

ABSTRACT

Title of Dissertation: UNDERSTANDING DISPROPORTIONATE
SUSPENSIONS OF MINORITY STUDENTS
AND STUDENTS WITH DISABILITIES: A
MULTILEVEL APPROACH

Michael P. Krezmien, Doctor of Philosophy, 2007

Dissertation directed by: Professor Peter E. Leone.
Department of Special Education

This paper presents the findings from an investigation of suspension practices in Maryland. Logistic regression analysis, discriminant analysis, and hierarchical generalized linear modeling were employed to understand the individual characteristics and school characteristics associated with risk for suspension of secondary age students in Maryland public schools. The findings from the HGLM analyses revealed substantial variability in the suspension practices of schools, and indicated that school-level characteristics accounted for a majority of the explained variance in the suspensions of youth in Maryland. A number of school factors were significantly associated with suspensions of youth when Race and Disability were controlled as level-1 predictors. Race and

Disability were significant and robust predictors of the suspensions even when school-level factors were controlled. Results from this investigation are reported and discussed, and limitations to interpretation of the findings are described.

UNDERSTANDING DISPROPORTIONATE SUSPENSIONS OF
MINORITY STUDENTS AND STUDENTS WITH DISABILITIES:
A MULTILEVEL APPROACH

by

Michael P. Krezmien

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2007

Advisory Committee

Professor Peter E. Leone, Chair
Professor Patricia Alexander
Professor David Cooper
Professor Andrew Egel
Professor Deborah Speece

©Copyright by

Michael P. Krezmien

2007

Dedication

To Dad

“Thanks Pop. In all ways, this is for you. I love you.”

To My Pal

“This has been possible because of you.”

TABLE OF CONTENTS

List of Tables.....	vi
List of Figures.....	viii
Chapter I: Introduction.....	1
Disciplinary Removal.....	1
Race, Disability, and Exclusion.....	2
Consequences of School Exclusion.....	4
Suspensions: Impact of Individual and School Factors.....	6
Individual-Level Models.....	7
School-Level Models.....	9
The Multilevel Approach.....	12
Purpose.....	13
Odds Ratios.....	14
Research Questions.....	14
Definition of Terms.....	15
Chapter II: Literature Review.....	18
Methods for Review of Studies.....	19
Findings.....	20
Suspension Rates.....	20
School Level.....	23
Race.....	23
Disability.....	24
SES.....	24
Achievement.....	25
School Factors.....	26
Interactions.....	27
Multilevel Analysis.....	29
Summary of Findings.....	29
Methodological Review.....	31
Rationale, Research Questions, and Hypotheses.....	31
Sample Description.....	33
Summary.....	36
School and Setting Descriptions.....	36
Sampling Procedures.....	37
Adequacy of Variable Description.....	39
Data Collection.....	43

Data Analysis.....	43
Descriptive Designs.....	43
Correlational Designs.....	44
Sample Size.....	44
Adequacy of Statistical Procedures.....	47
Descriptive Statistics.....	47
Statistical Analysis.....	48
Main Effects.....	49
Follow-up Procedures.....	50
Effect Sizes and Confidence Intervals.....	51
Limitations.....	51
Summary.....	52
Chapter III: Method.....	55
Data Collection.....	55
Participants.....	55
Suspension Data.....	57
School Characteristics.....	58
School Data.....	59
Data Accuracy.....	59
Variables.....	60
Predictor Variables.....	61
MSA Basic.....	61
Attendance.....	61
Percent NHQ.....	61
Mobility.....	63
Enrollment.....	63
Suspension Rate.....	63
Percent SPED.....	63
Student-Teacher Ratio.....	63
SES.....	64
Percent White.....	64
Design and Data Analysis.....	64
Analysis 1.....	64
Analysis 2.....	65
Analysis 3.....	65
Analysis 4.....	67
Chapter IV: Results.....	68
Analysis 1.....	68
Analysis 2.....	73

Analysis 3.....	78
Race.....	80
Predictors of Suspension.....	82
Level-1 Predictor.....	84
Level-2 Predictors.....	85
Disability.....	86
Predictors of Suspension.....	86
Level-1 Predictor.....	88
Level-2 Predictors.....	88
Explanation of Variance.....	89
Analysis 4.....	89
Race.....	89
Disability.....	92
Chapter V: Discussion.....	96
Analysis 1.....	97
Analysis 2.....	98
Analysis 3.....	103
School-Level Predictors.....	104
Analysis 4.....	110
Limitations.....	112
Recommendations for Practice.....	114
Recommendations for Future Research.....	116
Conclusions.....	117
Appendix 1.....	121
Appendix 2.....	121
Appendix 3.....	123
References.....	124

LIST OF TABLES

Table 2.1. Citation of References for Literature Review and Description of Purpose	20
Table 2.2. Reports of sample size and presence of components required for sample description.....	34
Table 2.3. Descriptions of variables of studies included in literature review.....	40
Table 3.1. Sources of data and types of data extracted displayed by source and by level of analysis.....	56
Table 3.2. Predictor variables at the school level of analysis.....	62
Table 4.1. Descriptive information about the number of suspensions and the number of students suspended in Maryland public schools: 2004.....	72
Table 4.2. Descriptive information about the number of suspensions and the number of students suspended in sample schools: 2004.....	74
Table 4.3. Odds ratios for suspensions and 95% confidence intervals for disability categories by Race.....	76
Table 4.4. Descriptive Statistics for Suspensions for HGLM Analysis.....	81
Table 4.5. Log-Linear Models for Log Suspensions in Maryland with Race as Level-1 Predictor.....	83
Table 4.6. Log-Linear Models for Log Suspensions in Maryland with Disability as the Level-1 Predictor.....	86
Table 4.7. Variance Explained from HGLM Models for Log Suspensions in Maryland when Race and Disability were entered as Level-1 Predictors.....	90
Table 4.8. Discriminant Function Analysis Predicting Race by Offense Resulting in Suspension.....	93

Table 4.9. Discriminant Function Analysis Predicting Disability by.....95
Offense Resulting in Suspension

LIST OF FIGURES

Figure 4.1. Number of students suspended per 100 students from.....	69
1995 to 2004: Maryland	
Figure 4.2. Number of students suspended by disability from.....	70
2000 to 2004: Maryland	

CHAPTER I: INTRODUCTION

During the past decade, public school disciplinary policies have been changed to respond to concerns about school safety. High profile school shootings and media coverage of those incidents have created the perception that many schools are unsafe (Brooks, Schiraldi, & Zeidenberg, 2001). The passage of the Guns Free Schools Act (1994) and provision of discretionary federal grants to schools to improve safety have led to the implementation of “zero tolerance” policies in response to serious student misbehavior (Brady, 2001). One significant problem with a zero tolerance posture is that serious punishments (i.e., suspensions from school) have been handed out in an arbitrary manner (Harvard University Civil Rights Project, 2000; Skiba & Peterson, 1999; Tebo, 2000). Suspensions are often meted out for minor offenses such as tardiness, absence, disrespect, and non-compliance (Skiba & Knesting, 2001). There is no evidence to suggest that disciplinary removal was applied in a rational and consistent manner before zero tolerance policies, but prior to zero tolerance policies school disciplinary practices allowed for flexibility in responses to disciplinary problems.

Disciplinary Removal

School disciplinary removal serves two important functions for schools: (a) ensure the safety of staff and students, and (b) create an environment conducive to learning (Gaustad, 1992). Serious student misconduct including weapons offenses, drug offenses, and physical attacks may interfere with these

objectives and may require removal from school. However, the most common discipline problems resulting in suspensions involve non-criminal violations of rules including disruptions, inappropriate language, and unexcused absences (Brady, 2001; Losen, Simmons, Staudinger-Poloni, et al, 2003; Moles 1989). These minor offenses do not appear to interfere with the aforementioned objectives, and school removal may not be required depending on the manner and context of the behavior. Nonetheless, under zero tolerance policies, school officials are required to use school removal for both serious and non-serious infractions regardless of the contextual factors (Brady, 2001; Skiba & Knesting, 2002). There is no evidence showing that schools are safer as a result of strict disciplinary policies; however, there is evidence of an increasing trend in overall suspension rates since the implementation of zero tolerance policies in schools (Imich, 1994; Zhang & Katsiyannis, 2005). Furthermore, evidence suggests that current disciplinary policies have had a disproportionate impact on minority youth and students with disabilities (Cooley, 1995; Civil Rights Project of Harvard University, 2000; Skiba & Knesting, 2002; Zhang, Katsiyannis, & Herbst, 2004).

Race, Disability, and Exclusion

Although disciplinary practices exclude students across racial and ethnic groups, they are of particular concern for African American students who continue to be disproportionately suspended and expelled (Cooley, 1995; Krezmien, Achilles, & Leone, 2006; Skiba, Michael, Nardo, & Peterson, 2002).

Additionally, students with disabilities appear to be at greater risk for disciplinary procedures than their peers without disabilities (Cooley, 1995; Krezmien, Achilles, & Leone, 2006; Leone et al., 2000; Zhang, Katsiyannis, & Herbst, 2004). Disciplinary provisions under the Individuals with Disabilities Education Improvement Act of 2004 (IDEIA) and accompanying regulations are designed to ensure that a student with a disability has protections from disciplinary removal from school if the behavior resulting in a suspension is a manifestation of a student's disability. As a result, the current legislation as well as previous versions of the law has led to a perception that students with disabilities are less likely to be suspended than their peers without disabilities (Morrison, Anthony, Storino, et al., 2002). Some critics believe that special education rules and regulations have tied school principals' hands with regard to discipline and students with disabilities (Hymowitz, 2000).

However, the IDEA 1997 and the regulations of the IDEIA (2004) provided several options for responding to disciplinary problems exhibited by students with disabilities. Principals can unilaterally remove special education students involved in weapons or drug offenses and those at risk of harming themselves or others and place them in interim alternative programs (Bear, 1999). Administrators also have the ability to use short-term suspensions of special education students for serious or minor infractions. In 2001, the U.S. General Accounting Office (GAO) examined school discipline in the context of IDEA

regulations. The GAO reported that students with disabilities were disciplined in a manner similar to other students by school administrators and that 74% of the administrators they surveyed indicated that special education regulations had a neutral or positive effect on their ability to discipline students (GAO, 2001). The perception that schools are unable to equitably discipline students with disabilities reported by Hymowitz (2000) is not supported by available evidence (GAO, 2001).

Consequences of School Exclusion

It is important to understand the impact school exclusions have on students, and to investigate whether the rates of suspensions should be a concern for general educators and special educators as well as parents and policy makers. The current policies designed to meet troubling behavior with harsh punishments have been ineffective for reducing or eliminating the behaviors, and may exacerbate the problems they are designed to punish (Leone et al., 2003). Nonetheless, zero tolerance policies continue to dominate public school disciplinary policies despite an almost complete lack of documentation to support their effectiveness (Skiba & Peterson, 1997).

Costenbader and Markson (1994) found that 40% of school suspensions are delivered to repeat offenders, suggesting that suspension is ineffective for those students for whom it is most commonly prescribed. This finding is particularly problematic for special educators who are responsible for promoting

prosocial behaviors and eliminating troubling behavior through sustained and systematic behavioral interventions. Exclusionary practices that fail to improve behavior may actually inhibit the effectiveness of special education behavioral programming because they remove students from necessary behavioral and educational services and because they interrupt sustained service delivery. For some students, exclusions may accelerate the course of delinquency by decreasing educational opportunities and increasing occasions to associate with deviant peers.

School exclusions put children at risk for a host of negative social outcomes. This is one of the reasons that the No Child Left Behind Act of 2002 (NCLB) requires suspensions and expulsions to be reported to the federal government, and these reports are used to determine which schools meet criteria for persistently dangerous schools. Despite high expectations for schools to decrease suspension rates under the mandates of the NCLB (2002), increasing numbers of students continue to be suspended for a number of infractions. As a result, more students who are excluded from school and have an increased likelihood to engage in delinquent activity, experience academic failure, and drop-out (Leone et al., 2003; Losen et al., 2003; Skiba & Rausch, 2006). These failures place them at great risk for involvement with the juvenile justice and the criminal justice systems (Leone et al., 2000). Although public schools are not responsible for the underlying risks associated with negative outcomes, they can ameliorate or

exacerbate the vulnerability of children to those negative outcomes (Leone et al., 2003).

Suspensions: Impact of Individual and School Factors

Disproportionate exclusions of minority students and students with disabilities have been consistently documented over the past decade (Zhang, Katsiyannis, & Herbst, 2005). However, researchers have been unable to understand how the factors associated with disproportionate discipline contribute to the disparity in treatment or how the factors interact across individuals and school levels. Most of the current research in this area has involved the examination of individual factors associated with disproportionate school exclusion. These studies have demonstrated an overrepresentation of African American students (Costenbader & Markson, 1998; Raffaele Mendez, Knoff, & Ferron, 2002; Skiba & Peterson, 1997; Zhang et al., 2005), and students with disabilities (Cooley, 1995; Skiba et al, 2002; Zhang et al., 2005) in the suspension roles of schools, school districts, and states. Fewer researchers (Bruns et al., 2005; Christle, Jolivette, & Nelson, 2005; Skiba et al. 1997; Skiba & Peterson, 2002) have examined the relation between school characteristics (e.g., attendance rate, teacher quality, and SES) and disproportionate suspensions of minority students and students with disabilities.

Researchers have typically employed two quantitative approaches to investigating disproportionate suspensions of minority students and students with

disabilities. One group of researchers has examined the individual level characteristics that are associated with risk for suspension while the other group has examined the school factors.

Individual-Level Models

Quantitative researchers who examine suspension practices at the individual level have investigated the association between student characteristics and risk for suspension (Costenbader & Markson, 1998; Fasfo, Grubb, & Osborne, 1995; Losen et al., 2003; McFadden & Marsh, 1992; Skiba, Peterson, & Reece, 1997; Zhang, Katsiyannis, & Herbst). These researchers did not typically provide a theoretical basis for their investigations. Rather, the researchers examined the existence or the magnitude of disproportionate suspensions of minority students or students with disabilities, not the underlying causes of the phenomena. In one sense, the approach is representative of basic research rather than applied research.

Most of the researchers employed correlational models to identify the individual-level variables associated with suspension rates or to understand the strength of the association between individual predictors and risk for suspension. Although the quality and methodological rigor of these investigations varied, all of the researchers relied on an assumption that the association between individual characteristics and risk for suspension is independent from school-level factors. Skiba and colleagues (2002) suggested that identifying bias against specific

groups through quantitative approaches is difficult because sources of bias are difficult to identify and measure. Nonetheless, they argued that bias can be verified if all other explanations of disproportionate suspensions can be systematically eliminated as underlying causes.

The authors of two additional, non-quantitative articles proposed causes for disproportionate suspensions of minority students or students with disabilities. Vavrus and Cole (2002) suggested that disproportionate discipline of minority students is a result of sociocultural factors within the classroom that influence a teacher's decision to remove a student from the classroom. This proposal was supported by Bullara (1993) who contended that racial and cultural differences between teachers and students result in mistreatment of minority students and the overrepresentation of minority students in school suspension rolls.

Vavrus and Cole (2002) suggested that the decision to suspend a student is a contextualized decision based on subtle race relations that cannot be addressed in school discipline policies. The authors conducted a qualitative investigation of suspension practices in a single urban high school and examined disciplinary approaches in two classrooms. At the time of the investigation, the school had implemented a zero tolerance discipline policy that mandated suspensions for serious and minor infractions. They found that suspensions were typically administered when a teacher was unable to manage incidental student behaviors (talking out, questioning without raising hands, and inappropriate attention

seeking) that the authors believed were culturally normative and poorly understood by minority students. The authors contended that the incidental behaviors of minority students were disciplined with suspensions, while the incidental behaviors of White students were efficiently managed by the teacher. As a result, the authors maintained that suspensions in the school were disproportionately administered to minority students, particularly African American students, because of the attitudes and shortcomings of the teacher. Similarly, students with disabilities often exhibit inappropriate behaviors that teachers are ill-equipped to manage (Bullara, 1993). Instead, teachers refer these students for disciplinary removal from school.

Although the views of Vavrus and Cole (2002) and Bullara (1993) are compelling, their proposals lack sufficient empirical support. Their investigations do not include the rich and thick descriptions of rigorous qualitative investigations (Huck, 2004). Additionally, these authors failed to consider the impact that the environmental and administrative characteristics of schools have upon the decisions and behaviors of students and teachers within the schools.

School-Level Models

The second group of quantitative researchers examined school-level factors as predictors of suspensions, investigating the association between school characteristics and risk for suspension (Bruns et al., 2005; Christle et al., 2004; Cooley, 1995; Rafaelle Mendez & Knoff, 2003; Rausch & Skiba, 2004). Like the

authors who researched individual-level models, these researchers employed atheoretical approaches to examine correlations between school characteristics and suspensions of minority students or students with disabilities. Nonetheless, all of the authors tested the assumption that disproportionate rates of suspensions are attributed to school-level factors, not individual-level factors. From this perspective, disproportionate suspensions of minority students and students with disabilities at the state level are due to disproportionate contribution of suspensions from schools with high rates of suspensions and (a) prevalence of minority students, and / or (b) an inability to adequately respond to the behavioral concerns of students with disabilities.

Those who have studied suspension practices at the school level proposed that poverty, low percentages of White students, and low teacher expectations were characteristic of high suspending schools (Christle et al., 2004; Flannery, 1997; Imich, 1994, Skiba et al., 1997). Christle and her colleagues (2004) found that high suspending schools relied primarily on exclusionary practices to maintain school safety and order while low suspending schools utilized school-wide behavioral intervention programs to promote appropriate prosocial behaviors. Flannery (1997) identified several school-level factors related to risk for suspension, including high student/teacher ratios, insufficient curricular and

course relevance, weak, inconsistent adult leadership, high suspension rates at schools that have high rates of minority students, and limited academic opportunities for students.

Although none of the authors of the school-level investigations articulated a theory that grounded their investigations, they relied on a premise that understanding characteristics of schools is integral to understanding risk for suspension. This premise is consistent with educational development models that hypothesize that an individual's cognition, affect, volition, and behavior develop as a result of transactions among the various components of the mind and environmental contexts. An extension of this model is that individuals do not develop in isolation; they develop in a variety of contexts in which the individual is in constant interaction (Bridge, Judd, & Moock, 1979; Bronfenbrenner, 1977; Bronfenbrenner, 1989; Lee, 2000).

One of the most important contexts contributing to the development of human learning and behavior is the school. The nature of the relations between children and schools changes over time. Unlike young children who are typically nested in a single classroom with a single group of class peers, adolescents are exposed to multiple specialized teachers and a more diverse group of classmates (Lee, 2000). Although the learning context of adolescents is complex and varied, Lee (2000) proposed the school itself, not the classroom, is the appropriate organizational unit able to define the major educational context for adolescents,

and that hierarchical linear modeling is most appropriate for understanding educational outcomes for adolescents. In this investigation, I have extended the framework proposed by Lee to include the behaviors that result in suspensions. I believe that the context of the school contributes to the behaviors of adolescents, and that school characteristics will impact the extent to which students are suspended.

The Multilevel Approach

One problem with the current lines of school discipline research is that researchers have not examined the complex ways that the interactions between individual characteristics and school factors affect patterns of suspensions. Most researchers have either examined data at the individual level (usually aggregate data from school districts or states) or at the school level. A limited number of researchers (Bruns et al., 2005; Christle et al., 2004; Skiba et al., 1997; & Skiba & Peterson, 2002) have examined suspension practices data at both the individual level and the school level, but the authors have not utilized multilevel procedures in their analysis of data. These authors employed procedures that measured disproportionate suspensions using consecutive analyses with the same data set, relying on an assumption that the individual characteristics are independent from the school characteristics.

In this investigation I have examined the suspension practices in Maryland using a multilevel approach. Specifically, I have employed the theoretical and

methodological framework proposed by Lee (2000). I examined the risk of suspension for minority students and students with disabilities using a two level hierarchical model that includes individual characteristics as the level-1 predictors of suspension and school characteristics as the level-2 predictors. The multilevel analyses employed in this investigation allow me to determine the extent to which school factors and individual factors contribute to risk of suspension for youth in a single analytical model.

Purpose

The existing research has identified a number of individual and school characteristics related to disproportionate suspension practices. However, there has been limited research examining the complex interactions of these factors across various levels. The purpose of this investigation was to understand the multi-level factors associated with disproportionate suspension rates in Maryland. This research builds upon the findings from the pilot investigation for this study (Krezmien, Achilles, & Leone, 2006) that found that African American students and students with disabilities were disproportionately suspended in Maryland. The study presented here reports how factors at two levels of education (individual and school) interact to explain which factors or combinations of factors are most predictive of disproportional suspensions of

minority students and students with disabilities. Specifically, I am interested in determining if race and disability status are predictive of risk for suspension when school characteristics have been accounted for.

Odds ratios. The primary analyses used in this investigation involved the use of the logistic regression model. A logistic regression equation yields odds ratios. The odds ratio is the increase or decrease (if the ratio is less than 1.0) in the odds of being in the outcome category when the value of the predictor increases by one unit (Tabachnick & Fidell, 2001). Odds ratios that are greater than 1.0 indicate that members in the group have an increased odds (or likelihood) of the outcome. Odds ratios that are less than 1.0 indicate that members in the group have a decreased odds (or likelihood) of the outcome. In this investigation the odds ratio represented the odds of being a suspended student (in Analysis 2) or the odds of being suspended (in Analysis 3). For example, if the odds of being suspended for a student in Group A is 1.5, the odds for an individual in Group A is 1.5 times the likelihood of an individual in the comparison group.

Research Questions

This investigation will be organized around several research questions related to suspension practices in Maryland.

Question 1: What are the current suspension patterns in Maryland?

Question 2: What were the odds of being a suspended student by race and by disability without controlling school factors?

Question 3. What were the odds of being suspended by race and by disability status when school characteristics were controlled?

Question 4: What characteristics of schools were associated with odds of being suspended when race and disability were controlled?

Question 5: How were offenses resulting in suspensions predictive of race and disability in the State when school factors were not controlled?

Definition of Terms

Disability - As defined by IDEA, the term "child with a disability" means a child: with mental retardation, hearing impairments (including deafness), speech or language impairments, visual impairments (including blindness), serious emotional disturbance, orthopedic impairments, autism, traumatic brain injury, other health impairments, or specific learning disabilities; and who, by reason thereof, needs special education and related services.

Disproportionate – When the proportion of one group that has been suspended is substantially different from the proportion of another group that has been suspended. In this investigation, disproportionate is a statistical difference in the proportions.

EBD – Emotional and behavioral disorder as defined by the IDEA of 2004. It refers to a condition in which behavioral or emotional responses of an individual in school are so different from his/her generally accepted, age-appropriate, ethnic, or cultural norms that they adversely affect educational performance in such areas as self-care, social relationships, personal adjustments, academic progress, classroom behavior, or work adjustment.

Enrollment – The total number of students enrolled in a school, school district, or state. Enrollment numbers are based on the initial enrollment numbers at the beginning of the school year.

Exclusion – Any removal from school for a disciplinary infraction.

Expulsion – Permanent removal from school as a consequence for a disciplinary infraction. According to the Maryland State Department of Education, expulsions are indirect measures of student misbehavior. They represent the response of the school to the behavior.

Infraction – A violation of a school rule.

Individual Factors – Characteristics of individuals (e.g., race, disability).

Levels of Analysis – One of two types of characteristics investigated in this study. The individual level includes characteristics of individual students.

NCES – The National Center for Education Statistics. The primary federal entity for collecting and analyzing data related to education.

Offense – Any infraction at a public school that results in a disciplinary action.

Race – One of two racial groups of students defined by the Maryland State Department of Education. Includes White and African American.

Referral – Any time a teacher or school staff sends a student to the school disciplinarian for an infraction.

School Factors – Characteristics of school (e.g., school performance on State assessments, average daily attendance rate).

State Report Card – An annual state report that includes information about school performance, school demographics and the like.

Suspension -Temporary removal from school as a consequence for a disciplinary infraction. According to the State Department of Education, suspensions are indirect measures of student misbehavior. They represent the response of the school to the behavior. Suspensions refer to any disciplinary removal from school. They do not represent expulsions.

Zero Tolerance -The policy or practice of not tolerating undesirable behavior. With regards to school discipline, zero tolerance refers to the use of automatic school exclusions for both minor and serious disciplinary infractions.

CHAPTER II: LITERATURE REVIEW

School discipline is an essential component of school safety. However, considering the impact that removal from school has on at-risk children and youth, understanding how disciplinary practices disproportionately affect minority students and students with disabilities may be critical for developing more effective responses to deviant behaviors in school. Understanding the factors that underlie disproportionate disciplinary practices requires a strong foundation of empirically validated research that documents how school exclusions are applied and how they affect specific groups of children. Furthermore, researchers must identify the factors that contribute to disproportionate suspensions and expulsions in schools.

The purpose of this review of the related literature is two-fold. First, I will examine the findings of the related research and examine the adequacy of the research base. Second, I will examine the quality of research conducted in the area of school discipline among populations of children and adolescents. This chapter evaluates fifteen studies for methodological rigor and places the major findings in the context of disproportionate disciplinary exclusions in school. The rationale and theoretical basis, participant descriptions and selection procedures, descriptions of predictor and criterion variables, presence and adequacy of variable descriptions, use of appropriate statistical procedures, and interpretation of findings of each study is briefly discussed. Major findings

and the implications for practice and future research are also reported and discussed.

Methods for Review of Studies

The review of the literature was conducted using the Academic Search Premier, the Educational Resources Information Center (ERIC), the SocIndex, and the PsychINFO electronic databases from 1984 to 2004. The twenty-year time frame was selected in order to understand issues of disproportionate suspension practices prior to and after the implementation of zero tolerance policies in schools in the mid to late 1990s. However, the review did not identify any research studies conducted on the topic prior to 1992. Although some investigations of disproportionate exclusionary practices occurred prior to this time, it appears that rigorous examinations of this issue occurred substantially later. My search terms included school discipline, suspension, race, school exclusion, zero tolerance, disability, disproportionate, bias, expulsion, referrals, and special education. My searches were completed using all possible combinations and sequences of terms, but every search included the term school discipline or the term suspension. In addition to the electronic search of the databases, I also conducted a hand search of the journals that yielded at least one article for review. My initial search yielded 25 articles.

The articles were then examined to determine their adequacy for inclusion in the comprehensive review. The criteria used were: (a) publication in a peer

reviewed journal or as a report commissioned by a state government or the federal government, (b) report of descriptive or quantitative research studies, and (c) examination of disproportionate suspension rates among the sample population. Fifteen of the 25 studies met all of the criteria for inclusion in the methodological review (Table 2.1).

Findings

A major purpose of this review of the related literature was to explore the major findings and implications of studies that examined school suspension practices. Through my examination of the findings, I identified a number of individual and school factors that have a demonstrated relationship to disproportionate suspension practices. I incorporated these factors as independent variables in this investigation. In this section I report the results from prior investigations with regards to rates of suspensions, types of offenses resulting in suspensions, and differences in suspension rates by individual or school characteristics.

Suspension Rates

Authors of only two of the studies examined changes in suspension rates over time. Zhang et al (2005) reported an increase in overall rates of suspensions from 2001 to 2003. Rausch & Skiba (2004) found a similar increase in suspension rates over time, but found decreases in expulsions over the same period of time.

Table 2.1. Citation of References for Literature Review and Description of Purpose

Citation	Description of Purpose
Bruns et al., 2005	Investigate whether presence of school based mental health services in an urban school district were associated with suspension rates
Christle, Jolivette, & Nelson, 2004	Examine suspension rates in Kentucky Middle Schools
Constenbader & Markson, 1998	Investigate important variables associated with the population of students who have been suspended and perceptions of events.
Cooley, 1995	Examine whether acts leading to suspension or expulsion were different from those committed by other students.
Fasko, Grubb, & Osborne, 1995	Examine suspension rates of a school district in Eastern Kentucky. Determine differences by gender, race, disability, school level.
Losen, Simmons, Staudinger, Rausch, & Skiba, 2003	Explore the hypothesis that low teacher quality is an important predictor of a student's risk for suspension
McFadden & Marsh, 1992	Assess race and gender differences in the occurrence and treatment of school children's (a) rates of referrals, (b) types of violations, (c) types of punishments
RafaelleMendez & Knoff, 2003	Examine out-of-school suspensions in a large ethnically diverse school district by race, gender, school level, and infraction type
Rafaelle Mendez, Knoff, Ferron, 2002	Examined OSS in a large divers school district using quantitative and qualitative measures.

Table 2.1. Citation of References for Literature Review and Description of Purpose (Cont.)

Citation	Description of Purpose
Raffaele Mendez, 2002	Examine a) characteristics of students with differing rates of suspensions, b) elementary predictors for students who receive suspensions, c) how number of 6th grade suspensions lead to later school outcomes
Rausch & Skiba, 2004	Describe trends in state for OSS and expulsion in 2002, 203 school year.
Skiba & Michael, 2002	Explore the extent to which racial and gender referrals are artifactual or possible indicators of bias
Skiba, Peterson, & Reece, 1997 Study 1	Examine disproportional representaiion of youth in disciplinary referrals and suspensions.
Skiba, Peterson, & Reece, 1997 2	Study Examine referrals and suspensions in one middle school.
Zhang & Katsiyannis, & Herbst, 2004	Examine disciplinary exclusions in special education over four years to understand trends by minority and disability status

School Level

Authors of three of the studies (Rafaelle Mendez & Knoff, 2003; Rafaelle Mendez et al., 2002; & Rausch & Skiba, 2004) examined differences in suspension rates by school level. Authors of all three studies found that approximately 24 percent of middle school students were suspended during the years examined in the respective studies. They all found slightly lower suspension rates for high school students and low rates of suspensions in elementary schools.

Race

Authors of all but one of the studies reported differences in suspension rates by race. Fasko et al. (1995) found that White students were disproportionately suspended; however, the author reported less than one percent of the student population was non-White. This makes interpretation of this finding questionable. Authors of all but two of the studies (Skiba et al, 1997(1); & Zhang et al., 2005) reported that African American students were suspended more than students from any other racial group. Authors of both of these studies (Skiba et al, 1997(1); & Zhang et al., 2005) reported that American Indian students were as likely to be suspended as African American students. In contrast, Cooley (1995) found that American Indian students were less likely to be suspended than any group other than Asians. The small numbers of American Indian students in the samples in each of these studies limit the strength of the findings. Authors of six studies examined suspension rates of Hispanic students. Authors of five of these

studies (Cooley, 1995; Losen et al., 2003; Rafaelle Mendez et al., 2002; Rausch & Skiba, 2004; & Zhang et al., 2005) found that Hispanic students were more likely to be suspended than White students, while McFadden and Marsh (1992) reported that Hispanic students were no more likely to be suspended than White students. The small numbers of Hispanic students in the samples of all but one of these studies (Zhang et al., 2005) limits the strength of these findings.

Disability

Authors of five of the studies (Cooley, 1995; Fasko et al., 1995; Skiba et al, 1997(1); Skiba et al, 1997(2); & Zhang et al., 2005) examined suspension rates of students with disabilities. All of these authors found that the rates of suspension for students in special education were higher than the rates of students in general education. Authors of three of the studies (Cooley, 1995; Skiba et al, 1997(1); & Zhang et al., 2005) found that students with EBD were suspended more often than any other group of students. Cooley (1995) also found that students with LD were also suspended at higher rates than any other group except students with EBD.

SES

Authors of seven of the studies (Bruns et al., 2005; Christle et al., 2004; Rafaelle Mendez et al., 2002; Rafaelle Mendez, 2002; Skiba et al., 2002; Skiba et al, 1997(1); & Skiba et al, 1997(2)) examined the impact of SES on suspension rates. In each of the studies, students receiving free or reduced lunch was used as a proxy for SES. Authors of all but one of the studies found that low SES

increased the risk of suspension for all students. Rafaelle Mendez (2002) found that SES was negatively correlated with the risk for suspension, but the correlation was not significant. Additionally, Skiba et al. (2002) found that the effects of SES did not reduce the effects of race on suspension rates.

Achievement

Authors of five of the studies examined the impact individual student achievement on risk for suspension. Rafaelle Mendez (2002) found only a small negative correlation between student reading scores and risk for suspension. Rafaelle Mendez et al. (2002) found similar small negative correlations between student reading scores and student math scores and the risk for suspension. They found a moderate and significant correlation between student writing performance and risk for suspension. Authors of two of the studies (Christle et al., 2004; & Rausch & Skiba, 2004) examined the relationship between school achievement scores and whether the school was classified as a high suspending school or a low suspending school. Authors of both studies found that high suspending schools had substantially lower scores on state achievement tests than low suspending schools. However, the inappropriate use of univariate post hoc analyses by Christle and colleagues (2002) compromised their findings. Losen and his colleagues (2003) examined academic achievement at the state level. They found that states with the highest achievement scores on state assessments had the lowest rates of suspensions.

School Factors

Authors of five of the studies (Bruns et al., 2005; Christle et al., 2004; Cooley, 1995; Rafaelle Mendez & Knoff, 2003; & Rausch & Skiba, 2004) examined whether school factors affected risk for suspension. Authors of two of the studies (Christle et al., 2004; & Rausch & Skiba, 2004) compared the characteristics of high suspending schools to those of low suspending schools. As described previously, authors of both studies found that achievement had a significant relationship to suspending practices of schools. Christle et al. (2002) also reported that high suspending schools had higher numbers of low SES students, higher drop out rates, lower attendance rates, higher retention rates, and lower percentages of White students than low suspending schools. Cooley (1995) only examined the relationship between school size and suspension rates. He found that school size was not related to the patterns of suspension practices of school. Rausch and Skiba (2005) examined effects of school locale on suspension practices, and found differences in the rates of suspensions by locale. They reported that urban schools had the highest rates of suspensions, followed by suburban schools, town schools, and rural schools. Each of these findings should be viewed with caution considering the methodological problems noted earlier.

Bruns and Moore (2005) examined whether there were differences between a group of schools with a specialized mental health program and schools

without the program. They found that the presence of the mental health program was not predictive of suspension rates. They did find that schools with low attendance rates and high enrollments predictive of high suspension rates. Rafaelle Mendez et al. (2002) reported that the percent of a school population receiving free lunch, the percent of the population that was Black, and the school mobility rate were moderately positively correlated with rates of suspensions. They found that school performance on standard tests of achievement, percent of the school population that was White, and the percent of the school population that was Hispanic were moderately negatively correlated with rates of suspensions. When they examined secondary schools only, they found that the percent of the school staff that were new had a strong and significant correlation to suspension rates in the schools. They also found that writing achievement had a strong and significant negative correlation with suspension rates.

Interactions

Authors of six of the studies (Cooley, 1995; McFadden & Marsh, 1992; Rafaelle Mendez & Knoff, 2003; Rafaelle Mendez, 2002; Skiba et al., 2002; & Zhang et al., 2005) looked at the interaction effects of multiple independent variables on the dependent variables. Authors of four of the studies (Cooley, 1995; McFadden & Marsh, 1992; Rafaelle Mendez & Knoff, 2003; & Skiba et al., 2002) examined the interaction of race and gender. All of these authors found that African American males were suspended most often, followed by White males,

African American females, and White females. Rafaelle Mendez & Knoff (2003) also examined the interaction between race, gender, and school level. They found the same order of risk for suspensions at every grade level, but found that middle school students were suspended the most often, followed by high school students and then elementary students.

Zhang and colleagues (2005) examined the interaction between race and disability category. They found that Black students with disabilities were suspended more than other students with disabilities. They also found that Black students with emotional disturbance were twice as likely to be suspended as any other students.

Rafaelle Mendez (2002) examined the effects that the interaction of race, gender, special education status, and SES had on risk for suspension. She found that males in special education who received free lunch had the highest risk for suspension. She also found a tremendous overrepresentation of African American males in special education who received free lunches among those suspended. She found that 66.27% of these students were suspended, compared to 44.12% for White males with similar characteristics. She also found that Black girls receiving free lunch had more suspensions regardless of special education status. She also found that White males in general education who paid

for lunch were more likely to be suspended than Black males with the same characteristics. This result should be accepted with caution because of the small number of Black students in general education who paid for lunch.

Multilevel Analysis

None of the authors used a multilevel approach in their data analysis. This is problematic because the authors failed to account for variables in ways that school level and school district level factors may interact with individual characteristics. In education, students are members of schools, and school factors that impact individuals who are members of the school are best analyzed through multilevel analyses such as hierarchical linear modeling.

Summary of Findings

Authors of all of the studies reviewed in this chapter that examined racial differences in suspension practices reported that African American students were disproportionately suspended. Results from all of those studies indicated that African Americans were more likely to be suspended than any other racial group with the occasional exception of American Indian students. Results about other racial groups were inconsistent. Authors of all of the studies that examined gender differences reported that male students were more likely to be suspended than female students. All of the authors who examined differences in suspensions practices by disability found that students in special education were suspended more than students in general education. Additionally, students with EBD were

found to be suspended at much higher rates than all other students. Authors that examined interactions of individual characteristics consistently reported that African American males and in particular, African Americans in special education, had highest rates of suspensions.

Several studies examined school factors related to suspension practices. Authors of these studies consistently reported that schools with high percentages of students receiving free and reduced lunch, high percentages of African American students, and low attendance rates suspended more students than schools with low numbers of students receiving free and reduced lunch, low rates of African American students, and high attendance rates. Additionally, there was limited evidence to support that teacher quality, student performance on school assessments, per pupil expenditures were related to school suspension practices.

In this review I identified a number of factors that had a documented relationship to disproportionate suspension rates. I will include each of the available factors as predictor variables in this investigation. I also identified a number of factors that may be associated with disproportionate suspension practices, but lacked sufficient empirical support. I will also include these factors as predictors in this investigation to better understand their impact on disproportionate suspension rates.

Methodological Review

After I examined the studies to better understand the major findings, I reviewed each of the studies to establish the quality of the research design. Guidelines provided by Isaac and Michael (1997), Odom, Brantlinger, Gersten, Horner, Thompson, and Harris (2005), and Thompson, Diamond, McWilliam, Snyder, and Snyder (2005) were used to develop the framework for the evaluation of the studies. The critical factors for a methodologically rigorous investigation begin with the development of research questions and a rationale for the investigation that relies on a complete understanding of the literature in the field of study (Huck, 2004). Next, researchers should identify the relevant variable or variables, select appropriate subjects, select or develop the appropriate instruments for measuring the variables, and determine the appropriate type of analysis for the study (Isaac & Michael, 1997). After data are collected, researchers should conduct the appropriate tests of statistical assumptions of the analysis (Thompson et al., 2005; Tabachnik and Fidel, 2000). Finally, researchers should use proper statistical procedures, and report both clinical and practical significance of findings (Thompson et al., 2005).

Rationale, Research Questions, and Hypotheses

Each of the fifteen studies (See Table 2.1) included a well described and clear purpose. Authors in all of the studies provided a strong rationale for their research. Each of the authors included reviews of the related literature and

provided ample evidence of the dearth of information in the area of school discipline they examined. All of the authors included a description of how school exclusion practices can affect children and youth, and identified the need of the research community to better understand disproportionate disciplinary practices among various school populations.

Authors of eight of the studies (Bruns et al., 2005; Losen et al., 2003; Rafaele Mendez & Knoff, 2003; Rafaele Mendez, 2002; Rausch & Skiba, 2004; Skiba et al., 2002; Skiba et al, 1997(1); & Skiba et al, 1997(2)) explicitly described the research questions that guided their investigations. Authors of only two of the studies (Losen et al., 2003; & Skiba et al., 2002) clearly described their research hypotheses, and these authors reported both their research questions and their hypotheses. Authors of the remaining seven the studies failed to report either the questions or the hypotheses. Huck (2004) suggests that authors may not need to identify their hypotheses because operating with hypotheses may bias data collection and analysis. This is particularly true for correlational investigations that are often exploratory in nature. However, the inclusion of clearly stated research questions or hypotheses allows the reader to understand a researcher's line of inquiry. The failure of a number of the authors to clearly establish the research questions or hypotheses make their interpretation of the data questionable.

Sample Description

Comprehensive participant descriptions are critical to good research examinations. According to Huck (2004), the results of a study are meaningless unless the reader is provided with a clear description of the population from which the sample was drawn or the sample itself. Detailed descriptions facilitate replication, appropriateness of the intervention, and allow researchers to better understand the areas in need of further investigation (Mooney, Epstein, Reid, & Nelson, 2003). Since most of the articles I reviewed involved children who exhibited behavior problems, I used the guidelines established by Mooney, Epstein, Reid, & Nelson (2003) in their examination of the methodological issues of research of students with emotional disturbance. They report that authors should report sample size by gender, race or ethnicity, chronological age, grade level, locale, SES, and disability status. I also included criteria from Rosenberg et al. (1994) who suggest that authors also report measures of aptitude and achievement, and report the subtests of the measures as well as overall scores. I evaluated each article to determine which of these participant characteristics were reported. Table 2.2 displays the characteristics reported by authors of each of the investigations.

The level of the descriptive information of the studies varied. Authors of 10 of the studies reported the racial composition of the participant group (See Table 2.2). However, one of these studies included a sample that was over 99%

Table 2.2. Reports of sample size and presence of components required for sample description

Study	Sample Size	Sample description							
		Race	Gender	SPED	SES	Age	Grade	IQ	Ach.
Bruns et al., 2005	82 Schools	0	0	0	x	0	x	0	0
Christle et al., 2004	161 schools	0	0	0	0	0	x	0	0
Constenbader & Markson, 1998	4 schools	x	x	0	x	0	x	0	0
Cooley, 1995	441 principals	x	x	x	0	0	0	0	0
Fasko et al., 1995	3019 students	x	x	x	0	0	0	0	0
Losen et al., 2003	not included	0	0	0	0	0	0	0	0
McFadden & Marsh, 1992	4391 students	x	0	0	0	0	0	0	0
Rafaelle Mendez & Knoff, 2003	142 schools. 138,761 students.	x	x	0	x	0	x	0	0
Rafaelle Mendez et al., 2002	not provided	x	x	0	x	0	0	0	x
Raffaele Mendez, 2002	8268 students	x	0	0	0	0	x	0	0
Rausch & Skiba, 2004	not provided	0	0	0	0	0	0	0	0
Skiba et al., 2002	11,001 students	x	x	x	x	0	x	0	0
Skiba et al., 1997 - Study 1	11,001 students	x	x	x	x	0	x	0	0
Skiba et al., 1997 - Study 2	610 students	x	0	0	x	0	x	0	0
Zhang et al., 2004	All students with disabilities	0	0	0	0	0	0	0	0

x = present; 0 = not present

White, and the authors of one other study only reported the percent of students in the sample that were not White. The authors of the remaining five studies failed to include any information about the racial composition of the sample. Although three of these studies included nationally representative samples, the authors should have provided information about the racial composition of the school-aged population for the nation for the years examined. Authors of seven of the studies reported gender differences for all of the groups examined. Authors of six of the studies reported SES of the sample, each using the percent of the population receiving free and reduced lunch as a proxy for SES. Costenbader & Markson (1998) included a description of SES, but it was not clear whether the description was for the sample or for the community from which the sample was drawn.

None of the authors reported the mean age of the sample. This may be due to the difficulty in obtaining data from extant data sources. Due to the difficulty in obtaining ages of participants, grade level is an appropriate replacement for age. Grade levels of the participants were reported in eight of the studies (See Table 2.2), but none of the studies reported both the mean grade levels and standard deviations for all groups examined. The authors of just four of the studies reported the disability status of the participants of all groups included in the analyses. However, the authors of two of the studies only reported the presence or absence of a disability. Authors of the other two of the studies reported the disability categories of the sample. None of the authors reported scores from intelligence

instruments for the participants. The authors of only one study reported mean achievement scores for elementary and middle school students, but not for the high school students.

Summary. Most of the authors failed to include adequate participant descriptions. Authors of only two of the studies include more than half of the components of sample descriptions that I examined (See Table 2.2). Authors of four of the studies included half of the components in the participant descriptions. Authors of three of the studies failed to include information from any of the components in their description of the participants. The failure to include detailed descriptions clouds the meaning of the findings, and prevents replication. The shortcomings of a number of the studies reviewed negatively impact communities' understanding of disciplinary exclusions in schools. Methodologically rigorous studies must include detailed and comprehensive participant descriptions.

School and Setting Descriptions

Authors of all but four of the studies (Cooley, 1995; Losen et al., 2003; Rausch & Skiba, 2004; & Zhang et al., 2005) included descriptions of the setting examined. The authors of three of these studies used national data and did not have any setting data to describe. Christle and her colleagues (2004) failed to include a description of the setting from which their sample was drawn. This is problematic for replication and for interpretation of their findings. Authors of five

of the studies (Bruns et al., 2005; Rafaelle Mendez et al.; 2002; Skiba et al., 2002; Skiba et al, 1997(1); & Skiba et al, 1997(2)) included descriptions of the schools from which students were suspended. The authors of these studies provided descriptions of the school size, the school locale, and sufficient descriptions of the school. The authors of four of the studies (Cooley, 1995; Losen et al., 2003; Rafaelle Mendez & Knoff, 2003; & Zhang et al., 2005) did not examine school-level data, so school descriptions were not necessary. The authors of the remaining studies did not include adequate descriptions of schools included in their investigations. This is problematic considering the variation in schools and school characteristics across districts, states, and the nation. Only Cooley (1995) failed to report any information about the setting without providing an appropriate explanation.

Sampling Procedures

Sampling procedures are critical to research methodology because they impact the generalizability of the findings (Huck, 2004). To evaluate the adequacy of the sampling procedures, I looked at three questions that Huck (2004) regarded as critical to investigatory research: a) what is the relevant population, b) how was the sample extracted from the population, and c) what characteristics of the sample were measured. Authors of all but three of the studies reported the same characteristics in the population that they reported in the sample. Authors of two of the remaining studies (Losen et al., 2003; & Zhang et

al., 2005) used national data and included all available subjects in their studies, but they failed to report relevant information about the population. Cooley (1995) sampled all principals in Kansas, and all principals responded. However, the author failed to report information about the principals. Bruns et al. (2005) developed matched groups, and they clearly described the participant selection procedures.

Authors of three of the studies (Bruns et al., 2005; Costenbader & Markson, 1998; & Cooley, 1995) included volunteer participants. In these studies, participants were recruited in various ways and from a variety of settings. For example, Cooley (1995) recruited all school administrators in Kansas via standard mail. He got 100 percent response rate, so the sample was the same as the population. In contrast, Costenbader & Markson (1998) sampled students from three different schools. The authors surveyed all students, but did not report the response rate. Additionally, the authors failed to report the differences on variables of importance between participants and non-participants. According to Isaac and Michael (1997), there may be differences between participants and non-participants, and these differences should be examined and reported. Failure of the authors to report these differences threatens the generalizability of the findings. Bruns and Moore (2005) examined all students in two groups of schools. The schools were matched on percent of students in poverty, average attendance rates, and percent of the students who were White. The authors compared the

groups on these characteristics and found no differences. However, the schools within each group differed on a number of characteristics making analysis by school level problematic.

Adequacy of Variable Descriptions

Wilkinson (1999) stated that researchers should explicitly define the variables, demonstrate how they are related to the purpose of the study, and describe how the variables are measured. The quality and degree of variable description varied across the studies. Table 2.3 displays a brief description of the dependent variables and the independent variables examined in each of the studies. Authors of all of the fifteen studies included a description of the dependent variables. Authors of all but five of the studies (Cooley, 1995; Fasko et al., 1995; Losen et al., 2003; McFadden & Marsh, 1992; & Rausch & Skiba, 2004) also provided operational definitions of the dependent variables. The quality of the operational definitions also varied, but all were clear. The authors of the remaining five studies failed to provide adequate operational definitions of the dependent variable. The lack of operational definitions seriously impacts the interpretability of the findings (Gersten et al., 2005; & Huck, 2004).

Authors of all of the studies provided descriptions of the independent variables (See Table 2.3) and authors of all but 4 of the studies (Costenbader & Markson, 1998; Cooley, 1995; McFadden & Marsh, 1992; & Rausch & Skiba, 2004) also included operational definitions of their independent variables. The quality and

Table 2.3. Descriptions of variables of studies included in literature review

Study	Dependent Variable Description	Independent Variable Description
Bruns et al., 2005	Number of suspensions, Average length of suspension, Total suspension days	Enrollment, Attendance rate, Poverty rate, Percent non-white,
Christle et al., 2004	Suspension Rate	Board violations, Law violations, Enrollment, Attendance rate, Achievement scores (CTBS), Retention rate, Percent males, Percent Caucasian, Percent free lunch, Rop-out rate, Per pupil expenditure, Teacher student ratio, Average teacher salary
Constenbader & Markson, 1998	In-school suspension, Out-of-school Suspension	Race, Gender, School level
Cooley, 1995	Suspension, Expulsion, Reasons for suspension or expulsion	Race, Gender, Disability, Disability Category, Grade
Fasko et al., 1995	Suspensions	Race, Gender, School level
Losen et al., 2003	Suspension rate, Expulsion rate	4th and 8th grade achievement (math, science, writing), Percent of classes taught by teacher without major in that subject, Percent of classes taught by a teacher without a certificate in subject, Percent of classes taught by a teacher without major or certificate in the study, Percent of secondary teachers with less than 3 years experience

Table 2.3. Descriptions of variables of studies included in literature review (Cont.)

Study	Dependent Variable Description	Independent Variable Description
McFadden & Marsh, 1992	Referrals, Suspensions, Type of violations	Race, Gender
Rafaelle Mendez & Knoff, 2003	Unduplicated suspensions, Duplicated suspensions, Offenses	Race, Gender, School level
Rafaelle Mendez et al., 2002	Duplicated suspensions	Percent White, Percent Black, Percent Hispanic, Percent free lunch, Teacher absences, Teacher experience, Percent new staff, Enrollment, Percent capacity, Operating cost, Class size, Promotion rate, Mobility rate, Kindergarten readiness, Writing, Stanford reading, Stanford Math, Parent life involvement, Parent educational involvement, Parent conferences, Parent volunteers
Raffaele Mendez, 2002	Out of school suspensions Grade 7-8, Out of school suspensions Grade 9-12	Race; Gender; Self-Esteem; Early Delinquency; Reading Achievement; Math Achievement; Teacher Ratings of Behavior; School Adjustment (Grade 5); SES (FRL); SPED Status; Concerns About Middle School (Grade 6); Reading Achievement (7-8); Math Achievement grades 7-8; On-time graduation.

Table 2.3. Descriptions of variables of studies included in literature review (Cont.)

Study	Dependent Variable Description	Independent Variable Description
Rausch & Skiba, 2004	Suspension Rates, expulsion rates, types of incidents	Locale (urban, suburban, town, rural); School level (elementary, middle, high); Race (AA, W, H, AMIND, Asian, Multi); Percentage passing state assessments
Skiba et al., 2002	Referrals, Suspensions, Expulsions	Race, Gender, SES
Skiba et al., 1997 - Study 1	number of referrals, number of suspensions	gender, ethnic status, disability label, SES
Skiba et al., 1997 - Study 2	number of referrals, number of suspensions	gender, ethnic status, disability label, SES
Zhang et al., 2004	removal by school personnel, short term suspension, long term suspension	Region, Race, Disability type, State

specificity of the operational definitions varied, but all of the descriptions were clear. The failure of the authors to operationalize the independent measures makes interpretability of the findings problematic.

Data Collection

The authors of all fifteen studies provided clear and thorough descriptions of the data collection procedures. The clarity of the descriptions for data collection procedures allow for replication and verification of the findings.

Data Analysis

Descriptive designs. Three of the studies I reviewed included only descriptive analyses of the data. Fasko, et al. (1995) examined percentages of suspensions by race, gender, and school level. They compared percentages of within group rates of suspension across the various groups. This was appropriate for gender and school level factors, but was not appropriate for examining differences by race. More than 99% of the sample was White, making comparisons across groups impossible and making interpretability of findings spurious. Rausch & Skiba (2004) presented suspension data for one state. They were only interested in reporting percentages as part of an evaluation of State practices. They did not attempt to make comparative statements beyond those appropriate to descriptive designs. In another study, Skiba, Peterson, & Williams

(1997) reported the percentages of students suspended in one middle school. Like Rausch & Skiba (2004), they did not make any comparative statements beyond what was appropriate to the analysis.

Correlational designs. The authors of the remaining 12 studies utilized correlational approaches to analyze data. Results obtained from correlational research require a sophisticated process of data collection and analysis. Researchers should report all statistical procedures used and describe their findings in a clear and consistent manner. In order to obtain interpretable findings, researchers must take care to appropriately apply statistical procedures based upon their sample size and research questions, and report those procedures completely and with clarity (Thompson et al., 2005). I evaluated the articles based on the appropriateness of the sample, the use of appropriate descriptive statistics to clean the data, and the appropriate use of statistical procedures for analysis.

Sample size. Sample sizes varied across studies. I examined two important factors regarding the sample sizes. First, I looked at the equality of sizes in the different groups being compared by the investigators. Each of the studies examined disproportionate disciplinary rates across groups. One of the primary differences examined were across racial groups. Since most states, school districts, and schools are disproportionately White, it was important to understand if there were enough students in each racial group to make comparisons. Six of the studies (Cooley, 1995; Fasko et al., 1995; Rafaelle Mendez, 2002; Skiba et al,

1997(1); Skiba et al, 1997(2); Zhang et al., 2005) looked at differences between students in special education and students in general education. Students in special education typically represent 10 to 15 percent of the total student population in public schools. As a result, it was important to understand if there were enough students in the special education group to make comparisons. Second, I investigated the adequacy size of the sample and the size of the groups examined to determine if the appropriate statistical procedures were conducted. I examined each study to determine if they used appropriate statistical procedures to account for differences in group sizes.

Authors of three of the correlational studies (Bruns et al., 2005; Costenbader & Markson, 1998; Fasko et al., 1995; & Losen et al., 2003) failed to report the racial composition of the sample, but they did use race as a predictor. The failure to report makes replication difficult. Fasko et al. (1995) did not use race as an independent variable in the data analysis. This failure prevents any comparisons of suspensions by race, and would threaten the validity of any of the findings, but the authors reported that the school was nearly all White. Bruns et al. (2005) reported similar racial compositions in the two groups examined. However, the authors used percent of the population that was non-White as the only measure of racial composition. Considering that race is known to have a relationship to risk for suspension, and that the risk is different across racial groups, the authors' use of percentage of the sample that was non-White as a

proxy for racial composition threatens the interpretability of the findings. Authors of the remaining seven studies reported differences in the racial composition of the respective samples.

Authors of three of the correlational studies (Fasko et al., 1995; Rafaelle Mendez, 2002; & Skiba et al, 1997(1)) reported the special education status of the respective samples. In each of the studies, the special education students represented about 10 percent of the sample.

Authors of all but three of the correlational studies (Christle et al., 2004; Costenbader & Markson, 1998; & Rafaelle Mendez, 2002) used appropriate analyses or correction procedures to control for unequal groups sizes. Authors of five of the studies (Cooley, 1995; Losen et al., 2003; McFadden & Marsh, 1992; Rafaelle Mendez & Knoff, 2003; & Rafaelle Mendez et al., 2002) used nonparametric statistics. Authors of two studies (Skiba et al., 2002; & Zhang et al., 2005) used corrective procedures in their analyses. Authors of two studies (Skiba et al., 2002; Zhang et al., 2005) used multivariate analyses appropriate for comparing groups with unequal sizes.

Rafaelle Mendez (2002) used separate analyses for each racial group which is not a recommended practice (Huck, 2004). No rationale was provided for this approach. Costenbader & Markson (1998) failed to account for differences in group size. They ran an ANOVA to look at differences between groups although they had unequal cell sizes. They failed to report the follow-up tests they used, but

they did have some cells with no members. This is a flaw in the analysis, making interpretation of the results difficult. Christle et al., (2004) failed to report the unequal sizes of the groups as a problem, and failed to report how they accounted for this problem. The failure to address differences in group sizes compromised the findings of each of these studies.

Adequacy of statistical procedures. Methodologically strong research should include sample sizes that are appropriate to the research design and statistical procedures used to analyze the data. Wilkinson (1999) provided some guidelines for reporting information regarding the sample size. He indicated that researchers should report the sample size, a power analysis, and the analytical procedures used in power calculations. None of the authors using correlational designs reported power analyses. This is problematic for interpreting the adequacy of the statistical procedures.

Descriptive Statistics

All quantitative researchers should examine their data using descriptive procedures prior to analysis, and should demonstrate that their data meet the statistical assumptions necessary for the statistical procedures used to interpret the data. Tabachnick and Fidel (2001) suggested that researchers examine the data for normality of each of the variables, the presence of outliers, and group differences. They also suggested statistical procedures for controlling abnormal distributions of data, responding to outliers, and accommodating missing data. Researchers

should examine their data through descriptive procedures, report their findings, and report statistical procedures used to respond to aberrations in the data (Tabachnick & Fidell, 2001). I examined each of the articles for reports of descriptive procedures, reports of statistical procedures to normalize data, reports of outliers, and reports of statistical procedures for responding to outliers.

Authors of eight of the studies (Bruns et al., 2005; Christle et al., 2004; Rafaelle Mendez & Knoff, 2003; Rafaelle Mendez et al., 2002; Rafaelle Mendez, 2002; Rausch & Skiba, 2004; Skiba et al, 1997(1); & Skiba et al, 1997(2)) reported the use of descriptive procedures to examine the data prior to analysis. None of these authors described the procedures completely, but provided support for the use of the respective procedures. None of the authors discussed issues of normality of data distribution, and none of the authors discussed issues associated with statistical outliers. The failure to report and utilize normalization procedures is a methodological shortcoming that may inhibit interpretation of the findings. Authors of the remaining seven studies failed to report any descriptive procedures, a major methodological flaw.

Statistical Analysis

Each of the articles was reviewed to determine the adequacy of the statistical procedures used by the researchers. I looked at four factors to

determine the appropriateness or adequacy of the statistical procedures: a) the unit of analysis, b) multivariate procedures when necessary, c) univariate analyses when necessary, and d) clarity of the results.

Authors of all but two of the studies (Losen et al., 2003; & Skiba et al, 1997(1) reported and used the appropriate level of analysis in their investigations. The authors of the two remaining studies failed to report sufficient data to determine whether they used the appropriate unit of analysis.

Main effects. Authors of six of the studies (Bruns et al., 2005; Cooley, 1995; Losen et al., 2003; McFadden & Marsh, 1992; Rafaelle Mendez et al., 2002; & Skiba et al, 1997(1)) used the appropriate univariate procedures in their analysis of the data. Each of these studies included clear descriptions of the univariate procedure, and provided sufficient support from the research for utilizing the chosen procedure. Authors of two of the studies (Costenbader & Markson, 1998; & Rafaelle Mendez, 2002) did not use appropriate univariate statistics, and failed to explain the rationale for this decision. Fasko et al. (1995) failed to use any statistical procedure at all, but drew conclusions about group differences. Rafaelle Mendez et al. (2002) used multiple univariate procedures with multiple outcome variables. The failure to use multivariate procedures in a study that has multiple outcome variables is inappropriate unless the researcher can demonstrate sufficient cause for using univariate procedures (Thompson et.al., 2005, Tabachnick & Fidell, 2001). The use of univariate methods inflates

the possibility of Type I errors and negates the reality that outcome variables interact in unique ways that may affect statistical findings (Thompson et al., 2005). The failure of the authors to use appropriate multivariate approaches limits the interpretability of their findings.

Authors of four of the correlational studies (Christle et al, 2004; Costenbader & Markson, 1998; Skiba et al., 2002; & Zhang et al., 2005) examined multiple outcome measures, and used appropriate multivariate procedures in initial analyses. Zhang his colleagues (2005) only examined the main effects, and did not conduct any post hoc tests.

Follow-up procedures. Authors of three of the studies included follow-up procedures to further examine the main effects from the initial analyses. Univariate procedures should never be used in research that contains multiple outcome variables (Thompson et al., 2005). Thompson and colleagues (2005) contend that univariate procedures, often used as post hoc procedures to refine the researchers' understanding of statistical phenomena, is inappropriate. Instead, researchers should use descriptive discriminant analysis or a similar approach to describe multivariate dynamics. Of the authors who included appropriate multivariate procedures, only Christle and her colleagues (2004) inappropriately used univariate post hoc procedures. Authors of two of the studies (Costenbader & Markson, 1998; & Skiba et al., 2002) used appropriate multivariate procedures to explore multivariate dynamics.

Effect Sizes and Confidence Intervals

Thompson and colleagues (2005) argued that all manuscripts describing quantitative studies should include effect sizes. They also indicate that researchers should be careful to appropriately interpret the meaning of the confidence intervals for the reader. Authors of just three of the studies (Bruns et al., 2005; Skiba et al., 2002; & Zhang et al., 2005) reported effect sizes. The authors of these three studies described how the effect sizes were computed. The authors of two of these studies (Bruns et al., 2005; & Skiba et al., 2002) interpreted the effect sizes by comparing the findings to the findings of previous research on disproportionate treatment of school children. The failure of most of the authors to report effect sizes jeopardizes the adequacy and interpretability of the findings. The computation and reporting of confidence intervals of the effect sizes is necessary to inform judgment regarding the plausibility of the findings (Thompson et al., 2005). None of the authors reported confidence intervals of the effect sizes, a serious flaw in data analysis.

Limitations

Researchers should include accurate and appropriate reports of the limitations of their studies. I examined the articles and evaluated the extent to which limitations were described and whether the limitations included a description of the methodological flaws of the research. Authors of four of the studies (Cooley, 1995; Fasko et al., 1995; Losen et al., 2003; & McFadden &

Marsh, 1992) failed to report any limitations of their investigations. Authors of two of the studies reported several key limitations, but failed to address critical limitations of their respective investigations. Costenbader & Markson (1998) failed to explain the reason for using ANOVA when they had unequal cell sizes and multiple empty cells. Furthermore, they failed to address the problems associated with their failure to report response rates for their survey or their failure to report reliability of the instrument. Raffaele Mendez (2002) reported most of the limitations, but failed to identify why she chose to run separate analyses for White and Black students rather than using race as an independent variable and looking for differences between groups. Authors of the remaining nine studies described several of the key limitations of their study and addressed almost all of the major methodological problems I identified in this paper. However, none of the authors identified the failure to report descriptive procedures to examine and clean the data as a limitation.

Summary

The consistency of the findings with regard to individual characteristics is compelling. However, many of the studies had serious methodological flaws, making interpretation of the findings difficult. A number of studies also examined the relationship between school characteristics and school suspension practices. However, two of the studies examining school characteristics had numerous methodological flaws which seriously impacted the validity of the

findings. None of these studies examined whether school factors affected disproportional suspensions in school districts.

One of the major issues with school suspension research is the use of extant data. Often times the data are incomplete, the source of the data is questionable, and the available information about the participants is limited. As a result, investigators of suspension research must carefully describe the sources of data, and fully discuss the limitations of the research investigation. Additionally, investigators must use the appropriate research designs and employ statistical analyses appropriate to the design.(Thompson et al., 2004). In order to more clearly describe the methodological quality of the studies reviewed, I examined the degree to which the studies had problems specifically related to data collection and description and the degree to which the investigations had problems with design and / or analysis.

Authors of all but two of the studies had problems with data collection and study descriptions (Skiba & Knesting, 2002; Skiba et al., 1997). For all but five of the studies, the problems were serious enough to seriously impair interpretation of the findings (Costenbader & Markson, 1998; Rafaelle Mendez & Knoff, 2003; Rafaelle Mendez et al., 2002; Skiba et al., 2002; Skiba et al., 1997). All but five of the studies (Bruns et al., 2005; Rausch & Skiba, 2004; Skiba et al., 2002; Skiba et al., 1997; Zhang et al., 2004) had problems with design and / or data analysis. Only two of the studies (Skiba et al., 2002; Skiba et al., 1997) did not have

problems with either the data collection and description and the research design and analysis. These studies were methodologically strong investigations. The remaining studies had either serious problems with the data collection and / or descriptions or problems with the design or analysis. As a result, five of the studies (Bruns et al., 2005; Rausch & Skiba, 2004; Skiba et al., 2002; Skiba et al., 1997; Zhang et al., 2004) were determined to be methodologically strong, while ten had serious methodological shortcomings, with spurious results that should not be accepted.

None of the studies I reviewed examined if disproportionate suspensions of minority students and students in special education was due to direct bias on the part of school personnel or if disproportionate suspension was due to an indirect bias related to complex factors at the school and school district levels. Furthermore, none of these studies examined the way that multilevel factors influence each other in complex ways. Research in the area of school suspensions is needed to examine the cause and nature of biased treatment of minority students and students in special education. Additionally, studies should examine multilevel factors to determine how school factors and individual characteristics interact to promote or inhibit disproportionate suspension of youth.

CHAPTER III: METHOD

Data Collection

Data used in this investigation were drawn from four databases: State Enrollment Reports, State Reports of Suspensions and Expulsions, the State Report Card, and the National Center on Education Statistics (NCES) database. Specific types of data extracted are summarized in Table 3.1.

Participants

The demographic characteristics of the students enrolled in the state were obtained from the State Enrollment Report for the 2003-2004 school year. This report includes the number of students enrolled in the State public schools disaggregated by race, disability category (as defined by the IDEA), and the combination of race and disability category. There were 869,113 students enrolled in the State at the beginning of the 2003-2004 school year. The majority of the students were White (50.4%) with a large percentage of African American students (37.9%). 13.1% of the students were identified with a disability under the IDEA 1997. Males comprised 50.9% of the population, and females comprised 49.1%.

Demographic data for each school were obtained from the NCES database. This database includes reports of the number of students enrolled in each school by race, gender, and special education status. These data were electronically transferred into the main SPSS database by the investigator. The schools used for

Table 3.1. Sources of data and types of data extracted displayed by source and by level of analysis

	State Enrollment Report	State Suspension and Expulsion Report	State Report Card	NCES
Individual Level	Number of students in the State disaggregated by Race and Disability Category	Number of students suspended disaggregated by Race, and Disability Category Number of suspensions disaggregated by Race, Disability, and Offense		
School Level		Number of students suspended in each school disaggregated by Race and by Disability Offenses resulting in suspensions at each school disaggregated by Race and by Disability	Student enrollment at each school disaggregated by Race and by Disability MSA Basic, Attendance, Percent NHQ, Mobility, Percent SPED, Percent White	Student enrollment at each school disaggregated by Race and by Disability Student Teacher Ratio, SES, Percent White

the third and fourth analyses only included middle schools and high schools. There were 239,053 (60.6%) White participants and 155,666 (39.4%) African American participants included in the investigation. There were 48,944 (11.1%) participants identified with a disability under the IDEA of 1997, and 393,963 (89.9%) participants not identified with a disability included in the analysis.

Suspension Data

Suspension data were drawn from State Reports of suspensions and expulsions from the 2003-2004 school year. For the purposes of this investigation, a suspension refers to any disciplinary removal from school. This includes removal for one day as well as for multiple days. Suspension does not include expulsions which are permanent disciplinary removals from school. The data included the number of suspensions as well as the number of students suspended. Both short-term and long-term (more than ten days) suspensions were included in this investigation. The data include the number of suspensions disaggregated by five racial groups (i.e., White, African American, Hispanic, Asian, American Indian) and by disability. There were low percentages of Hispanic students, Asian students, and American Indian students in the population and in the sample. The small numbers were insufficient for the multilevel analyses utilized in this investigation.

Additionally, my prior investigation of suspensions in Maryland revealed that the primary differences in suspension practices were across the White and

African American groups. As a result, only African American and White students were used in the analysis for this investigation. The data also included the number of students suspended disaggregated by the same racial groups, by disability category (according to the IDEA definition), and by the combination of race and disability category. Consistent with a prior investigation (Krezmien, Leone, & Achilles, 2006), six of the categories of disabilities (Autism, Mental Retardation, Other Health Impairment, Emotional Disturbance, Learning Disability, Speech Language Disability) extracted from the State databases were used in this investigation. The remaining disability categories were combined into the Other category because they contributed minimally to the total numbers of students suspended in the prior investigation (Krezmien, Leone, & Achilles, 2006). These data were electronically transferred into the main SPSS database by the investigator.

School Characteristics

Information about the school characteristics were obtained from the State Report Card and from the NCES database. Information about the school performance on state assessments, school attendance rates, teacher qualifications at each school, student mobility at each school, percent of school enrollment that is White were obtained from the State report card. Data were transferred by hand from the State Report Card into the main SPSS database by the investigator. Information about school student-teacher ratios, percentages of students receiving

free and reduced lunch at each school, and the locale of the school were obtained from the NCES database. Those data were transferred electronically from the NCES database into the main SPSS database by the investigator.

School Data

There were 1,254 public primary, intermediate, and high schools in Maryland at the beginning of the 2003-2004 school year. For the purposes of this investigation, I examined the suspension practices of middle and high schools only. I did not investigate suspension practices in primary schools because those schools typically have very low rates of suspensions. The low rate of suspensions in primary schools has been documented by a number of researchers (Cooley, 1995; Christle et al., 2004). I also found very low suspension rates in the pilot investigation of suspensions in Maryland this study (Krezmien et al., 2006), and in my initial examination of the suspension data used for this investigation. I also eliminated 12 alternative schools from the analysis. These schools had small numbers of students, and special education students generally represented over 90% of the enrollment. The final analysis in this investigation included 405 schools.

Data Accuracy

To verify the accuracy of the data transfer, data were checked on all of the schools. After the main database was completed, two trained graduate research assistants (GRA) compared the information in the database to the

information from the State Report Card. Each GRA marked each item that was incorrectly transferred from the State Report Card with a slash and recorded the correct information on the datasheet. The investigator and each GRA met to confirm the inaccurate items, and the investigator corrected the items in the main database. Additionally, identical fields of data from multiple databases were checked for consistency across databases. There were no errors across databases.

Variables

Three criterion variables were included in the analyses. The first criterion variable was Students Suspended and was used in Analysis 1 and Analysis 2. Students Suspended represents the number of students suspended, not the number of suspensions. Students suspended more than one time were only counted once. The second criterion variable was Suspensions and was used in Analysis 1 and Analysis 3. In the analysis, suspensions represented the number of suspensions, not the number of students suspended. Students suspended more than once were counted as additional suspensions. The third criterion variable was Offense and was used in Analysis 4. Offense represented the type of school infraction that resulted in a suspension. Offense is related to the number of suspensions, not the number of students suspended. Students suspended more than once were counted as additional offenses and may represent more than one category of offense.

Predictor Variables

There were two levels of predictor variables included in this investigation. Race and Disability are the predictor variables for the individual level of analysis. Table 3.2 displays the predictor variables for the school level of analysis. Each of these predictors is a variable that was included in at least one other investigation. Table 3.2 provides a brief description of each of the variables which are described more fully below.

MSA Basic. MSA Basic represents the percentage of students at a specific school that did not meet proficiency on the Maryland State Assessment in mathematics. Mathematics was used because it was collected in each of the school levels. I used the percentage for the highest grade within each school level. For middle schools, I used the MSA Basic for eighth grade students and for high schools I used the MSA Basic for 10th grade students,

Attendance. Attendance represents the average daily attendance for each school included in the investigation. Attendance was reported by the MSDE, and represented the average daily attendance divided by the number of students enrolled.

Percent NHQ. Percent NHQ represents the percentage of the teachers at each school who were not highly qualified under the NCLB Act of 2001.

Table 3.2. Predictor variables at the school level of analysis

Variable	Description
MSA Basic	School performance on State Assessments. Represented as the percent of student enrollment performing below proficient levels as measured by the State.
Attendance	Average yearly attendance rate for the school as reported in the State Report Card.
Percent NHQ	Percent of classes at the school that are not taught by highly qualified teachers as defined by the State.
Mobility	Percent of the initial enrollment at the school that withdrew from the school prior to the end of the school year.
Enrollment	The total student enrollment at the beginning of the school year.
Suspension Rate	The number of suspensions in each school divided by the total enrollment in the school
Percent SPED	The percent of the enrollment that had an identified disability under the IDEA
Student-Teacher Ratio	The number of students in the school divided by the number of teachers in the school.
SES	The percent of the school receiving free or reduced lunch.
Percent White	Percent of the total school enrollment that is White.

According to the MSDE, highly qualified status is provided to teachers who have successfully met the requirements under the NCLB Act of 2001, and who have filed all necessary paperwork with the MSDE.

Mobility. Mobility represents the variability of the school population due to student movement out of schools. For the purposes of this investigation, I used the percent of the students initially enrolled in the school that withdrew from the school before the final day of the school year.

Enrollment. Enrollment represents the number of students enrolled in each school on the first day of school.

Suspension Rate. Suspension rate represents the number of suspensions in each school divided by the enrollment of the school on the first day of school. Suspension rate is based on suspensions, not the number of students suspended. As a result, students suspended more than once are counted multiple times.

Percent SPED. Percent SPED represents the percent of the students enrolled that have an identified disability under the IDEA of 1997.

Student-Teacher Ratio. The Student-Teacher Ratio represent the ratio of students to teachers at each school. This number was reported by the MSDE, and was calculated by using the number of teachers and the number of students on the first day of school.

SES. SES represents the socioeconomic status of the students enrolled at each school. SES represents the number of students receiving free or reduced lunches at each school divided by the enrollment at that school.

Percent White. Percent White represents the racial composition of each school. The percentage of the population that was White was used because White is the predominant racial group in the states enrollment. Percent White represents the percentage of the enrollment of each school that is White according to the MSDE definition.

Design and Data Analysis

This investigation answered the research questions using correlational statistical analyses of extant data. The questions were answered using four separate analyses of the data.

Analysis 1

The first part of this study involved a descriptive analysis of the suspension rates in the State. Numbers of suspensions disaggregated by race and by disability and the numbers of students suspended disaggregated by race and disabilities were reported. Number of students suspended per 100 students and number of suspensions per 100 students were calculated and reported to show differences in rates of suspensions by race and by disability (Rafaelle Mendez, 2002; & Skiba, et al., 2002).

Analysis 2

In the second analysis, the logistic regression model was used to examine disproportionate suspension rates for students by Race, Disability, and the combination of Race and Disability. Unduplicated suspensions were used for the logistic regression analyses because these data were the only numbers disaggregated by Race, Disability Category, and the combination of Race and Disability Category in the State. This analysis did not include any of the school level factors considered in Analysis 3 and Analysis 4 because these data were not disaggregated at the school level or at the district level. Analysis 2 only reports information at the individual level across the entire state. Suspension was the criterion variable (0 = not suspended, 1 = suspended). Race by disability was a categorical variable and was entered as a predictor. Each disability type for each racial group had a unique category. For instance, White students with mental retardation were a distinct category with a unique code. There were a total of 40 categories representing each disability category for each racial group. White students with No Disability were the reference group because it represented the largest group in the population. In the model, each of the categories was compared to the White students in the No Disability category.

Analysis 3

In the third analysis, hierarchical generalized linear modeling (HGLM) was employed to determine if there were differences in rates of suspensions by

race and by disability and to understand the factors related to differing rates of suspensions. Traditional statistical techniques used in prior investigations (e.g., linear regression, logistic regression, MANOVA, non-parametric statistics) examined individual characteristics related to suspension practices. Such techniques are unsuitable for addressing the multi-layered quality of disciplinary practices examined in this investigation because they do not correctly account for effects of individual-level variables that vary according to contextual and / or organizational factors (Pardoe & Weidner, 2004, Raudenbush & Bryk, 2002). Furthermore, the error variance obtained through multiple regression analyses are from the individual level of analysis, and are not interpretable at the school level. To properly account for covariates having a multilevel nature such as those in this investigation, hierarchical modeling is appropriate.

Two separate analyses were conducted for the two separate individual predictor variables (Race and Disability). Suspension was the criterion variable (0 = not suspended, 1 = suspended). In each of the analyses, the individual predictor was the only predictor in the first level of analysis. In the first analysis, Race was the predictor (0 = White, 1 = African American). For Race as a predictor, White was the reference category because it is the largest group in the State (Hosp & Reschly, 2003). In the second analysis, Disability was the predictor (0 = No Disability, 1 = Disability). For Disability as a predictor, No Disability is the

reference category because it is the larger group. The second level included the 10 proposed predictor variables at the school level of analysis (See Table 3.2).

The HGLM analysis involved an examination of four different regression models. I generated an unconditional model that displayed the percent of unexplained variance in the number of suspension for the two groups across schools without any predictors. The unconditional model allowed me to understand the proportion of variance accounted for when no predictors were included in the model (Raudenbush & Bryk, 2002). I also generated a model that included only the first level predictor. This model displayed the percent of unexplained variance in the number of suspensions for the two groups when only the individual predictor was included in the model. This model allowed me to understand the proportion of variance explained when only the level-1 predictors were included in the model. I then generated a unit specific and a population specific model that included all of the level 1 and level 2 predictors. The two models are described more fully in the Results.

Analysis 4

In the fourth analysis I employed discriminant analysis to explore the extent to which the types of offenses resulting in suspensions differed for White and African American students and for students with and without a disability. The sample was identical to the sample used in Analysis 3. The grouping variables were Race (0 = White, 1 = African American) and Disability (0 = No Disability,

1 = Disability). The response variables were the six categories of offenses resulting in a suspension.

CHAPTER IV: RESULTS

Analysis 1

Analysis 1 involved an examination of statewide enrollment and suspension records to understand the number of students suspended and the number of suspensions by race and by disability. Figure 4.1 displays the trend in the number of students suspended per 100 students for White students and for African American students from 1995 to 2004. The trend line for White students was stable across the ten year period, with approximately six White students per 100 students suspended each year. The trend line for African American students increased over time from approximately eight African American students per 100 students suspended in 1995 to more than 14 students per 100 students suspended in 2004. Figure 4.2 displays the number of suspensions per 100 students for students with disabilities and students without disabilities from 2000 to 2004. The trend lines for both groups increased over time, but the number of suspensions per 100 students was higher for students with disabilities each of the five years.

Table 4.1 displays the enrollment, number of students suspended and the number of suspensions for all public school children during the 2003-2004 school year. About 9% of the population was suspended at least one time. The table shows that African American students have been suspended at more than twice the rate of White students. The table also shows that students with disabilities were suspended at nearly twice the rate as students without disabilities.

Figure 4.1. Number of students suspended per 100 students from 1995 to 2004: Maryland

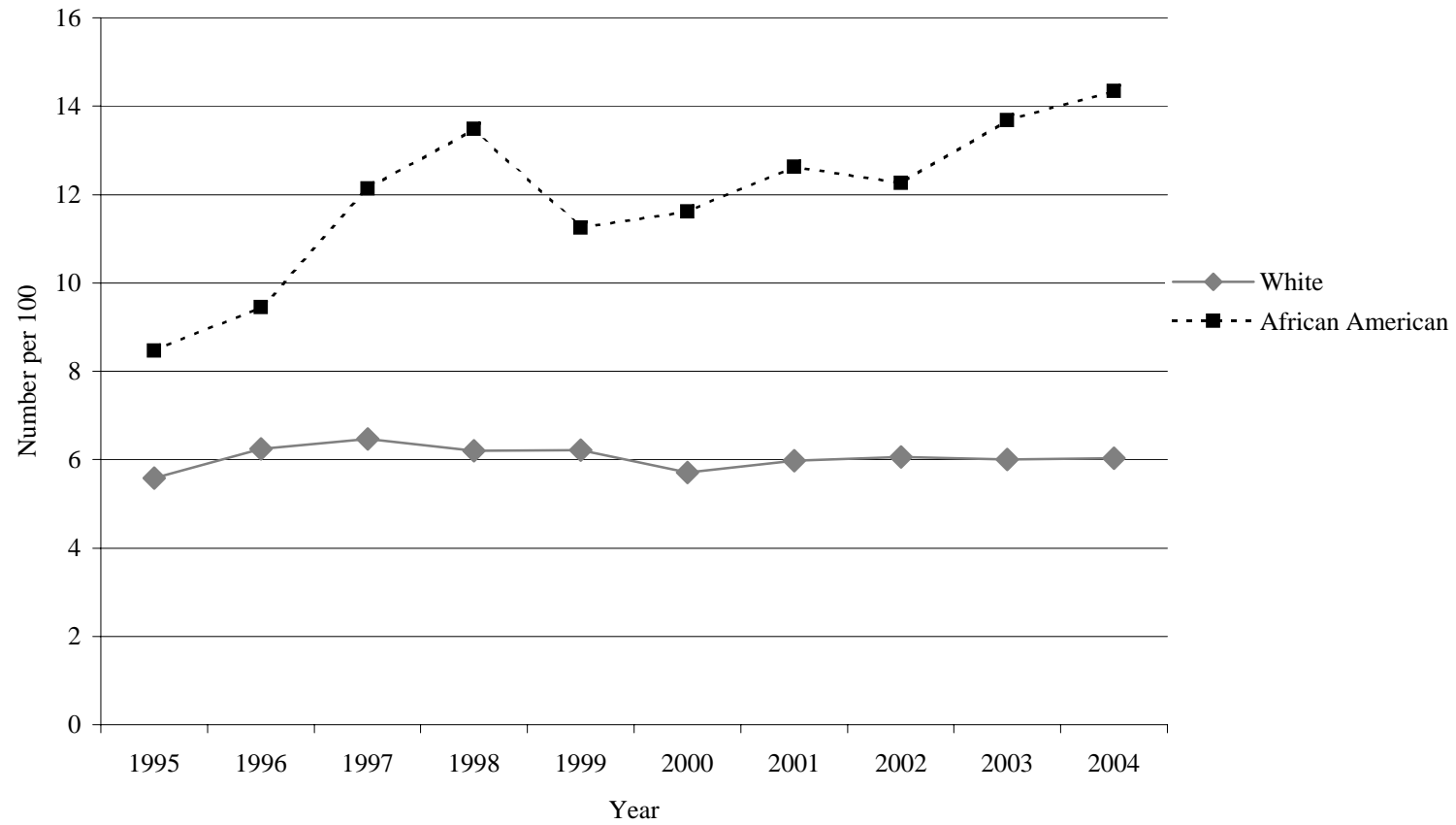


Figure 4.2. Number of students suspended by disability from 2000 to 2004: Maryland

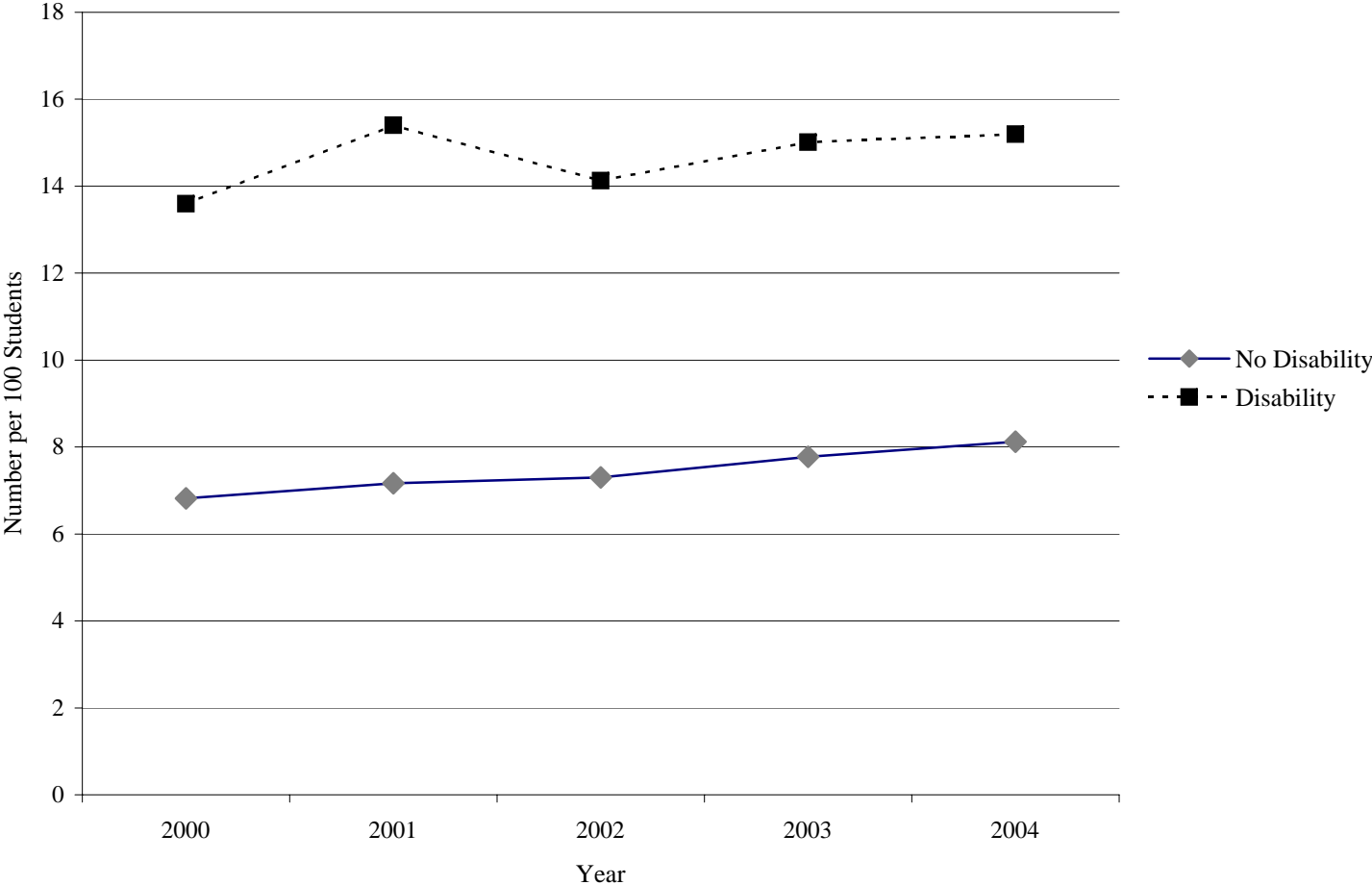


Table 4.1. Descriptive information about the number of suspensions and the number of students suspended in Maryland public schools: 2004

	Population	Students Suspended	Students suspended per 100	Suspensions	Suspensions per 100
Total	869113	78551	9.04	141556	16.29
White	438450	26470	6.04	46975	10.71
African American	329526	47278	14.35	87086	26.43
Disability	113760	17150	15.08	34784	30.58
No Disability	755353	61401	8.13	106772	14.14

Table 4.1 also displays the number of suspensions per 100 students within each category. The number of suspensions per 100 students was higher than the number of students suspended per 100 students for the population, for White students, African American students, and students with disabilities. The number of suspensions per 100 students for African American students was about two and half times the number of suspensions per 100 students for White students. The number of suspensions per 100 students for students with disabilities was more than two times the number for students without disabilities.

Table 4.2 displays the number of suspensions per hundred students for White students, African American students, and students with and without a disability from the sample. This table does not include the number of students suspended because those data were not available. The total suspensions per hundred students was 27.1, nearly twice that for the population. The suspensions per 100 students for African American students were nearly two and a half times the number for White students. Similarly, the number of suspensions per 100 students for students with disabilities was almost two and a half times the number for students with no disabilities.

Analysis 2

I used the logistic regression model to examine differences in suspensions for students by the combination of Race and Disability for 2004. This approach was identical to the approach I used in the pilot investigation for this study

Table 4.2. Descriptive information about the number of suspensions and the number of students suspended in sample schools: 2004

	N	Suspensions	Suspensions per 100 students
Total	394719	107395	27.21
White	239053	41152	17.21
African American	155666	66243	42.55
Disability	48944	25959	53.04
No Disability	393963	87913	22.32

(Krezmien, Leone, & Achilles, 2006). Since suspension data are not normally distributed, logistic regression was appropriate for this analysis. Logistic regression does not assume linearity of relationship between the independent variables and the dependent variables. Furthermore, logistic regression makes no assumption about the distribution of the independent variables. They do not have to be normally distributed, linearly related or of equal variance within each group.

Suspension was the criterion variable (0 = not suspended, 1 = suspended). Race by Disability was a categorical variable and was entered as a predictor. Each disability type for each racial group had a unique category. For instance, White Students with Mental Retardation was a distinct category with a unique code. There were a total of 16 categories representing each disability category for each racial group. White Students with No Disability was the reference group because it represented the largest group in the population. In the model, each of the categories was compared to the White Students with No Disability category.

Table 4.3 displays the odds ratios and the 95% confidence intervals for the odds ratios for the 16 categories of Race by Disability in 2004. The overall model was significant ($\text{Chi}^2 = 28,258, p < .001$). In the table, the No Disability category for White students is empty because it is the reference group.

The odds ratio for the African American group in the No Disability category was 2.61 with a small 95% confidence interval. The model predicted that students in the African American group with no disabilities were 2.61 times more

Table 4.3. Odds ratios for suspensions and 95% confidence intervals for the odds ratios for disability categories by Race

Disability Category	White			African American		
	Odds Ratio	95% CI	WALD	Odds Ratio	95% CI	WALD
No Disability				2.61	(2.57 - 2.65)	14147.7***
Mental Retardation	1.22	(1.04 - 1.43)	6.0*	3.02	(2.78 - 3.28)	671.9***
Speech / Language Impairment	0.36	(0.33 - 0.40)	381.4***	1.6	(1.49 - 1.72)	173.8***
Emotional Disturbance	7.25	(6.79 - 7.74)	3521.2***	11.72	(11.07 - 12.40)	7192.6***
Other Health Impairment	3.07	(2.87 - 3.28)	1081.4***	7.29	(6.82 - 7.80)	3359.1***
Learning Disability	2.48	(2.38 - 2.58)	1864.1***	5.9	(5.69 - 6.12)	9133.0***
Autism	0.77	(0.63 - 0.93)	7.3**	0.71	(0.55 - 0.92)	6.6*
Other Disability	0.84	(0.75 - 0.95)	8.3**	1.35	(1.22 - 1.50)	33.5***

* significant to the .05 level

** significant to the .01 level

*** significant to the .001 level

likely to be suspended than students in the White group in the no disability category. This predicted odds ratio was slightly higher than the ratio of the students suspended per 100 students for the African American group to the students suspended per hundred for the White group ($14.35 / 6.04 = 2.38$) from table 4.1.

The odds ratios were highest for students with emotional disturbance (ED). The odds ratios were significant and large for students with ED from both racial groups, and the 95% confidence intervals were small. The odds ratio was highest for the African American group in the ED category. The model predicted that students with ED were more likely to be suspended than White students without disabilities and the odds ratios for African American students with ED were higher than those from any other category.

The odds ratios were high for students in the Other Health Impairment (OHI) category for both racial groups with small 95% confidence intervals. The odds ratio for the African American group in the OHI category was higher than the odds ratio for any other group except the African American group in the ED category, and was more than twice the odds ratio for the White group in the OHI category.

The odds ratios were high for the Learning Disability (LD) category for both racial groups with small 95% confidence intervals. The odds ratio was higher for the African American group in the LD category and was nearly twice

the odds ratio of the White group in the LD category. The model predicted that students with OHI and students with LD from each racial group were more likely to be suspended than White students without disabilities.

The odds ratios for students in the Autism category were less than the odds ratios for White students without disabilities for both racial groups, but the 95% confidence intervals approached 1.0 for both racial groups. The odds ratio for the White group in the Other Disability category was low, but the 95% confidence interval approached 1.0. The odds ratio for the African American group in the Other Disability category was slightly above 1.0, with a lower 95% confidence interval of only 1.22.

The odds ratio for the White group in the Mental Retardation category was above 1.0, but the 95% confidence interval approached 1.0. In contrast, the odds ratio for the African American group in the Mental Retardation category was 3.02 with a small 95% confidence interval. The odds ratio for White students in the Speech/Language category was low, with a small confidence interval. In contrast, the odds ratio for the African American group in the Speech/Language category was above 1.5 with a small 95% confidence interval.

Analysis 3

I employed hierarchical modeling to understand the degree to which school-level factors and individual-level factors explained variation in suspensions. This analysis explained the extent to which school characteristics

contributed to suspension practices across racial categories and disability categories. The data in this analysis were counts of suspensions within schools, reported by race and by disability. The use of hierarchical generalized linear modeling (HGLM) was appropriate for this analysis (Lee, 2000). This approach is employed when data are not normally distributed. The data in this investigation had a poisson distribution, a discrete probability distribution most commonly used to model the number of occurrences of some phenomenon in a specified unit of space or time (Raudenbush & Bryk, 2002). The distribution is particularly useful for understanding rare events. It is appropriate for this investigation because suspensions are a relatively rare event, and most students experience no suspensions or few suspensions.

I examined the distribution of the data and verified the poisson distribution of the outcome data. A histogram of the suspension rate (number of suspensions within each school divided by the school enrollment) shows that the data have a poisson distribution (Appendix 1). Additionally, I examined the distribution of the level-1 residuals of the total suspensions (Appendix 2). The residuals were normally distributed, indicating that the assumptions about the poisson distribution of the data appear to have been correct (Raudenbush & Bryk, 2002).

I examined the correlations of the level-2 predictors to insure that I did not have a problem with multicollinearity, a problem when variables are too highly correlated. Appendix 3 contains the correlation matrix. A number of the variables

have significant correlations, but none of the correlations are strong. According to Tabachnick and Fidell (2001), variables with correlations above .90 are problematic. None of the level-2 predictors have such strong correlations, so multicollinearity is not a problem for this analysis.

I used HLM6 (Raudenbush, Bryk, & Cheong et al., 2004) to analyze the data, using the poisson model with variable exposure. Variable exposure was utilized because the schools varied by enrollment. It was therefore appropriate to weight the cases by enrollment. One analysis was conducted using Race at the level-1 predictor, and one analysis was conducted using Disability as the level-1 predictor as described in the Method section.

Race

Table 4.4 displays descriptive statistics for the level-1 and level-2 variables. The level one model is $E(Y_{ij} | \lambda_{ij}) = m_{ij}\lambda_{ij}$ $\text{Var } E(Y_{ij} | \lambda_{ij})$, with Y_{ij} being the number of suspensions in racial group i of school j and m being the population size of that racial group in that school. According to the model, the predicted value of Y_{ij} when $m_{ij} = 1$ will be the event rate λ_{ij} . Because the level 1 data are poisson distributed, they must be transformed using the log function. In HGLM, the log link function when the level-1 model is poisson is $\eta_{ij} = \log(\lambda_{ij})$. In this equation, η_{ij} is the log of the event rate. When the event rate is one, the log is zero. When the event rate is less than one, the log is negative and when the event rate is greater than one the log is positive. The B coefficients from the

Table 4.4. Descriptive Statistics for Suspensions for HGLM Analysis

Level-1 Descriptive Statistics					
Variable	N	Mean	SD	Minimum	Maximum
Enrollment	809	482.71	447.08	1	2681
Suspensions	809	131.68	178.42	0	1342
Level-2 Descriptive Statistics					
Variable	J	Mean	SD	Minimum	Maximum
Enrollment	405	1083.04	509.1	138	3293
Suspension Rate	405	0.27	0.22	0	1.45
Percent White	405	0.56	0.33	0	1
MSA Basic	405	54.54	21.57	8.4	100
Attendance	405	92.58	4.8	50.7	97.2
NHQ	405	35.84	18.55	3.5	100
Percent SPED	405	11.63	4.24	0	28.6
Mobility	405	11.09	8.81	0.2	93.4
Student-Teacher Ratio	405	16.52	2.19	8.6	23.7
SES	405	0.28	0.2	0	0.89

model are reported as log event rates. The predicted log event rates can be converted back to an event rate by computing $\lambda_{ij} = \text{event rate} = \exp(\eta_{ij})$, and λ_{ij} will be positive whatever the value of η_{ij} .

The level 2 model is formulated by using the level-1 intercept (B_{0j}) and slope (B_{1j}) as outcomes in the model. I ran analyses for three models, the unconditional model (a), the level-1 model with no level-2 predictors (b), and the full model (c). The unconditional model was used to gauge the magnitude of variation between schools in numbers of suspensions using a model with no predictors at either level. The level-1 model with no predictors was used to gauge the magnitude of the variation between schools when only the level-1 predictor was entered into the model. Finally, the full model was used to determine the association between the level-1 and level-2 factors on the risk of being suspended when all variables were entered into the model. Each of the models was used to calculate the proportion of variance explained at each step.

Predictors of Suspension

To assess the contribution of level-1 and level-2 predictors to differential suspension rates of White students and African American students across schools, I examined the coefficients from the full model (c) that included all level-1 and level-2 predictors. Table 4.5 displays the coefficients, the standard errors, and the exponentiated coefficients from the unit-specific and population-average level-2 models. The coefficients are the log-odds obtained from the HLM output, and the

Table 4.5. Log-Linear Models for Log Suspensions in Maryland with Race as Level-1 Predictor

Fixed Effect	Unit-Specific Model			Population-Average Model		
	Coefficient	se	exp(coeff)	Coefficient	se	exp(coeff)
Level-1 Variables						
Intercept, B_0	-3.930	0.620	0.020	-3.715	0.619	0.024
Race	-0.920 ***	0.008	0.399	-0.920 ***	0.008	0.399
Level-2 Variables						
Enrollment	0.000	0.000	1.000	0.000	0.000	1.000
Suspension	3.140 ***	0.104	23.123	3.205 ***	0.104	24.660
Percent White	1.040 ***	0.107	2.830	0.965 ***	0.107	2.625
MSA Basic	0.008 ***	0.002	1.009	0.007 ***	0.002	1.007
Attendance	0.009	0.005	1.009	0.011	0.007	1.011
NHQ	-0.002	0.002	0.998	-0.002	0.002	0.998
Percent SPED	-0.007	0.006	0.993	-0.006	0.006	0.994
Mobility	0.002	0.004	1.002	0.001	0.004	1.001
Student Teacher Ratio	0.007	0.010	1.007	0.007	0.011	1.007
SES	0.071	0.190	1.074	0.026	0.189	1.026

*** significant to the $p < .001$ level

exponentiated coefficients are the odds. The unit-specific model describes a process that is occurring in each level-2 unit (Raudenbush & Bryk, 2002, p. 304). In the unit-specific model, the level-2 model describes how differences in the explanatory variables at level-2 relate to differences in the level-1 processes at the level-1 unit.

The coefficients associated with each level-2 predictor represent the association between the predictor and the odds of being suspended for African American students (the reference group) holding all other level-2 predictors constant. The coefficients in the unit-specific model are interpreted as the expected difference in the log-odds of suspension associated with a unit increase in the predictor, holding constant all other predictors. The population-average model provides answers to population questions. In contrast to the unit-specific model, the coefficients in the population-average model provides the expected difference in the log-odds of suspension associated with a unit increase in the predictors holding constant all other predictors but averaging over the distribution of level-2 effects. In Table 4.5, the directions of all findings for each predictor from both models are identical, and the statistical significance nearly identical for each predictor in both models. However, there are slight differences in the coefficients of Suspension and Percent White.

Level-1 predictor. Table 4.5 shows a strong association between Race and Suspensions in the unit-specific model. Because White is represented by 1 in the

analysis, the $\exp(\text{coeff})$ of 0.399 indicates that White students have a decreased odds of suspension of 0.399 compared to African American students, holding all other predictors constant. Because I was interested in understanding the odds of being suspended for African American students, I calculated the odds for African American students. The odds for African American students is the inverse of the odds for White students $\{\text{odds} = 1 / 0.399 = 2.51\}$, or an odds of 2.51.

Level-2 predictors. The table also shows that the level-2 predictors of Suspension Rate, MSA Basic, and Percent White all have strong positive associations with Suspension in the unit-specific model. A one standard deviation increase in Suspension Rate (.22) multiplies the Suspension Rate by $\exp\{(.22)*(3.141)\} = 2.00$, or a 100% increase in the suspension rate. A one standard deviation increase in MSA Basic (21.57) multiplies the suspension rate by $\exp\{(21.57)*(.009)\} = 1.21$, or a 21% increase in the suspension rate, and a one standard deviation in Percent White (0.33) multiplies the suspension rate by $\exp\{(0.33)*(2.83)\} = 2.54$, or a 154% increase in the suspension rate. None of the other predictors in the unit-specific model were significantly related to Suspensions.

The coefficients from the population-average model are similar to those from the unit-specific model. In the population-average model, Race and MSA Basic have nearly identical associations with Suspensions. Suspension Rate and Percent White have different associations. In the model, a one standard deviation

in Suspension Rate multiplies the suspension rate by 2.02, or a 112% increase in suspensions, slightly greater than in the unit-specific model. In contrast, a one standard deviation increase in Percent White multiplies the suspension rate by 2.67, or a 141% increase in suspensions. This is a slight decrease from the unit-specific model.

Disability

I used the same procedures with Disability as the level-1 predictor as I used for Race as the level-1 predictor. With Disability as the level-1 predictor, I used No Disability as the reference group. Since Disability represented such a small percentage of the sample, it was not appropriate to use Disability as the reference group. Therefore, when considering the impact of the level-2 factors on risk of being suspended for students with a disability, it is necessary to multiply findings for students with No Disability by the level-1 odds for Disability, which represents the odds of being suspended for a student with a disability.

Predictors of Suspension

To assess the contribution of level-1 and level-2 predictors to differential suspension rates of students with and without disabilities, I examined the coefficients from the full model that included all level-1 and level-2 predictors. Table 4.6 displays the coefficients, the standard errors, and the exponentiated

Table 4.6. Log-Linear Models for Log Suspensions in Maryland with Disability as Level-1 Predictor

Fixed Effect	Unit-Specific Model			Population-Average Model		
	Coefficient	se	exp(coeff)	Coefficient	se	exp(coeff)
Level 1 Variables						
Intercept, B_0	-1.793	0.019	0.166	-1.739	0.019	0.176
Disability	0.799 ***	0.007	2.220	0.799 ***	0.007	2.220
Level 2 Variables						
Enrollment	0.000	0.000	1.000	0.000	0.000	1.000
Suspension	3.217 ***	0.106	24.960	3.301 ***	0.106	27.155
Percent White	0.294 **	0.109	1.340	0.216 **	0.108	1.241
MSA Basic	0.010 ***	0.002	1.011	0.009 ***	0.002	1.009
Attendance	0.010	0.007	1.011	0.009	0.007	1.009
NHQ	-0.002	0.002	0.998	-0.002	0.002	0.998
Percent SPED	-0.021 ***	0.006	0.979	-0.019 **	0.006	0.981
Mobility	0.003	0.004	1.003	0.002	0.004	1.002
Student Teacher Ratio	0.008	0.011	1.008	0.008	0.011	1.008
Free and Reduced Lunc	0.062	0.193	1.064	0.026	0.193	1.026

** significant to the $p < .01$ level

*** significant to the $p < .001$ level

coefficients from the full model with Disability as the level-1 predictor. The coefficients are the log-odds obtained from the HLM output, and the exponentiated coefficients are the odds.

Level-1 predictor. Table 4.6 shows a strong association between Disability and Suspensions in the unit-specific model. Because students with a disability is represented by 1 in the analysis, the $\exp(\text{coeff})$ of 2.220 indicates that students with a disability have an increased odds of suspension of 2.220 over students with no disability, holding all other predictors constant.

Level-2 predictors. The table also shows that Suspension Rate, MSA Basic, Percent White, and Percent SPED all have strong positive associations with Suspension in the unit-specific model. A one standard deviation increase in Suspension Rate (.22) multiplies the Suspension Rate by $\exp\{(.22)*(3.217)\} = 2.03$, or a 103% increase in the suspension rate. A one standard deviation increase in MSA Basic (21.60) multiplies the Suspension Rate by $\exp\{(.010)*(21.60)\} = 1.24$, or a 24% increase in the suspension rate. A one standard deviation in Percent White (0.33) multiplies the suspension rate by $\exp\{(0.33)*(.294)\} = 1.10$, or a 10% increase in the suspension rate. A one standard deviation increase in Percent SPED (4.25) multiplies the suspension rate by $\exp\{(4.25)*(-.021)\} = 0.91$, or a 9% decrease in the suspension rate. None of the other predictors in the unit-specific model are significantly related to Suspensions. The coefficients from the population-average model are all similar to those from the unit-specific model.

Explanation of Variance

Table 4.7 displays the variance components associated with the three HGLM models for race as the level-1 predictor and for disability as the level-1 predictor. To determine the percent of level-2 variance explained by the models, the variance component for the model is subtracted from the component of the unconditional model and then divided by the variance component of the unconditional model. Table 4.7 shows that the level-1 predictor accounts for 15.8% of the level-2 variance. The full model accounts for 83.2% of the level 2 variance, an increase of 526% from the model with the level-1 predictor only.

Table 4.7 displays the variance components associated with the three HGLM models for Disability as the level-1 predictor. Table 4.7 shows that the level-1 predictor accounts for 2.3% of the level-2 variance. The full model accounts for 82.6% of the level12 variance, an increase of a factor of 35.1 from the model with the level-1 predictor only.

Analysis 4

Race

A discriminant function analysis (DFA) was used to explore the extent to which the types of offense resulting in suspension predicted group membership by race and by disability. A DFA is appropriate to understand the dimensions along which groups differ (Tabachnick & Fidell, 2001). According to Tabachnick and Fidell (2001), the following conditions must be met in order to conduct a

Table 4.7. Variance Explained from HGLM Models for Log Suspensions in Maryland when Race and Disability were entered as Level-1 Predictors

Model Using Race as Level-1 Predictor

	(a)	(b)	(c)
Variance Component	0.800	0.673	0.134
% of Variance Explained		0.158	0.832

Model Using Disability as Level-1 Predictor

	(a)	(b)	(c)
Variance Component	0.798	0.780	0.139
% of Variance Explained		0.023	0.826

(a) = unconditional model

(b) = model with level-1 predictor only

(c) = model with level-1 and level-2 predictors

discriminant analysis: (a) there are two or more mutually exclusive groups, (b) there are at least two subject per group, (c) any number of variables can be included as long as cases exceed variables by more than two, (d) no discriminating variable can be a linear combination of others, and (e) each group is drawn from a multivariate normal distribution on the discriminating variables. Each of the conditions is met except for the final condition. However, DFA is robust to failures of normality if the violations are caused by skewness rather than by outliers (Tabachnick & Fidell, 2001). I used the Mahalanobis distance test to determine significant outliers, and none were identified (Tabachnick & Fidell, 2001). I therefore proceeded with the DFA.

I used Direct DFA, entering all predictors at one time, and each predictor is assigned only the unique association it has with groups (Tabachnick & Fidell, 2001). The grouping variable was Race (0 = White, 1 = African American). The independent variables were the eight offense categories as reported by the MSDE. All variables significantly entered and remained in the discriminant function at $p < .01$ level or better.

With only two conditions for the criterion variable (Race), the analysis yielded a single canonical discriminant function. The canonical correlation coefficient associated with the eigen value was moderate (0.627), indicating that a large amount of the variance in offenses was not explained by the function. The Wilks' lambda associated with the function was moderate in size (.607), and

significant (X^2 (df = 8) = 197165.5, $p < .001$). The size of the Wilk's lambda indicates that the proportion of the overall variance accounted for was $\{1 - (.607)^2\} = 0.63$ (Tabachnick & Fidell, 2001).

Of greater interest were the specific offenses that significantly differentiated between White students and African American students. Variables entering the equation and the measures of their respective strength are displayed in Table 4.8. Positive and negative values are arbitrary, based on the coding of African American students as 1 and White students as 0. A positive value indicates a significantly higher mean offenses for African American students, and a negative value indicates a significantly higher mean offenses for White students.

White students appear to be suspended for Dangerous substances, and to a lesser degree, for Sex offenses. African American students appear to be suspended for Attacks and Threats, Weapons, Other Offenses, and to a lesser degree Disrespect. However, examination of the structure matrix indicates that only Dangerous Substances loaded on the White students. Examination of the Classification results indicates that the model was better at predicting group membership for White students (94.5%) than for African American students (59.8%).

Disability

A DFA was also used to explore the extent to which the types of offense resulting in suspension predicted group membership by Disability. The grouping

Table 4.8. Discriminant Function Analysis Predicting Race by Offense Resulting in Suspension

Offense	Variables Predicting African American Suspension		Variables Predicting White Suspension	
	DFA Coefficient	Structure Matrix	DFA Coefficient	Structure Matrix
Attendance		0.688	-0.079	
Weapons	0.346	0.590		
Disrespect	0.185	0.493		
Other	0.354	0.451		
Fire		0.346	-0.053	
Attack / Threat	0.489	0.135		
Sex		0.102	-0.124	
Dangerous Substance			-0.726	-0.342

DFA coefficient represents the standardized canonical discriminant function coefficient, transformed so that all variables have a mean of 0 and a standard deviation of 1. This coefficient can be regarded as an index of the relative importance of each variable in the function.

Structure Matrix represents pooled within-group correlations between discriminating variables and standardized canonical and represents an index of the degree of correlation of the variable with the function within each group

variable was Disability (0 = No Disability, 1 = Disability). All variables significantly entered and remained in the discriminant function at $p < .01$ level or better. The Wilks' lambda associated with the function was large in size (.877), and significant (X^2 (df = 8) = 57954.4, $p < .001$). The size of the Wilk's lambda indicates that the variance accounted for was small $\{1 - (.877)^2\} = 0.23$ (See T & B). Additionally, the canonical correlation was small (0.350), indicating the findings were not robust. Variables entering the equation and the measures of their respective strength are displayed in Table 4.9. Positive and negative values are arbitrary, based on the coding of African American students as 1 and White students as 0. A positive value indicates a significantly higher mean offenses for students with a disability, and a negative value indicates a significantly higher mean offenses for students with no disability.

Students with no disability appear to be suspended for Attendance and Disrespect, but only to a small degree. Students with a disability appear to be suspended for Dangerous Substances, Attacks and Threats, and to a lesser degree for Sex Offenses, Fire Offenses, Other Offenses, and Weapons Offenses. Examination of the structure matrix indicates that none of the offense categories loaded for students with no disabilities. Examination of the Classification results indicates that the model predicted group membership for both groups, with 88.9% of the original grouped cases correctly classified.

Table 4.9. Discriminant Function Analysis Predicting Disability by Offense Resulting in Suspension

Offense	Variables Predicting Disability Suspension		Variables Predicting No Disability Suspension	
	DFA Coefficient	Structure Matrix	DFA Coefficient	Structure Matrix
Attack / Threat	0.531	0.698		
Dangerous Substance	0.644	0.659		
Disrespect		0.592	-0.156	
Weapons	0.12	0.576		
Other	0.155	0.498		
Fire	0.219	0.483		
Sex	0.239	0.354		
Attendance		0.275	-0.145	

DFA coefficient represents the standardized canonical discriminant function coefficient, transformed so that all variables have a mean of 0 and a standard deviation of 1. This coefficient can be regarded as an index of the relative importance of each variable in the function.

Structure Matrix represents pooled within-group correlations between discriminating variables and standardized canonical and represents an index of the degree of correlation of the variable with the function within each group

CHAPTER V: DISCUSSION

The findings of this study were consistent with a body of research that found overrepresentation of African American students (Bruns et al., 2005; Christle et al., 2004; Costenbader & Markson, 1998; Cooley, 1995; Losen et al., 2003; McFadden & Marsh, 1992; Rafaelle Mendez & Knoff, 2002; Rafaelle Mendez, 2002; Rafaelle Mendez et al., 2002; Rausch & Skiba, 2004; Skiba et al., 2002; Skiba et al., 1997; Zhang et al., 2004) and students with disabilities (Cooley, 1995; Fasko et al., 1995; Skiba et al., 1997; Zhang et al., 2005) suspended from schools. The findings were also consistent with an emerging body of research that identified school factors associated with suspension rates (Bruns et al., 2005; Christle et al., 2004; Rafaelle Mendez & Knoff, 2003; Rausch & Skiba, 2004). The present investigation contributed to the current body of research as an initial examination that employed multilevel modeling to understand how individual and school factors are associated with risk for suspension. As such, the findings from this study represent initial empirical evidence that both school-level characteristics and student-level characteristics have significant correlations with student suspensions when examined in a single multilevel model.

The findings from the HGLM analyses revealed substantial variability in the suspension practices of schools, but more importantly, they indicated that school-level characteristics accounted for a majority of the explained variance in the suspensions of youth in Maryland.

Nonetheless, the odds of being suspended for African American students were 2.51 times the odds for White students, and the odds for students with disabilities were 2.22 times the odds for students without disabilities when school characteristics were controlled. The magnitude of the odds was high for both groups, and was consistent with findings from other investigations (Bruns et al., 2005; Costenbader & Markson, 1998; Cooley, 1995; Losen et al., 2003; Rafaelle Mendez, 2002; Rausch & Skiba, 2004; Skiba et al, 2002; Skiba et al, 1997; Zhang et al., 2004) and with the findings from the pilot study for this investigation, all single-level investigations.

Furthermore, racial compositions of schools and the percentage of students with disabilities were also significantly associated with odds of being suspended, consistent with findings from two prior investigations (Christle et al., 2004; Rafaelle Mendez et al., 2002). The individual factors of Race and Disability were significant predictors of suspension in the full model, but accounted for only a limited proportion of the explained variance. Considering the magnitude of the odds of being suspended associated with the school factors, the findings from the individual-level logistic regression analysis (Analysis 2) and the individual-level discriminant analysis (Analysis 4) should be considered with caution.

Analysis 1

The descriptive data revealed that the increasing trend in the number of students suspended in Maryland public schools over the past decade continued

during the 2004 school year. The number of students suspended per 100 students was the highest on record in Maryland. The rate of African American students suspended was more than twice the rate of White students suspended in 2004. Additionally, African American students experienced a steady increase in the rates of suspensions since 1995, while the rates for White students remained stable over the same 10 year period. It is difficult to interpret the increasing trend in suspensions for African American students, but the trend does suggest that the zero tolerance policies implemented in Maryland have had a disproportionate impact on African American students, but no noticeable impact on White students.

It is impossible to tell if the problem behaviors of African American children and youth have increased consistent with this trend, if the behaviors have remained constant but the implementation of zero tolerance policies have forced administrators to suspend African American students for behaviors that were previously managed through alternate disciplinary procedures, or if some combination of these phenomena may explain the difference in trends. Students with disabilities were also suspended at much higher rates than students without disabilities, although both groups experienced increases in rates from 2000 to 2004.

Analysis 2

The findings from the logistic regression analysis examining the number of students suspended by race and disability categories indicated differences in the odds of being suspended by race, by disability category, and by the combination of race and disability category. The findings from this analysis did not employ multi-level procedures because the data did not allow for hierarchical analysis. As a consequence, the odds of being suspended for African American students and students with disabilities may be inflated. However, since the findings from the HGLM indicated that both Race and Disability were significantly associated with risk for suspension in the hierarchical model, and since the odds for both groups were large, I have included a discussion of this analysis. These findings, however, should be considered with caution.

I found that that youth identified as having disabilities experienced higher rates of suspension than youth not identified as having disabilities, consistent with other current reports (Cooley, 1995; Zhang et al., 2004). The risk of being suspended among students with EBD was high across both racial groups, not a surprising finding considering that problem behaviors are a defining characteristic of the disability. However, the disproportionate odds of being suspended for students with ED may indicate that the behaviors associated with this disability may be poorly managed by schools or that behaviors associated with the disability are not considered when determining disciplinary consequences. The high suspension rates of students with ED are problematic because these students

require intensive behavioral interventions implemented consistently over time. Interruption of these interventions because of referrals to the office, suspensions, or expulsions negatively impacts the effectiveness of special education behavioral interventions and supports (Losen et al., 2003).

I also found that students with OHI and LD had higher risks of being suspended than their same race peers without disabilities. One possible explanation in the elevated odds for students with LD is that students with these disabilities often find academic tasks aversive and may respond to difficult academic tasks with disruptive behaviors that result in disciplinary referrals and exclusions (Scott et al., 2001). Additionally, most students with LD in Maryland are instructed in general education classrooms. As a result, many of these students may not be getting the quality and intensity of instruction that they would receive from trained special educators who can deliver the specialized and targeted instruction required by the needs of many LD students.

If the increased odds of being suspended are due to increases in the problem behaviors of these youth, the behaviors could be associated with frustration from their academic difficulties in these general education classrooms. Students with ADHD, the students most commonly identified with an OHI, tend to exhibit a number of behaviors (e.g., out of seat, off-task, talking out) that are typically perceived as disruptive (Skiba et al., 2002). These are the types of

behaviors that have increasingly resulted in suspensions as a result of zero tolerance policies (Losen et al., 2003).

Removal does not, however, promote prosocial behaviors and it limits student access to the behavioral or academic supports that may decrease future problem behaviors. Instead, exclusion of students with disabilities reduces their exposure to instruction, makes academic tasks more aversive, results in further negative behaviors, and increases the risk for further exclusions (Scott, Nelson, & Liaupsin, 2001). In effect, the disciplinary policies that contribute to high rates of suspension for students with ED and other disabilities such as ADHD may conflict with the underlying principles of effective and empirically supported behavioral interventions validated through special education research and practice.

The odds of being suspended for African American students were higher than the odds of being suspended for White students for all disability categories except the Autism category. This was most clearly evident for African American students with mental retardation whose odds of being suspended were approximately three times the odds for White students with mental retardation. I found no evidence to indicate that suspensions are an appropriate form of discipline for students with mental retardation. Students with mental retardation have impaired cognitive faculties that may interfere with the ability to understand disciplinary practices or consequences. Disciplinary removal of students with

mental retardation is not an appropriate response considering the nature of the disability, and such practices likely inhibit appropriate and prosocial behaviors among this population.

Furthermore, the increased risk for removal from school for disciplinary purposes faced by students with mental retardation who happen to be African American is difficult to understand. There is no evidence that African American students with mental retardation exhibit a higher frequency of suspendable behaviors than White students with disabilities. The racial disparity, however, was not limited to students with mental retardation. African American students with ED, OHI, LD, and Speech/Language Impairments were also disproportionately at risk for suspensions. This finding was consistent with the findings from the pilot study for this investigation (Krezmien, Achilles, & Leone, 2006), and was consistent with the disproportionate risk of being suspended for African American students and students with disabilities identified in the HGLM.

One possible explanation is that schools with high numbers of African American students with disabilities may have high suspension rates that artificially inflate their odd of being suspended in the aggregate (Christle et al., 2002; Krezmien et al., 2006; Rafaelle Mendez et al., 2002). However, the results from the HGLM analysis are not consistent with this perspective.

Analysis 3

The HGLM analyses allowed me to examine the impact of both individual and school factors on risk of suspensions for middle school and high school students in Maryland. In contrast to the previous two analyses, the HGLM involved numbers of suspensions as the dependent measure. The use of suspension count data is different from the use of the number of students suspended, particularly because the suspension data accounts for multiply suspended students. Both dependent measures have been used in previous research, and have been determined to be effective for examining rates of suspensions and disproportionate impact of suspension practices (Rafelle Mendez 2002).

Analysis 3 was also different from the previous analyses because it only includes middle and high school students. Authors of previous research have consistently demonstrated that middle and high school students are much more likely to be suspended than primary school students (Cooley, 1995; Rafaele Mendez & Knoff, 2003; Rafaele Mendez et al., 2002; Rausch & Skiba, 2004). The differences in these risks are also evident in the suspensions per 100 students for the sample and for the entire population (Table 4.1 and 4.2). The suspension per 100 students was substantially higher for each group in the sample than for the respective group in the population which included primary level students. The differences in suspensions per 100 students between the sample and the

population were particularly large African American students (Mean = 42.55) and for students with disabilities in (Mean = 53.04) in 2004.

School-Level Predictors

I found several important school level predictors associated with suspensions in schools when either Race or Disability was included as the level-1 predictor of suspensions. Not surprisingly, Suspension Rate was the strongest predictor of suspensions. This finding is consistent with the findings from Christle and colleagues (2004) and Rausch and Skiba (2004) In the HGLM model, the coefficients associated with the school-level predictors represent the effect on the odds for suspension for the African American students when Race was the level-1 predictor and students with no disability when Disability was the level-1 predictor holding all other level-2 predictors constant.

The high odds of being suspended associated with suspension rate indicates high variability in suspension rates across school The strong relation between high suspension rates in schools and risk of suspensions indicate that schools differ in their approach to disciplinary removal from school or that some schools have much higher rates of suspendable student offenses. However, the current model did not account for the administrative structures or school disciplinary policies in the analysis. Furthermore, there was no way to determine whether the frequency or intensity of behaviors resulting in suspensions were different across the different schools. A follow-up analysis that included these

additional sources of data would be necessary to better understand the nature of the disproportionate suspension rates.

The racial composition of schools also contributed to odds of suspensions when Race was entered as the level-1 predictor. The increase in the odds of being suspended associated with an increase in the percent of the school enrollment that was White should be considered in the context of the odds for African American students. It indicates that as the percentage of White students enrolled in a school increase, the odds of being suspended for African American students, the reference group increases. Because White students have 0.399 the odds of being suspended compared to African American students, the percent White had a slightly negative impact of the odds of being suspended for White students. There are no other studies that have investigated the school racial composition and the risk of being suspended for individuals within a specific racial category making an explanation difficult. This finding appears to contradict the assumption that schools with high percentages of African American students disproportionately suspend students resulting in higher odds of being suspended for African American students at the state level. One problem with the use of Percent White as the level-2 predictor is that it relies on the assumption that Percent White also measures Percent African American since the racial compositions of all schools examined are primarily composed of these two racial groups. However, if the percentage of African American students was entered into the model, there is a

possibility that it could also have been a significant predictor of suspensions for African American students and / or White students. An examination of the suspension rates of schools that are primarily White reveals that they have suspension rates that are substantially below the mean rate for all schools. In contrast, schools that have a majority of African American students have the highest suspension rates in Maryland. Further investigation of this finding is necessary to better understand the phenomena.

The racial composition of schools and the percentage of students with a disability in the schools were also significantly associated with the risk of being suspended for students with and without disabilities. However, neither of these factors had a strong association with risk for suspension. I found that a 33% increase in the percent of enrolled students that were White only accounted for a 10% increase in the suspension rate for students without disabilities, while a 4.25% increase in the percent of enrolled students with a disability was associated with a 9% decrease in the suspension rate for students without disabilities.

These increases in the composition of the schools were quite large, but the relative contribution to the risk of being suspended was small, particularly compared to the contributions from the other school-level factors. Nonetheless, it appears that a predominantly White school increases the risk of suspension for students without a disability, and schools with a high percentage of students with disabilities enrolled decrease the risk. Again, it is important to consider the impact

on students with a disability, who were more than two times as likely to be suspended as their peers without a disability.

Similarly, I found that the percent of the enrollment that had a disability was significantly and negatively associated with odds of being suspended for students with no disability, although the strength of the association was not strong. This finding is difficult to interpret, but it may indicate that schools with high percentages of students with disabilities identify a large percentage of students and provide adequate educational and related services. Consequently, such schools may be better prepared to manage the difficult behaviors that students with disabilities are known to exhibit. Further investigation of this phenomena should be conducted at the school level.

The percent of the students who did not meet proficiency on the MSA was also significantly and positively associated with risk of suspension when either Race or Disability was entered as the level-1 predictor. This finding was consistent with the findings of a number of researchers (Christle et al., 2002; Losen et al., 2003; Rafaelle Mendez, 2002; Rafaelle Mendez et al., 2002; Rausch & Skiba, 2004). The authors proposed that students in schools with high rates of students who perform well on their state assessments have a lower risk for suspension. They maintain that academically high performing schools emphasize the importance of instruction, implement high quality educational practices, and keep students actively engaged. As a consequence, students are more successful

in the classroom and have limited opportunities or reasons to act out in a manner that could result in a suspension. This is also consistent with a demonstrated correlation between academic achievement and behavior (Benner, Epstein, & Nelson, 2002).

The proposals by these authors may contribute to the understanding of the findings from this analysis. Schools with low percentages of students who meet proficiency on the state assessments are faced with the difficult task of increasing student performance while managing student frustration and disciplinary infractions. These schools may not have the resources or experience to respond to these challenges appropriately, or they may lack effective and well-trained teachers. As a result, there is a chance that these schools rely upon zero tolerance policies to remove students who misbehave, and consequently diminish the negative impact of disruptive behavior on the learning environment of the other students. Further research is required to better understand the relationship between school performance on state standard assessments and the increased risk of suspension.

A number of school-level predictors were not significantly associated with risk of being suspended. The most interesting factor that was not related to suspensions was socioeconomic status. A number of other investigators previously found that socioeconomic status was significantly predictive of suspensions (Bruns et al., 2005; Christle et al., 2004; Rafaelle Mendez et al.,

2002; Skiba et al, 1997). In general, these authors suggested that schools with high percentages of students who came from low socioeconomic backgrounds likely had limited funding and may have been poorly equipped with staff and other resources to adequately manage inappropriate behaviors through alternative means such as in school suspension, peer mediation, mentoring, etc. In contrast, Skiba and his colleagues (2002) found that disproportionate risk of being suspended was not diminished when socioeconomic status was controlled. The authors argued that schools with limited resources may suspend a higher percentage of youth, but that there was no evidence that socioeconomic status affected the disproportionate suspension of minority youth. The findings from this investigation suggest that socioeconomic status is unrelated to rates of suspension. This may be due to the multilevel analysis utilized in this analysis, but replications of this finding are necessary if the inconsistency across studies is to be resolved.

School size was not significantly associated with risk of suspensions consistent with other research (Christle et al., 2002; Cooley, 1995; Raffaele Mendez et al, 2002; Rausch & Skiba, 2004). I also found that the percentage of teachers who were not highly qualified, student mobility rates, the student to teacher ratio, and student attendance rates were not significantly associated with the risk of being suspended. These findings were not consistent with the findings

of a number of prior investigations (Bruns et al., 2005; Christle et al., 2004; Rafaelle Mendez et al., 2002).

Once again, these authors suggested that most of these factors were associated with limited resources of schools. However, none of the researchers used the multilevel modeling used in this investigation. While a number of explanations may explain the inconsistency across studies, but it appears that a number of significant predictors of suspension found in prior investigations may have been inaccurately identified because of the single-level designs utilized in the analyses. Replications of this investigation across multiple states and locales are necessary to eliminate the inconsistencies.

Analysis 4

The discriminant analysis for offenses by Race revealed differences in the types of offenses that resulted in suspensions for White students and African American students. As discussed previously, the findings from this analysis were not obtained using multilevel modeling, and should be considered with caution. White students were overrepresented for some serious categories of offenses (Dangerous Substances and Sex Offenses), and African American students were overrepresented for other serious categories (Attacks/Threats and Weapons Offenses).

This finding was not consistent with the findings of Skiba and colleagues (2002) who found that African American students were primarily suspended for

minor infractions such as loitering and disruptions. The differences may be due to differences in the locales or to differences in the disciplinary policies of different states and / or different schools and school districts. African American students were also disproportionately suspended for Disrespect, a less serious offense, but the association was not strong. This model was useful for predicting group membership for White students, but it only correctly classified 58.9% of the African American students. This substantially limits the interpretability of the finding.

The discriminant analysis for offenses by Disability status was less robust than the analysis for Race, and the model only accounted for a small percentage of the variance. Students with no disability tended to be suspended for less serious offenses, while students with disabilities tended to be suspended for more serious offenses such as Dangerous Substances, Attacks and Threats. One major problem with interpretation of the findings was the inability to include disability type in the analysis. Considering the differences in the odds of being suspended associated with disability type obtained from Analysis 2, including disability type into the model is recommended.

Considering the high odds of being suspended for students with ED obtained from Analysis 2, a disproportionate number of the suspensions for students with disabilities were accounted for by this group of students. Students with ED have consistently been linked to serious misconduct in school (Benner et

al., 2001; Bruns et al., 2005), and they have an increased risk for drug use and delinquent activity (Leone et al., 2003). The disproportionate suspensions for substance use and attacks and threats may be due to the behaviors of these youth, but further investigation is required to better understand this finding.

Limitations

A number of factors limit the interpretation of the findings. The information available to complete the analysis presented here were obtained from extant datasets and were limited in a number of ways. Information about sex, grade level, or socioeconomic status of students who were suspended was unavailable. While the single predictors of race and disability were available and could be linked to specific infractions the data could not be analyzed within the hierarchical model. Thus, it was impossible to examine the impact that school context had on the types of offenses resulting in suspensions. The school-level factors from the level-2 of the HGLM analysis were also limited. If I had been able to include information about the administrative structures and disciplinary policies I could have constructed a model that could better explain how schools contribute to disproportionate suspensions of African American students and students with disabilities. Additionally, it was impossible to understand the extent to which school factors actually represented community factors. As a result, some of the variance associated with schools may in fact be associated with the

community. Due to the nature of the data available to this and other similar investigations, it may be impossible to adequately resolve this issue,

There may be differences in the way that schools and school districts identify and report suspensions, limiting the interpretability of our findings. Schools may use different approaches for interpreting and reporting suspension data to the MSDE. There may also be differences in the accuracy of the data across schools. School suspension rates are now an important measure used in determining if schools are persistently dangerous, which may impact what information schools report.

The data used in this investigation were obtained from the Maryland State Department of Education that collected and compiled suspension reports from schools. There is no way to verify the accuracy of these reports. Additionally, the MSDE may have an inadequate system for managing and reporting the data from the schools. The inadequacies of The MSDE's data collection and reporting process is evidenced in the fact that the type of suspension data that The MSDE reports each year has changed over time, which also make analysis of trends impossible. Future investigations should include a system for the verification of data across schools and school districts. Theoretically such an approach could be accomplished by randomly selecting and verifying data from a sample of schools, but obtaining records from schools is difficult as those records are not part of the public record.

The use of parallel analyses in my HGLM was also a limitation. Because of the nature of the data, I had to conduct a separate analysis for each of the individual-level factors. The practice of using simultaneous analyses using the same dataset is problematic, so the findings must be considered carefully.

Finally, each of the analytical procedures employed in this investigation were correlational. The analyses allow for statements of predictability, but do not support statements of causality. Future experimental research must be conducted to answer questions about causal Relationships between school and individual characteristics and the risk of being suspended.

Recommendations for Practice

Policymakers, researchers, and educators must develop ways to keep schools safe without compromising the quality of education services. Comprehensive and preventative approaches to maintaining school safety and discipline need to replace punitive and exclusionary procedures currently in place. By targeting all students, not just “problem students,” comprehensive approaches encourage positive alternatives to maladaptive behavior. One of the primary components of such a strategy includes the development of assessment procedures to identify and intervene with students at-risk for disruptive or anti-social behavior (Walker & Severson, 1992). Accurate collection and reporting of suspension data can be used by the states to track suspending practices of school

districts in order to monitor suspending practices of schools and to evaluate interventions to reduce disproportionate suspensions.

Under NCLB, schools are evaluated based upon their suspension rates, which is a principal factor in determining whether a school is identified as persistently dangerous. As a consequence, some school administrators may be tempted to keep students off of the suspension roles by placing students in an alternative placement that serves the same purpose as a suspension, but does not contribute to the school suspension rates. Inconsistencies of this type lead to inaccurate reporting of data, and subsequent misunderstanding of the magnitude of suspensions in schools, districts, and states. State agencies should develop clear guidelines for schools and school districts so that accurate data are collected, and problems can be adequately addressed. Additionally, state agencies should develop a system for monitoring suspension practices of schools and school districts, as well as a system for monitoring how schools and school districts collect and report suspension data. Developing a consistent approach to suspension practices and data collection will help with the development of disciplinary systems that can better support school personnel as well as students with disciplinary problems.

Additionally, special educators must be more involved in the development of school disciplinary policies. Students with disabilities are disproportionately suspended from school. I believe that special educators at the school level,

administrative level, and state level should become active in the development of disciplinary policies that promote school safety while limiting the influence of inflexible zero tolerance practices on special education students whose problem behaviors may be associated with their disability. While the manifestation determination procedures mandated under IDEIA are laudable, they are an insufficient response to a problem that has a broad and negative impact upon the population of students involved in the special education system.

Recommendations for Future Research

Future investigations in the area of school discipline should pursue several areas of examination. Researchers should find ways to include offenses resulting in disciplinary suspension into a multilevel model. This line of study would allow researchers to examine parity in offenses among students of different races and disabilities while understanding the importance of organizational context on suspension practices. Researchers should also attempt to specifically identify underlying factors associated with disproportionate suspensions of African American youth and youth with disabilities. Additionally, researchers should investigate if minority students and students with disabilities are disproportionately suspended for specific types of offenses using multilevel designs to examine the complex ways that the interactions between individual characteristics and school factors affect patterns of suspensions. In this investigation

I identified some individual and school characteristics that were associated highly with risk for suspension, but the participant and school characteristics included in this investigation lacked some important factors that should be included in future investigations. I believe that several individual characteristics would likely account for variance in single level and multilevel models. Future investigators should include (a) participant academic performance, (b) previous suspensions of participants, (c) disability categories of participants, (d) participant gender, and (e) participant age. Additionally, researchers should investigate the administrative structures or processes that contribute to suspension rates. Future investigators should identify schools that place specific groups of students at risk for suspension, and examine the internal processes that mitigate or ameliorate risk for disciplinary removal. State level data as currently reported in Maryland are inadequate for such an investigation. Finally, researchers should link suspensions of individuals with long-term problems including risk for future suspension or expulsion, dropping out of school, grade retention, and future involvement with the juvenile delinquency system.

Conclusions

The findings from this investigation represent an initial investigation of disproportionate suspensions of African American students and students with disabilities using multilevel analyses. Two important findings emerged from this investigation. First, I found that a number of school factors were significantly

associated with suspensions of youth when Race and Disability were controlled as level-1 predictors. This finding supports the model proposed by Lee (2000) that contends that the school is the appropriate organizational representing the major contextual context for adolescents. The school level factors accounted for the majority of the variance in suspensions. I examined the effects of school factors on suspensions, and not educational characteristics investigated by Lee. However, the robust findings I identified at the school level indicate that hierarchical modeling was the appropriate approach for understanding outcomes for adolescents in schools, and that future investigators of suspensions should include school factors in a multilevel analysis.

Second, I found that Race and Disability were significant and robust predictors of the suspensions even when school-level factors were controlled. This finding suggests that there is some systematic way that youth with disabilities and African American youth are disproportionately suspended from school. It was impossible to identify additional individual-level factors or school-level factors associated with this increased risk of being suspended. However, from the perspective of Skiba and his colleagues (2003), this investigation may have contributed to the current body of research by eliminating one more alternative explanation for the disproportionate suspensions of African American youth and youth with disabilities.

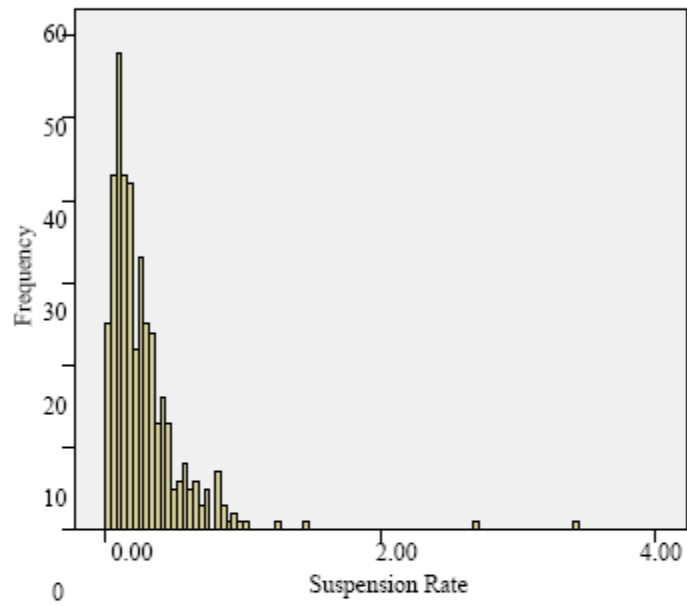
The findings reported here raise a number of issues about suspension practices in Maryland, and about the way suspension data are analyzed and reported. In successive investigations of suspension practices in the state, I have been unable to identify a consistent link between Race and Disability and the risk of being suspended. Instead, I have only been able to identify the disproportionately high rates of suspensions for these two groups. The results from the discriminant analysis suggest that the behaviors of African American students may be different from the behaviors of White students, but the analysis was not robust and I was unable to substantiate any differences. I was also unable to identify any structural characteristics of schools that may have contributed to the disproportionate rates of suspension.

Identifying the underlying factors associated with high rates of suspensions for African American students and students with disabilities is critical to developing programs or policies that can begin to decrease the disparity in the way students from particular groups are suspended. Changes in policies are essential because disciplinary removal from school is known to be associated with school failure, dropout, and involvement in the juvenile delinquency system (Leone et al., 2003). The current policies designed to meet troubling behavior with harsh punishments are ineffective for reducing or eliminating the behaviors, and may exacerbate the problems they are designed to punish (Leone et al., 2003; Costenbader & Markson, 1998), but without clear empirical evidence to

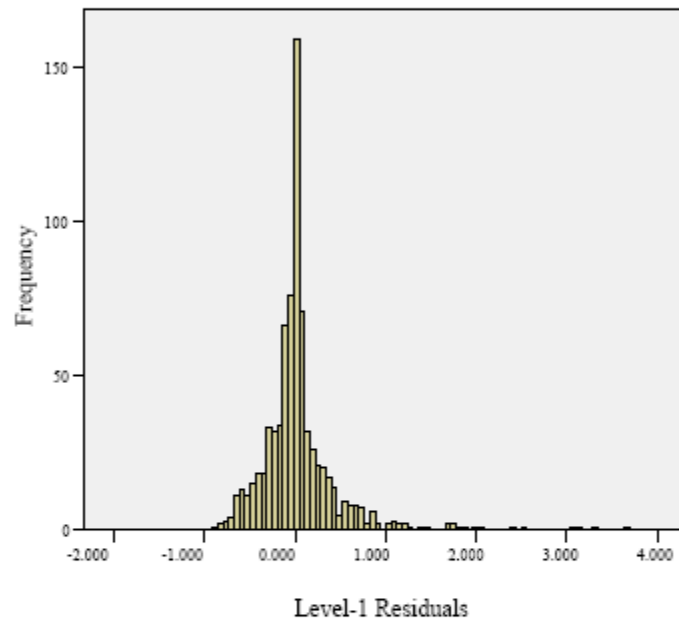
demonstrate how these practices disproportionately affect minority youth and youth with disabilities, zero tolerance policies will likely continue to dominate public school disciplinary policies despite an almost complete lack of documentation to support their effectiveness.

I believe this investigation has contributed to the current body of knowledge of school suspension practices. The findings from this investigation were important because they identified a number of individual and school factors associated with the way students are suspended in Maryland, but more research must be done in order to promote a system of school discipline aimed at decreasing the numbers of students that are suspended, and eliminating the practices that result in disproportionate suspensions of African American youth and youth with disabilities.

Appendix 1. Distribution of Suspensions by School: Maryland 2004



Appendix 2. Distribution of Level-1 Residuals: Suspensions by School in Maryland



Appendix 3. Correlations of Level-2 Predictors from HGLM Analyses

	% White	MSA	Attendance	NHQ	% SPED	Mobility	St-Teach	SES	Enrollment
Suspension Rate	-1.96**	.508**	-.434**	.382**	.176**	.477**	-.148**	.412**	-.130**
Percent White		-.677**	.463**	-.727**	-.064	-.564**	.006	-.645**	-.094
MSA Basic			-.624**	.677**	.270**	.693**	-.008	.745**	-.044
Attendance				-.516**	-.294**	-.817**	-.083	-.487**	-.057
NHQ					.193**	.591**	-.058	.668**	-.118**
Percent SPED						.261**	-.302**	.368**	-.245**
Mobility							-.049	.609**	-.003
Student-Teacher Ratio								-.287**	.542**
SES									-.335**

** Correlation is significant at the .01 level

References

- Bear, G.G. (1999). *Interim alternative educational settings: Related research and program considerations*. Alexandria, VA: Project FORUM, National Association of State Directors of Special Education.
- Benner, G. J., Nelson, R. J., & Epstein, M. H. (2002). The language skills of children with emotional and behavioral disorders: A review of the literature. *The Journal of Emotional and Behavioral Disorders, 10*, 43-56.
- Brady, K.P. (2001). Zero tolerance or (in)tolerance policies? Weaponless school violence, due process, and the law of student suspensions and expulsions: An examination of Fuller v. Decatur Public School Board of Education School District.
- Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist, 32*, 513-530.
- Bronfenbrenner, U. (1989). Ecological systems theory. In R. Vasta (Ed.). *Annals of child development, 6* (pp. 187-251). Greenwich, CT: JAI.
- Brooks, K., Schiraldi, V. & Zeidenberg, J. (2001). *School house hype: Two years later*. San Francisco, CA: Justice Policy Institute.
- Bullara, D. T. (1993). Classroom management strategies to reduce racially-biased treatment of students. *Journal of Education and Psychological Consultation, 4*, 357-368.

- Bridge, R., Judd, C., & Moock, P. (1979). *The determinants of educational outcomes: The impact of families, peers, teachers, and schools*. Cambridge, MA: Ballinger Publishing Co.
- Bruns, E.J., Moore, E., Stephan, S.H., Pruitt, D., & Weist, M.D. (2005). The impact of mental health services on out-of-school suspension rates. *Youth and Adolescence, 34(1)*, 23 – 30.
- Cauffman, E. (2004). A state-wide screening of mental health symptoms among juvenile offenders in detention. *Journal of Child and Adolescent Psychiatry, 43*, 430-439.
- Children's Defense Fund (1975). *School suspensions: Are they helping children?* Cambridge, MA: Author.
- Christle, C.A., Nelson, C. M., & Jolivette, K. (2004). School characteristics related to the use of suspension. *Education and Treatment of Children, 27*, 509-526.
- Civil Rights Project of Harvard University. (2000). *Opportunities suspended: The devastating consequences of zero tolerance and school discipline policies*. Report from a National Summit on Zero Tolerance June 15-16, 2000, Washington, D.C.
- Cooley, S. (1995). *Suspension/expulsion of regular and special education students in Kansas: A report to the Kansas State Board of Education*. Topeka, Kansas State Board of Education

- Costenbader, V.K., & Markson, S. (1994, October). School suspension: A survey of current policies and practices. *NASSP Bulletin*.
- Costenbader, V., & Markson, S. (1998). School suspension: A study with secondary school students. *Journal of School Psychology, 36*, 59-82.
- Fasko, D., Grubb, D. J., & Osborne, J. S. (1995, November). An analysis of disciplinary suspensions. Paper presented at annual meeting of MidSouth Educational Research Association, Biloxi, MS. (ERIC Document Reproduction Service No. ED 393 169)
- Flannery, D. J. (1997). *School violence: Risk, prevention, intervention, and policy*. (Report No. RR93002016) Retrieved September 27, 2002 from <http://ericweb.tc.columbia.edu/monographs/uds109>.
- Gun-Free Schools Act, Public Law 103-382, 108 Statute 3907, Title 14 (1994).
- General Accounting Office (2001). Student discipline: Individuals with Disabilities Education Act: Report to the Committees on Appropriations, U.S. Senate and House of Representatives. Washington DC: Author.
- Gregory, J.F. (1996). The crime of punishment: Racial and gender disparities in the use of corporal punishment in the U.S. Public Schools. *Journal of Negro Education, 64*, 454-462.

- Gunter, P. L., Hummel, J. H., & Venn, M. L. (1998). Are effective academic instructional practices used to teach students with behavior disorders? *Beyond Behavior, 9*(3), 5-11.
- Hosp, J.L., & Reschly, D.J. (2003). Referral rates for intervention or assessment: A metaanalysis of racial differences. *The Journal of Special Education, 37*(2), 67-80.
- Huck., S.W. (2004). *Reading statistics and research* (4th ed.). Boston, MA. Pearson Education Inc.
- Hymowitz, K. S. (2000, Spring) Who killed school discipline. *City Journal*. Retrieved July 15, 2005, from http://www.city-journal.org/html/10_2_who_killed_school_dis.html.
- Imich, I.J. (1994). Exclusions from school: Current trends and issues. *Educational Research, 36*(1), 3-11.
- Individuals with Disabilities Education Improvement Act of 2004(IDEIA),P.L. 108-446,20 U.S.C. § 1400 et. seq.
- Isaac, S. & Michael, W.B. (1997). *Handbook in research and evaluation* (3rd ed.). San Diego, CA. EdITS.
- Krezmien, M.P., Leone, P.L., & Achilles, G.A., 2006. Suspension, race, and disability: Analysis of statewide practices and reporting. *Journal of Emotional and Behavioral Disorders*. (In Press).

- Lee, V.E. (2000). Using hierarchical linear modeling to study social contexts: The case of school effects. *Educational Psychologist, 35* (2), 125 – 141.
- Leone, P.E., Christle, C.A., Nelson, C.M., Skiba, R., Frey, A. & Jolivette, K. (2003). School failure, race and disability: Promoting positive outcomes, decreasing vulnerability for involvement with the juvenile delinquency system. College Park, MD: The National Center on Education, Disability and Juvenile Justice.
- Leone, P. E., Mayer, M. J., Malmgren, K., & Meisel, S. M. (2000). School violence and disruption: Rhetoric, reality, and reasonable balance. *Focus on Exceptional Children, 33*(1), 1-20.
- Lietz, J., & Gregory, M. (1978). Pupil race and sex as determinants of office and exceptional educational referrals. *Education Research Quarterly, 3*(2), 61-66.
- Losen, D.J., Simmons, A.B., Staudinger-Poloni, L., Rausch, M.K., & Skiba, R. (2003). Exploring the link between low teacher quality and disciplinary exclusion. Harvard University Civil Rights Project and Northeastern University Institute on Race and Justice. Boston, MA
- McFadden, A.C., Marsh, G.E., Price, B.J., & Hwang, Y. (1992). A study of race and gender bias in the punishment of handicapped school children. *Urban Review, 24*, 239-251.

- Mendez, L.M., & Knoff, H.M. (2003). Who gets suspended and why: A demographic analysis of schools and disciplinary infractions in a large school district. *Education and Treatment of Children, 26(1)*, 30-51.
- Mendez, L.M., Knoff, H.M., & Ferron, J.M. (2002). School demographic variables and out-of-school suspension rates: A quantitative and qualitative analysis of a large, ethnically diverse school district. *Psychology in the Schools, 39(3)*, 259-277.
- Moles, Oliver C. *Strategies to Reduce Student Misbehavior*. Washington, D.C.: Office of Educational Research and Improvement, U.S. Department of Education, 1989. 187 pages. [ED311 608](#).
- Mooney, P. Epstein, M.H., Reid, R., & Nelson, J.R. (2003). Status and trends of academic research for students with emotional disturbance. *Remedial and Special Education, 24*, 273-287.
- Morrison, G.M., Anthony, S. Storino, M.H., Cheng, J.J., Furlong, M.J., & Morrison, R.L. (2002). In R.J. Skiba & G.G. Noam (Eds.). *New directions for youth development (no. 92: Zero tolerance: Can suspension and expulsion keep schools safe?)* (pp. 45-71). San Francisco: Jossey-Bass
- National Center on Educational Statistics (1998). *Violence and discipline problems in U.S. public schools: 1996-1997* (NCES 98-030). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

National Center for Education Statistics, (2002). National Forum on Education Statistics. *Safety in Numbers: Collecting and Using Incident Data to Make a Difference in Schools* (NCES 2002-312). Washington, DC.: U.S. Department of Education. Prepared by I. Hantman, G. Bairu, A. Barwick, B. Smith, B. Mack, S. Meston, L. Rocks, & B. James of the Crime, Violence, and Discipline Task Force.

National Research Council (2002). Minority students in special and gifted education. Committee on Minority Representation in Special Education, M.S. Donovan & C.T. Cross (Eds.): Washington, DC: National Academy Press.

Nelson, R.J., Gonzalez, J.E., Epstein, M.H., & Benner, G.J. (2003). Administrative discipline contacts: A review of the literature. *Behavioral Disorders, 28*, 249-281.

No Child Left Behind Act of 2001, Pub. L. No. 107-110, 115 Stat. 1425 (2002).

Odom, S.L., Brantlinger, E., Gersten, R., Horner, R.H., Thompson, B., & Harris, K. (2005). Research in special education: Scientific methods and evidence-based practices. *Exceptional Children, 71*(2), 137-148.

- Rabrenovic, G. & Levin, J. (2003, May). The quality of post exclusion assignments: racial and ethnic differences in post- exclusion assignments for students in Massachusetts. Paper presented at the School to Prison Pipeline Conference, Harvard University, Cambridge, MA.
- Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods, Second Edition*. Newberry Park, CA: Sage.
- Raudenbush, S.W., & Bryk, A.S., Cheong, Y.F., Congdon, R., & du Toit, M. (2004). *HLM 6: Linear and Nonlinear Modeling*. Lincolnwood, IL. Scientific Software International, Inc.
- Rosenberg, M.S., Bott, D., Majsterek, D., Chaing, B., Gartland, D., Wesson, C., Graham, S, Smith-Myles, B., Miller, M., Swanson, H. L., Bender, W., Rivera, & Wilson, R. (1994). Minimum standards for the description of participants in learning disabilities research. *Remedial and Special Education, 15*(1), 56-59.
- Rose, L.C., & Gallup, A.M. (2004). The 36th Annual Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward the Public Schools. Bloomington, IN: Phi Delta Kappa. [downloaded from the web August 11, 2005 at <http://www.pdkintl.org/kappan/k0409pol.htm>].

- Scott, T.M., Nelson, C.M., & Liaupsin, C.J. (2001). Effective instruction: The forgotten component in preventing school violence. *Education and treatment of Children, 24*, 309-322.
- Shaw, S. R., & Braden, J. B. (1990). Race and gender bias in the administration of corporal punishment. *School Psychology Review, 19*(3), 378-383.
- Skiba, R. J., & Knesting, K. (2002). Zero tolerance, zero evidence: An analysis of school disciplinary practice. In R.J. Skiba & G.G. Noam (Eds.). *New directions for youth development (no. 92: Zero tolerance: Can suspension and expulsion keep schools safe?)* (pp. 17-43). San Francisco: Jossey-Bass.
- Skiba, R. J., Michael, R. S., Nardo, A. C., & Peterson, R. (2002). The color of discipline: Sources of racial and gender disproportionality in school punishment. *Urban Review, 34*, 317-342.
- Skiba, R.J. & Peterson, R (2000). School discipline at a crossroads: From zero tolerance to early response. *Exceptional Children, 66*(3), 335-346.
- Skiba, R.J., Peterson, R. L., & Williams, T. (1997). Office referrals and suspension: Disciplinary intervention in middle schools. *Education & Treatment of Children, 20*(3), 295-315.

- Skiba, R.J., & Rausch, M.K. (2006). Zero tolerance, suspension, and expulsion: Questions of equity and effectiveness. In C.M. Evertson & C.S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 1063-1089). Mahwah, NJ: Lawrence Erlbaum Associates.
- Skiba, R., Simmons, A., Staudinger, L., Rausch, M., Dow, G., & Feggins, R. (2003, May). *Consistent removal: Contributions of school discipline to the school-prison pipeline*. Paper presented at the School to Prison Pipeline Conference, Harvard University, Cambridge, MA.
- Tabacnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics* (4th ed.). Boston, MA. Allyn & Bacon.
- Tebo, M.G. (2000). Zero tolerance, zero sense. *American Bar Association Journal*, 86(4048).
- Teplin, L.A., Abram, K.M., McClelland, G.M., Dulcan, M.K. & Mericle, A.A. (2002). Psychiatric disorders in youth in juvenile detention. *Archives of General Psychiatry*, 59(1133-1143).
- Thompson, B., Diamond, K.E., McWilliam, R., Snyder, P., & Snyder, S.W. (2005). Evaluating the quality of evidence from correlational research for evidence-based practice. *Exceptional Children*, 71(2), 181-194.

- Townsend, B. (2000). Disproportionate discipline of African American children and youth: Culturally-responsive strategies for reducing school suspensions and expulsions. *Exceptional Children*, 66, 381-391.
- Vavrus, F. & Cole, K. (2002). "I didn't do nothin'": The discursive construction of school suspension. *The Urban Review* 34(2), 7- 111.
- Velez, W. (1989). High school attrition among Hispanic and non-hispanic white youths. *Sociology of Education*, 62(2), 119- 133.
- Walker, H. M., & Severson, H. H. (1992). Systematic screening for behavior disorders (SSBD): Technical manual. Longmont, CO: Sopris West
- Wilkinson, L., and Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *The American Psychologist*, 54 (8), 594–604.
- Zhang, D.; Katsiyannis, A.; Herbst, M. (2004). Disciplinary Exclusions in Special Education: A 4-Year Analysis. *Behavioral Disorders*, 29, 337-347.