)	2.736*** (15.425)	2.6314*** (13.893)
Comparison Group All other students vs. White students	0.3418* (1.407)	0.342* (1.408)	0.582* (1.79)
Time			
Year 2 vs. 1		-0.049 (0.952)	0.1138 (1.121)
Year 3 vs. 1		0.073 (1.076)	0.2675 (1.307)
Comparison Group*Time			
All other students vs. White students*Year 2 vs. Year 1			-0.3425 (0.71)
All other students vs. White students*Year 3 vs. Year 1			-0.4072 (0.666)
Variance Components		<i>SD</i>	
Random Intercepts			
School Level	2.361	2.359	2.368
District Level	2.141	2.145	2.151
State Level	~0	0	
Note: A variable intercept was used, which varied at each level (i.e., school, district, and state). Comparison Group is a fixed effect, while Time is a random effect. Exponentiated values are presented in parentheses for ease of interpretation * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.			

Table 16 Model predicted probability of disproportionate status

Year	Comparison Group	
	White students	All other students
2012-2013	94%	96%
2013-2014	94%	95%
2014-2015	94%	96%
Note: Regression coefficient values are based on Model 2.2.		

References

- American Psychological Association. (2008). Are zero tolerance policies effective in the schools?: an evidentiary review and recommendations. *The American Psychologist*, 63(9), 852.
- Arcia, E. (2006). Achievement and Enrollment Status of Suspended Students: Outcomes in a Large, Multicultural School District. *Education and Urban Society*, 38(3), 359-369. doi:10.1177/0013124506286947
- Arcia, E. (2007). A Comparison of Elementary/K-8 and Middle Schools' Suspension Rates. *Urban Education*, 42(5), 456-469. doi:10.1177/0042085907304879
- B.C. Ministry of Education. (2016). How are we doing? Retrieved September 1, 2017, from http://www.bced.gov.bc.ca/reports/pdfs/ab_hawd/Public.pdf
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01
- Bellringer, C. (2015). An Audit of the Education of Aboriginal Students in the B.C. Public School System. Retrieved September 1, 2017, from http://www.bcauditor.com/sites/default/files/publications/reports/OAGBC%20Aboriginal%20Education%20Report_FINAL.pdf
- Bollmer, J. M., Bethel, J. W., Munk, T. E., & Bitterman, A. R. (2014). Methods for Assessing Racial/Ethnic Disproportionality in Special Education.
- Boneshefski, M. J., & Runge, T. J. (2014). Addressing Disproportionate Discipline Practices Within a School-Wide Positive Behavioral Interventions and Supports Framework A Practical Guide for Calculating and Using Disproportionality Rates. *Journal of Positive Behavior Interventions*, 16(3), 149-158.

- Brown v. Board of Education of Topeka, No. 483, 347 (U.S. Supreme Court 1954).
- Children's Defense Fund. (1975). School suspensions: Are they helping children?
- Chinn, P. C., & Hughes, S. (1987). Representation of Minority Students in Special Classes. *Remedial and Special Education (RASE)*, 8(4), 41-46.
- Christle, C., Nelson, C. M., & Jolivette, K. (2004). School Characteristics Related to the Use of Suspension. *Education & Treatment of Children*, 27(4), 509-526.
- Curry, C. (1997). The Federal Crime Bill: What Will it Mean for California? Retrieved September 1, 2017, from <http://www.lao.ca.gov/1994/pb092794.html>
- Davies, H. T. O., Crombie, I. K., & Tavakoli, M. (1998). When can odds ratios mislead? *BMJ*, 316(7136), 989-991.
- Department of Education. (2016). Persistent Disparities Found Through Comprehensive Civil Rights Survey Underscore Need for Continued Focus on Equity, King Says. Retrieved September 1, 2017, from <http://www.ed.gov/news/press-releases/persistent-disparities-found-through-comprehensive-civil-rights-survey-underscore-need-continued-focus-equity-king-says>
- Department of Education. (n.d.). EVERY STUDENT SUCCEEDS ACT Accountability, State Plans, and Data Reporting: Summary of Final Regulations. Retrieved September 1, 2017, from https://www.acsa.org/download_file/view/1858/1053
- Department of Education Office for Civil Rights. (2016). 2013-2014 CIVIL RIGHTS DATA COLLECTION. Retrieved September 1, 2017, from <https://www2.ed.gov/about/offices/list/ocr/docs/2013-14-first-look.pdf>
- DeRidder, L. M. (1991). How Suspension and Expulsion Contribute to Dropping Out. *Education Digest*, 56(6), 44-47.

- Dyches, T., & Prater, M. (2010). Disproportionate representation in special education: Overrepresentation of selected subgroups. In F. E. Obiakor & J. P. Bakken (Eds.), *Current Issues and Trends in Special Education: Identification, Assessment and Instruction* (pp. 53-71).
- Gibb, A. C., & Skiba, R. (2008). Using Data to Address Equity Issues in Special Education. Education Policy Brief. Volume 6, Number 3, Winter 2008. *Center for Evaluation and Education Policy, Indiana University*.
- Girvan, E. J., McIntosh, K., & Smolkowski, K. (2017). Tail, Tusk, and Trunk: An Examination of What Different Metrics Reveal about Racial Disproportionality in School Discipline.
- Greflund, S., McIntosh, K., Mercer, S. H., & May, S. L. (2014). Examining Disproportionality in School Discipline for Aboriginal Students in Schools Implementing PBIS. *Canadian Journal of School Psychology, 29*(3), 213-235.
- Gregory, A., Hafen, C. A., Ruzek, E., Mikami, A. Y., Allen, J. P., & Pianta, R. C. (2016). Closing the Racial Discipline Gap in Classrooms by Changing Teacher Practice. *School Psychology Review, 45*(2), 171-191.
- Gregory, A., Skiba, R. J., & Noguera, P. A. (2010). The Achievement Gap and the Discipline Gap: Two Sides of the Same Coin? *Educational Researcher, 39*(1), 59-68.
doi:10.3102/0013189x09357621
- KewelRamani, A., Gilbertson, L., Fox, M., & Provasnik, S. (2010). Status and Trends in the Education of Racial and Ethnic Groups. Retrieved September 1, 2017, from <http://nces.ed.gov/pubs2010/2010015.pdf>
- Kim, C., Losen, D. J., & Hewitt, D. (2010). The school to prison pipeline: structuring legal reform (Vol. 1). New York: New York University.

- Ladson-Billings, G., & Tate, W. F. (1995). Toward a Critical Race Theory of Education. *Teachers College Record*, 97(1), 47 - 68.
- Lee, T., Cornell, D., Gregory, A., & Fan, X. (2011). High suspension schools and dropout rates for black and white students. *Education and Treatment of Children*, 34(2), 167-192.
- Monroe, C. R. (2005). Why are "bad boys" always Black?: Causes of disproportionality in school discipline and recommendations for change. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 79(1), 45-50.
- Morris, E. W., & Perry, B. L. (2016). The Punishment Gap: School Suspension and Racial Disparities in Achievement. *Social Problems*, 63(1), 68-86. doi:10.1093/socpro/spv026
- Nicholson-Crotty, S., Birchmeier, Z., & Valentine, D. (2009). Exploring the Impact of School Discipline on Racial Disproportion in the Juvenile Justice System. *Social Science Quarterly*, 90(4), 1003-1018.
- Noltemeyer, A. L., Marie, R., McLoughlin, C., & Vanderwood, M. (2015). Relationship Between School Suspension and Student Outcomes: A Meta-Analysis. *School Psychology Review*, 44(2), 224-240. doi:10.17105/spr-14-0008.1
- Noltemeyer, A. L., & McLoughlin, C. S. (2010). Changes in Exclusionary Discipline Rates and Disciplinary Disproportionality over Time. *International Journal of Special Education*, 25(1), 59-70.
- Office of Special Education and Rehabilitation Services. (2016). Racial and Ethnic Disparities in Special Education. Retrieved September 1, 2017, from <https://www2.ed.gov/programs/osepidea/618-data/LEA-racial-ethnic-disparities-tables/disproportionality-analysis-by-state-analysis-category.pdf>

- Office of Special Education and Rehabilitative Services, & Department of Education. (2016). Assistance to States for the Education of Children With Disabilities; Preschool Grants for Children With Disabilities. Retrieved September 1, 2017, from <http://www.federalregister.gov/a/2016-03938/p-14>
- Office of Superintendent of Public Instruction. (n.d.). OSPI Performance Indicators. Retrieved September 1, 2017, from <http://www.k12.wa.us/DataAdmin/PerformanceIndicators/MeasuringDisproportionality.html>
- Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*(1), 85-112. doi:10.1016/j.jsp.2009.09.002
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2017). nlme: Linear and Nonlinear Mixed Effects Models.
- Puzzanchera, C. (2014). Juvenile Arrests 2012. Retrieved September 1, 2017, from <https://www.ojjdp.gov/pubs/248513.pdf>
- Quan, D. (2017). *Unlocking Student Potential Through Data: Final Report*. Retrieved from R Core Team. (2017). R: A language and environment for statistical computing (Version 3.4.0). Vienna, Austria: R Foundation for Statistical Computing.
- Reschly, D. J. (1997). Disproportionate Minority Representation in General and Special Education: Patterns, Issues, and Alternatives: Mountain Plains Regional Resource Center and Drake University.
- Seattle Times. (n.d.). Race dramatically skews discipline, even in elementary school. *Education Lab*. Retrieved September 1, 2016, from <http://www.seattletimes.com/education-lab/race-dramatically-skews-discipline-even-in-elementary-school/>

- Skiba, R. J., Horner, R. H., Choong-Geun, C., Rausch, M. K., May, S. L., & Tobin, T. (2011). Race Is Not Neutral: A National Investigation of African American and Latino Disproportionality in School Discipline. *School Psychology Review*, 40(1), 85-107.
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. L. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *The Urban Review*, 34(4), 317-342.
- Skiba, R. J., & Peterson, R. L. (2000). School Discipline at a Crossroads: From Zero Tolerance to Early Response. *Exceptional Children*, 66(3), 335-346.
doi:10.1177/001440290006600305
- Skiba, R. J., Shure, L., & Williams, N. (2012). RACIAL AND ETHNIC DISPROPORTIONALITY IN SUSPENSION AND EXPULSION. In A. L. Noltemeyer & C. S. McLoughlin (Eds.), *Disproportionality in education and special education: a guide to creating more equitable learning environments*. Springfield, Ill: Charles C. Thomas.
- Smith, E. J., & Harper, S. R. (2015). Disproportionate impact of K-12 school suspension and expulsion on Black students in southern states. *Centre for the Study of Race and Equity in Education*. Retrieved September 1, 2017, from https://www.gse.upenn.edu/equity/sites/gse.upenn.edu/equity/files/publications/Smith_Harper_Report.pdf
- Smolkowski, K., Girvan, E. J., McIntosh, K., Nese, R. N., & Horner, R. (2016). Vulnerable Decision Points for Disproportionate Office Discipline Referrals: Comparisons of Discipline for African American and White Elementary School Students. *Behavioral Disorders*.

- Snyder, H. N. (2005). Juvenile Arrests 2003. Retrieved September 1, 2017, from <https://www.ncjrs.gov/pdffiles1/ojjdp/209735.pdf>
- Spaulding, S. A., Irvin, L. K., Horner, R. H., May, S. L., Emeldi, M., Tobin, T. J., & Sugai, G. (2010). Schoolwide social-behavioral climate, student problem behavior, and related administrative decisions empirical patterns from 1,510 schools nationwide. *Journal of Positive Behavior Interventions*, 12(2), 69-85.
- Sugai, G., Sprague, J. R., Horner, R. H., & Walker, H. M. (2000). Preventing school violence the use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of Emotional and Behavioral Disorders*, 8(2), 94-101.
- Toronto District School Board. (2010). ACHIEVEMENT GAP TASK FORCE DRAFT REPORT. Retrieved September 1, 2017, from <http://www.tdsb.on.ca/Portals/0/Community/Community%20Advisory%20committees/1CAC/Subcommittees/AchievementGapReptDraftMay172010.pdf>
- Townsend, B. L. (2000). The disproportionate discipline of African American learners: Reducing school suspensions and expulsions. *Exceptional Children*, 66(3), 381-391.
- U. S. Department of Education. (2008). Coordinated Early Intervening Services (CEIS) Guidance. Retrieved December 11, 2016, from http://www2.ed.gov/policy/special/guid/idea/ceis_pg3.html
- U. S. Department of Education. (n.d.). IDEA Regulations DISCIPLINE. Retrieved September 1, 2017, from <http://idea-b.ed.gov/explore/view/p/,root,dynamic,TopicalBrief,6,.html>
- U.S. Department of Education. (2008). Policy Questions on the Department of Education's 2007 Guidance on Collecting, Maintaining and Reporting Data by Race or Ethnicity. Retrieved

September 1, 2017, from

<https://www2.ed.gov/policy/rschstat/guid/raceethnicity/questions.html>

Wald, J., & Losen, D. J. (2003). Defining and redirecting a school-to-prison pipeline. *New Directions for Youth Development*, 2003(99), 9-15. doi:10.1002/yd.51

Wallace, J. M., Goodkind, S., Wallace, C. M., & Bachman, J. G. (2008). Racial, Ethnic, and Gender Differences in School Discipline among U.S. High School Students: 1991-2005. *The Negro educational review*, 59(1-2), 47-62.

Wolf, K. C., & Kupchik, A. (2016). School Suspensions and Adverse Experiences in Adulthood. *Justice Quarterly*, 1-24. doi:10.1080/07418825.2016.1168475

Appendix A

List of possible infractions and outcomes that can be recorded in SWIS.

Infraction

Defiance/Insubordination/Non-Compliance
Physical Aggression
Disruption
Disrespect
Abusive Language/Inappropriate Language/Profanity
Tardy
Skip class
Harassment
Bullying
Fighting
Inappropriate Location/Out of Bounds Area
Truancy
Forgery/Theft/Plagiarism
Technology Violation
Property Damage/Vandalism
Lying/Cheating
Dress Code Violation
Inappropriate Display of Affection
Use/Possession of Tobacco
Use/Possession of Drugs
Use/Possession of Weapons
Use/Possession of Combustibles
Use/Possession of Alcohol
Gang Affiliation Display
Bomb Threat/False Alarm
Arson
Other Behavior

Outcome

Alternative Placement
Time Out/ Detention
Conference with Student
In-School Suspension
Loss of Privilege
*Out-of-School Suspension
Parent Contact
Time in Office
Individualized Instruction
Additional Attendance/Saturday School
Bus Suspension
Restitution/Community Service
Community Service
Expulsion
Action Pending
Other Action Taken