

## ABSTRACT

HYDE, JENNIE. The Concurrent Validity of Office Disciplinary Referrals Across a Two Year Period. (Under the direction of Scott Stage, PhD).

The aim of the present study was to examine the concurrent validity of office disciplinary referrals (ODR) as a data based decision making tool for a sample of students who received a complex positive behavior support Tier 2 intervention called Check, Connect, and Expect, and those who did not. To achieve this, the rates of ODRs received by students problem behavior, social skills, and academic skills were examined. Additionally, the rate of ODRs for ethnically and racially diverse students was also examined. Finally, an analysis was conducted to determine if the rates of ODRs differed between students who graduated from the Tier 2 intervention and those who did not. All analyses used a *poisson* based hierarchical nonlinear modeling, which is appropriate for use with ODR data because of the skewed nature of the distribution. Results found no relationship between ODRs and measures of academic achievement and social skills. Students at-risk for externalizing problem behavior and students who are ethnically diverse received more ODRs in comparison to the overall sample average. Students with internalizing problem behaviors received significantly fewer ODRs than the sample average. Additionally, ethnically diverse students continued to receive more ODRs despite completion of the Tier 2 intervention. Overall, results show poor concurrent validity between ODRs and other measures except for externalizing problematic behavior. Implications for schools using ODRs in Tier 2 decisions are discussed and alternative tools for data-based decision making are suggested.

© Copyright 2014 by Jennie Hyde

All Rights Reserved

The Concurrent Validity of Office Disciplinary Referrals Across a Two Year Period

by  
Jennie Hyde

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the degree of  
Doctor of Philosophy

Psychology

Raleigh, North Carolina

2014

APPROVED BY:

---

Scott Stage, PhD  
Committee Chair

---

John Begny, PhD

---

Douglas Cullinan, EdD

---

Mary Haskett, PhD

## **DEDICATION**

This dissertation is dedicated to my parents, James and Malinda, and my husband, Tres.

Thank you Tres for all your love and support! In return, I promise to get a job soon!

Thank you Mom for sparking my love of school and teaching me to treat the children I work  
with like I want my own children treated.

And thank you Dad for sharing all your Broughton stories with me and sparking my love of  
psychology. Even though, after all your years of working in mental health, you weren't too  
fond of psychologists, I know you are proud of me. Wish you were here to see this. I know  
you are looking down from above and you are always in my heart.

I love you all dearly.

## **BIOGRAPHY**

Jennie was born in Rutherford College, North Carolina. She attended the University of North Carolina at Chapel Hill where she earned bachelor's degrees in English and Psychology. After graduation, she worked as a research assistant at the Frank Porter Graham Child Development Institute with the Carolina Fragile X Project. She earned a master's and specialist's degree in school psychology from Appalachian State University. After working for one year with Guilford County Schools, she returned to graduate school to obtain her PhD in school psychology at North Carolina State University.

## ACKNOWLEDGMENTS

A PhD is a group effort and I have had several important people in my group.

First, I would like to express my deepest gratitude to my advisor, Scott Stage.

Without your guidance, patience, and support I would not have been able to complete this program. I appreciate all you have done for me.

A special thank you to my committee members, Dr. John Begney, Dr. Doug Cullinan, and Dr. Mary Haskett. Your guidance and support throughout this process has been invaluable.

I would also like to thank my cohort, Natalie, Jessica, Stacey, Kali, and honorary member Sally. Writing boot camps and library days are bearable with good company. A special thank you to Natalie for magic pasta during particularly difficult times.

## TABLE OF CONTENTS

LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
<b>Chapter One: Literature Review.....</b>	<b>1</b>
<b>Schoolwide Positive Behavior Intervention and Support (SWPBIS): An Overview... 1</b>	<b>1</b>
<b>Office Disciplinary Referrals.....</b>	<b>4</b>
<b>ODRs as a Tier 1 Outcome Measure: Research and Limitations.....</b>	<b>6</b>
<b>ODRs as a Tier 2 Outcome Measure: Research and Limitations.....</b>	<b>12</b>
<b>ODRs for Tier 2 Intervention Identification: Research and Limitations.....</b>	<b>15</b>
<b>Validity.....</b>	<b>19</b>
<b>Construct Validity.....</b>	<b>19</b>
<i>Messick’s unified construct approach.....</i>	<i>19</i>
<i>Nomological network.....</i>	<i>23</i>
<b>Concurrent Validity.....</b>	<b>27</b>
<i>Child Behavior Checklist.....</i>	<i>28</i>
<i>Behavior Assessment for Children.....</i>	<i>28</i>
<i>Social Skills Rating Scale.....</i>	<i>30</i>
<i>Systematic Screening for Behavior Disorders.....</i>	<i>31</i>
<i>Teacher Observation of Classroom Adaptation Checklist.....</i>	<i>32</i>
<b>Ethnicity and ODRs.....</b>	<b>34</b>
<b>Conclusions and Hypotheses.....</b>	<b>38</b>
<b>Chapter Two: Method.....</b>	<b>41</b>
<b>Participants.....</b>	<b>41</b>
<b>Procedure.....</b>	<b>44</b>
<b>Measures.....</b>	<b>48</b>
<b>Screening Measure.....</b>	<b>48</b>
<i>Systematic Screening for Behavior Disorders.....</i>	<i>48</i>
<b>Predictor Measures.....</b>	<b>49</b>
<i>Child Behavior Checklist Teacher Report Form.....</i>	<i>49</i>
<i>Daily Progress Report.....</i>	<i>50</i>
<i>Ethnicity.....</i>	<i>51</i>
<i>Social Skills Rating System-Teacher Report Form.....</i>	<i>51</i>
<i>Woodcock Johnson Tests of Academic Achievement.....</i>	<i>52</i>
<b>Dependent Variable.....</b>	<b>53</b>
<i>ODR.....</i>	<i>53</i>
<b>Statistical Analyses.....</b>	<b>54</b>
<b>Chapter Three: Results.....</b>	<b>55</b>
<b>Statistical Analyses Design.....</b>	<b>55</b>

<b>Null Model</b> .....	55
<b>Main Effects Model</b> .....	56
<b>Main Effects Nested within Graduate Status</b> .....	58
<b>Chapter Four: Discussion</b> .....	61
<b>The Null Model</b> .....	62
<b>The Main Effects Model</b> .....	63
<b>The Nested Model</b> .....	68
<b>The Concurrent Validity of ODRs</b> .....	70
<b>Functional Behavior Assessments (FBAs)</b> .....	73
<b>Poisson Hierarchical Nonlinear Modeling</b> .....	75
<b>Limitations</b> .....	76
<b>Future Directions</b> .....	77
<b>REFERENCES</b> .....	79
<b>APPENDICES</b> .....	104
SWIS ODR Form.....	105
Sample Daily Progress Report.....	106



**LIST OF TABLES**

Table 1.	<i>Frequency and Percentage for Treatment Level across the Two Years.....</i>	92
Table 2.	<i>Demographic Variable Percentages for Participants At-risk by Predictor Variables.....</i>	93
Table 3.	<i>Main Effects Hierarchical Nonlinear Modeling Estimates Predicting ODR Rate.....</i>	96
Table 4.	<i>Hierarchical Nonlinear Modeling Estimates Predicting ODR rate with Significant Main Effects Nested within Graduate Status.....</i>	97

**LIST OF FIGURES**

Figure 1.	<i>The distribution of ODRs within the sample.....</i>	101
Figure 2.	<i>The lognormal distribution.....</i>	102
Figure 3.	<i>ODR counts for non-graduates on externalizing problem behavior.....</i>	103
Figure 4.	<i>ODR counts for graduates on externalizing problem behavior.....</i>	104
Figure 5.	<i>ODR counts by ethnicity for non-graduates.....</i>	105
Figure 6.	<i>ODR counts by ethnicity for graduates.....</i>	106

## **Chapter One: Literature Review**

### **Schoolwide Positive Behavior Intervention and Support (SWPBIS): An Overview**

It is estimated that 16,000 schools nationwide are currently implementing SWPBIS to address the behavioral and emotional needs of students, with all 50 states implementing the initiative at some level (George, 2012). SWPBIS systems are modeled after a tiered public health framework that organizes prevention and intervention services across 3 levels of support. Movement through the tiers should occur through data based decision making (Sugai & Horner, 2006) and office disciplinary referrals (i.e., ODRs) are frequently considered a data based tool to make decisions concerning the level of interventions or behavior supports students received (Sherrod, Getch, & Ziomek-Daigle, 2009; Simonsen, Myers, Briere, 2011; Hawken, MacLeod, & Rawlings, 2007; Todd, Campbell, Meyer, & Horner, 2008). Given the prescriptive use of ODRs to determine level of school success and individual students' placement into more intensive interventions and the professional obligation to use valid forms of data for these purposes (National Association of School Psychologist, 2010), there is a need to determine the concurrent validity of ODRs in relation to other valid measures of student functioning.

Tier 1 of SWPBIS is designed for all students to learn the behavioral rules and expectations of both classroom and non-classroom settings (e.g., cafeteria, or hallways; Horner et al., 2009; Walker et al, 1996). This tier is considered a universal level of prevention, as the student population as a whole receives intervention. The ultimate goal of Tier 1 interventions is to establish a school culture in which appropriate student behavior is

reinforced and problematic behaviors are reduced (Horner et al., 2009). At this tier, school expectations and appropriate behaviors are positively and clearly defined through behavioral expectations the whole school follows. These expectations are stated positively (e.g., Be Respectful, and Work for Excellence; Bradshaw, Koth, Leaf, & Thorton, 2009), rather than negatively (e.g., Do not run in the hallways). Students demonstrating appropriate behavior are reinforced, often by receiving tickets that are linked to rewards (e.g., lunch with the principal or extra recess time). At the Tier 1 level of implementation, consistent strategies for addressing problem behavior are also developed. As a part of this process, staff must decide what type of behavioral violations can be managed within the classroom and what type constitute an ODR (Bradshaw et al., 2010). The intent of these interventions is to prevent students from becoming at-risk for emotional disabilities, which are characterized by problematic behavioral and emotional functioning (Walker et al., 1996).

Tier 2 is composed of more intensive and individualized interventions and is designed for students who need more behavioral support than provided at Tier 1 (Sugai & Horner, 2009). A core difference between Tier 1 and Tier 2 efforts is a shift in the unit of analysis from school-wide at the primary level to the individual student at the secondary level (Walker et al., 1996). General interventions at this level can include small group social skills, class-wide contingency plans, behavior contracts, peer mentors, homework club, at-risk counseling, conflict resolution skills, anger management, self-regulation skills, and academic support (Lane, 2007; Lane, Gresham, O'Shaughnessy, 2002; Todd et al., 2008; Walker et al., 1996). There are standard protocol interventions that have been used at Tier 2, such as First

Steps to Success (Walker et al., 1998) and Check, Connect and Expect (CCE: Cheney et al., 2009). Check-in/Check-out (CICO: Filter et al., 2007; Simonsen et al., 2011) interventions also have been used to reduce problem behavior for students at-risk for emotional disabilities. Within CICO systems, students check-in with a school staff member in the morning, receive behavior feedback throughout the day using a daily behavioral expectation card, and then check-out with the staff member in the afternoon. Reinforcement for appropriate behavior is tied to ratings on their daily behavior expectation report cards.

Tier 3 intervention is for students who are not responsive to either Tier 1 or Tier 2 intervention (Sugai & Horner, 2009); therefore, Tier 3 interventions are more specific to individual student needs. Supports at this level must be comprehensive, long-term, and may involve multiple agents (e.g., family, school, and community) to be successful (Walker et al., 1996). At this level, services are typically provided through special education and an individualized education plan, as the student may be identified as having an emotional disturbance (ED). Students are identified with the special education classification of ED when they display one or more of the following characteristics over a long period of time and to a marked degree: an inability to learn that cannot be explained by intellectual, sensory, or health factors; an inability to build or maintain satisfactory interpersonal relationships with peers and teachers; inappropriate types of behavior or feelings under normal circumstances; a general pervasive mood of unhappiness or depression; and/or a tendency to develop physical symptoms or fears associated with personal or school problems (IDEIA; Individuals with Disabilities Education Improvement Act, 20 U.S.C. § 1412, 2004). As in Tier 2, the level of

problem solving analysis is the individual student (Walker et al., 2009). Three main interventions described at Tier 3 include behavioral intervention plans developed from functional behavioral assessment (Bradshaw, Reinke, Brown, Bevans, & Leaf, 2008; Lane et al., 2002), MultiSystemic therapy (Nelson et al., 2009) and comprehensive school and community interventions delivered through wrap-around services (Eber, Sugai, Smith, Scott, & Terrance, 2002).

In sum, SWPBIS practices are widely used and aim to address the behavior and emotional needs of all students in a tiered approach. Decisions about a student's needed level of support and progress with SWPBIS interventions are made with data based decision making tools, such as ODRs.

### **Office Disciplinary Referrals**

According to Sugai, Sprague, Horner, and Walker (2000) an ODR is defined as the following:

.... an event in which (a) a student engaged in a behavior that violated a rule or social norm in the school, (b) the problem behavior was observed or identified by a member of the school staff, and (c) the event resulted in a consequence delivered by administrative staff who produced a permanent (written) product defining the whole event (p.91).

ODRs are frequently divided into two categories based on the severity of the behavior (May et al., 2000). "Major" ODRs typically involve actions that are a violation of another student or person's rights or are aggressive in intent (e.g., insubordination, blatant disrespect,

inappropriate language, harassment, physical aggression, safety violations, vandalism, theft, truancy, and illegal behaviors, such as smoking marijuana). “Minor” ODRs include behaviors that are less severe than major behaviors (e.g., being tardy or unprepared for class, violating classroom expectations, or lying and cheating; May et al., 2000). The current study focuses on “major” ODR data.

ODR data are used to make decisions in SWPBIS systems, both as an outcome measure for intervention effectiveness and for the identification of students in need of more intensive intervention. However, for ODR data to be valid for these purposes, schools must have clear operational definitions of ODRs and provide teachers with training on the exact types of behaviors that warrant an ODR (Sugai & Horner, 2002). Additionally, before the data can be used for decision making, ODRs must be collected and stored in a systematic manner, such as a web based tracking system to monitor ODRs (Sugai & Horner, 2002; Sugai & Horner, 2006). Schools without these elements in place may not be using valid forms of data when examining ODRs for the purpose of intervention decision making.

One method of collecting valid ODR data is through the use of the School-Wide Information System (SWIS), which is a frequently used web-based tracking system developed by the researchers at the University of Oregon and managed by the Positive Behavioral Interventions and Support Technical Assistance Center (May et al., 2000). SWIS is a web based computer program that is used to collect and report ODR outcome data. Data can be summarized at the individual, group, or whole student body level. Additional information about the location of the problem behavior, students involved in the problem

behavior, and staff that made the referrals can be generated from SWIS. The developers of SWIS report that is designed for decision making at both the individual student and school-wide behavioral support level (May et al., 2000). Currently, more than 8,000 schools use SWIS to manage ODR data (SWIS, 2013). Please refer to Appendix A for copy of a SWIS ODR.

**ODR as a Tier 1 Outcome Measure: Research and Limitations.** Several studies of Tier 1 SWPBIS interventions have reported ODR data as an outcome measure of effectiveness. For example, in response to a high number of ODRs, one school developed *The High Five Program* which taught five positively worded school wide behavioral expectations to middle school students. Students were reinforced with rewards, such as open gym time, for following school wide expectations. Results from this study indicated a 47% reduction in ODRs following the first year of the High Five implementation and a 68% reduction in ODRs five years after implementation. However, the authors of the study did not indicate if a web based tracking system was used to monitor the quality of ODR data (Taylor-Greene & Kartub, 2000).

As another example, Horner et al. (2009) utilized a randomized, wait-list control model to evaluate SWPBIS primary intervention outcomes including the effectiveness of state training, fidelity of implementation, perceived safety levels in the schools, reduction in ODRs, and academic achievement of third graders. Although ODR information was gathered as a part of this study, the authors stated that the data were inconsistent because of the quality of ODRs collected prior to SWPBIS implementation. In another study, Bradshaw, Mitchell,



and Leaf (2010) examined the differences between elementary schools trained in SWPBIS on implementation integrity and student outcomes. Using *School –Wide Evaluation Tool* (SET; Horner et al., 2004), a measure of SWPBIS implementation fidelity, the authors found that most schools could successfully implement SWPBIS with high fidelity within one to three years of initial efforts. According to Bradshaw and colleagues (2010), this long period of time is needed to implement SWPBIS practices correctly, as implementation often involves major organizational changes. In terms of ODR data, results showed schools trained in SWPBIS reported a significant reduction in ODRs across the three year study period. However, comparisons between SWPBIS schools and control schools on ODR data could not be made given the poor quality of ODR data in control schools (e.g., no clear operational definitions, lack of tracking system).

In addition primary grades, ODRs have been used as an outcome measure for SWPBIS programs at the secondary level. For example, Lane, Wehbby, Robertson, and Rogers (2007) examined outcome data on five different groups of high school students (i.e., at-risk internalizing problem behavior, at-risk externalizing problem behavior, comorbid at-risk problem behavior, typical behavior, and high-incidence disability) following Tier 1 SWPBIS interventions. Results from the study showed that the students in the at-risk internalizing, at-risk externalizing, and typical groups demonstrated increases in grade point average and decreases in unexcused tardies and suspensions following Tier 1 implementation. However, the number of ODRs did not significantly change for any of the groups except for the typical group who displayed only a low magnitude decrease. The

authors of the study concluded that ODRs were the least sensitive outcome measure and should not be used as a sole outcome measure when evaluating Tier 1 effectiveness.

In yet another study using ODRs, Lassen, Steele, and Sailor (2006) examined the efficacy of universal SWPBIS at an urban middle school. In this study, one middle school received training in universal SWPBIS. ODR, suspension rates, and reading and math scores on the state achievement test were gathered at baseline and three annual time points after universal implementation. Results showed that over the three years there were significant reductions in ODR and suspension rates. Math achievement scores significantly improved whereas reading scores did not. The authors concluded SWPBIS was effective at reducing problematic behavior after three years of implementation. The authors noted this conclusion depended on the assumption that a reduction in ODR reflected an increase in positive behavior. However, as these authors purported, the reduction in ODRs may actually reflect other contextual factors, such as school policy or teacher tolerance.

Finally, using a treatment and nonequivalent comparison group, Nelson, Martella, and Galand (1998) examined the effects of teaching school wide behavior expectations and establishing consistent consequences on ODRs for both rule and personal rights violations. A rule violation ODR was given for the following disruptive behaviors: late from recess/assembly, running, misuse of group equipment, throwing items/spitting/, gum/candy/food consumption, possession of inappropriate items, and interruption of learning environment. Rights violation included harming/threatening others, using verbal

abuse/inappropriate language, acting in a disobedient/disrespectful manner, destroying property/stealing, fighting/pushing, and harassing others.

In this study, Nelson et al. (1998) had one elementary school complete a baseline year in which traditional discipline practices were used. The following year, only school wide behavior expectations were taught without establishing consistent consequences for problematic behavior for a one year period. During the next year, both school wide expectations and consistent consequences for disruptive behavior were put into place. Archival ODR data for each year were analyzed. Results from the year immediately following baseline showed an increase in ODRs for rule violations while ODR rates for rights violations decreased in comparison to the baseline phase. The authors concluded teaching school-wide expectations may cause teachers to focus more on discretions in the rules violation category, resulting in an increase in ODRs in this category. After the implementation of consistent consequences for behavior, a reduction in ODR rates was seen for both rules and rights violations. According to the authors, this finding suggested systematic responses to disruptive behaviors may be key factor in reducing both rules and rights violations. However, given that the study did not use a condition in which only systematic responses were used, it is impossible to ascertain if only systematic responses would be responsible for the reduction in ODRs, or if the interaction between systematic responses and schoolwide expectations were needed. Overall, the results from this study indicated SWPBIS Tier 1 interventions may not have an impact on all types of ODRs and

that different components of SWPBIS may affect ODR data differentially. However, since the study utilized only one school, the generalizability of results is limited.

In total, the current review found six studies that collected ODR data as a possible outcome measure for Tier 1 interventions (Bradshaw et al., 2010; Horner et al., 2009; Lane et al., 2007; Lassen et al., 2006; Nelson et al., 1998; Taylor-Green & Kartub, 2000). However, there are some drawbacks to using ODRs as a Tier 1 outcome measure. For example, Kern and Manz (2004) indicated that the goal of Tier 1 is to increase pro-social behaviors and improve school culture. A reduction of ODRs may not be a valid measure of these behaviors, as ODRs are typically associated with more problematic behavior. The authors advocated that research examining the effectiveness of Tier 1 interventions should use outcome measures that are more sensitive to school climate and social competency.

Furthermore, as an outcome measure, ODR data may not be as sensitive to change as other measures. Results from the Lane et al. (2007) study showed that the students at-risk for internalizing and externalizing problem behavior, as well as a typical group demonstrated increases in grade point average and decreases in unexcused tardies and suspensions following Tier 1 implementation. However, the number of ODRs did not significantly change following Tier 1 implementation. Therefore, the authors of the study advocated for an examination of multiple outcome data, not just ODRs, when evaluating the effectiveness of Tier 1 interventions. Other sources of data suggested for use in determining intervention effectiveness included standardized measures of behavior and academic achievement, as well as direct observations of behavior. However, despite the need for multiple outcome

measures, some research examining Tier 1 intervention focused solely on ODRs as an outcome measure (e.g., Nakasato, 2000; Taylor-Greene & Kartub, 2000; Taylor-Greene et al., 1997).

Additionally, the quality of ODR data prior to SWPBIS implementation makes some analyses difficult to complete when attempting to determine the effectiveness of Tier 1 implementation. Because a core component of SWPBIS universal intervention is training in ODRs, the quality of ODRs prior to implementation may be drastically different and not suitable for pre and post intervention comparisons. For example, Horner et al. (2009) collected ODR information using SWIS after Tier 1 SWPBIS implementation for 30 schools, as well as 23 schools serving as controls. However, prior to universal interventions, ODR practices across schools were inconsistent, leading to invalid forms of data. Therefore, the authors could not compare pre and post SWPBIS ODR rates or compare control school ODR rates to experimental ODR rates. Bradshaw et al. (2010) noted a similar difficulty. Although the authors used experimental methods to compare SWPBIS and non-SWPBIS schools across a variety of outcome measures, ODR differences between the two could not be analyzed given the quality of data in control schools. Without adequate control groups, the assertion that SWPBIS Tier 1 interventions leads to a reduction in ODRs is difficult to make.

In summary, Tier 1 SWPBIS efforts focus on increasing pro-social behaviors and preventing students from becoming at-risk for behavioral and emotional disorders. Efficacy research at this tier has focused on ODRs either exclusively or in combination with other outcome measures. However, the use of ODRs in this manner may not be appropriate for

several reasons. Specifically, the quality of ODR data may be poor due to no systematic tracking methods or lack of professional development for teachers on ODRs. In addition, the outcome behaviors in Tier 1 may not reflect a school wide reduction in problem behavior. Other measures that are more sensitive to change and more reflective of an increase in pro-social behaviors may be more appropriate as an outcome measure for Tier 1.

**ODRs as Tier 2 Outcome Measure: Research and Limitations.** In addition to serving as an outcome measure for Tier 1 interventions, ODRs are also used as an outcome measure for Tier 2 interventions. For example, Filter et al. (2007) evaluated pre and post intervention ODR rates for students who participated in a CICO program. Nineteen students from one of three elementary schools were selected to participate in the Tier 2 CICO intervention. Each school developed its own method for identifying students in need of a Tier 2 intervention and, although no details were provided, the authors noted that a consideration of ODR data was used in every case.

At the end of the academic year, archival ODR data were obtained from the SWIS system at each school to compare rates of both minor and major ODRs pre and post CICO; however, data for both types of ODRs were only available for 12 participants. These 12 students showed a statistically significant reduction in ODRs while participating in CICO. Specifically, the combined average of ODRs for these students was .90 ODRs per week which was reduced to .59 after CICO. Within these 12 participants, 8 showed a reduction in ODRs, 1 showed an increase, and 3 never received an ODR during the school year. Major and minor ODRs were also examined as separate categories. When only major ODRs were

examined, there were no significant changes in ODR rate per week for the participants. In terms of minor ODRs, participants showed a significant decrease from an average of .65 ODRs per week to .47. Therefore the reduction in the combined ODR category analysis was attributed to the reduction in minor ODRs.

Although Filter et al. (2007) concluded CICO was effective at reducing minor ODRs, meaningful interpretations of these data are limited. First, although part of the selection criteria for participation was ODR data, all participants seemed to demonstrate a low rate of ODRs. The pre intervention ODR group average for minor ODRs was .65, major ODRs .22, and combined .59, which is surprising considering some research has suggested students who need targeted interventions like CICO receive more than 2 ODRs (Horner et al., 2002). Additionally, the interpretation and clinical meaningfulness of low rates of ODRs leads to more questions. For example, what does half an ODR practically mean? Is a reduction in ODRs from .65 to .37 enough to have a meaningful impact on a student's behavioral and academic performance? Finally, the low sample size limits the generalizability of findings. Therefore, although the authors of the study concluded CICO was effective at reducing ODRs, meaningful interpretation and practical consequences of this reduction is limited.

Fairbanks, Sugai, and Lathrop (2007) conducted a similar study examining the effects of a CICO program on ODRs. Ten second grade students were selected to participate in the CICO program based on teacher nomination. While the primary outcome measures were behavioral observation data (e.g., noncompliance and inappropriate physical contact), pre- and post- intervention ODR data were examined as secondary dependent variables. Statistical

analysis focused on the reduction of overall classroom rates of ODRs, rather than on individual students who were participating in the CICO program. Prior to CICO implementation, the average daily rate of ODRs for the second grade classrooms was .85. After CICO implementation, this average daily rate was reduced to .47. There was no mention of a data-based tracking system for ODRs or if teachers received proper training in ODRs; therefore, the quality of ODR data in this study is questionable. Also, the authors did not report if the change in ODRs between pre and post intervention was statistically significant. In addition, it is difficult to ascertain if the individual students participating in CICO demonstrated a reduction in ODRs, as the data were examined on a classroom level.

Together these two studies documented the use of ODR data as an outcome measure for Tier 2 interventions, but not without limitations. For example, the quality of ODR data is questionable without the use of a data management system to ensure quality of ODRs (Fairbanks et al., 2007). Also, meaningful interpretations and conclusions can be difficult to ascertain if ODR rates are low and not reported on an individual level (Fairbanks et al, 2007). Also, both of these studies failed to examine if the reduction in ODRs was maintained across time for students following intervention, to see if students could maintain appropriate behavior as measured by ODR count. Therefore, the sensitivity of ODRs to detect behavior change over time, as a result of a Tier 2 intervention, has not been adequately addressed in research.



**ODRs for Tier 2 Intervention Identification: Research and Limitations.** School staff may track ODR data for an individual student to determine the need for more intensive intervention. Specifically, a student with a significant number of ODRs may be considered to be unresponsive to Tier 1 intervention and in need of a Tier 2 intervention. SWPBIS researchers (Horner, Sugai, Todd, & Lewis-Palmer, 2002) have developed ODR cut points to guide intervention selection decisions. Specifically, students who receive two to five ODRs should be placed in a Tier 2 intervention, and students with six or more should receive a Tier 3 intervention. These cut points were developed through a study that used data collected from SWIS users (Horner et al., 2002). Utilizing SWIS data, approximately 300 elementary schools across 13 states reported the mean percentage for students with 0 to 1 ODRs was 87%, 2 to 5 ODRs was 9%, and 6 or more was 4%. The authors of the study concluded that these percentages were similar to the percentages of students that are typically served at each tier of support in three tiered models (e.g., 80% of students are typically served in Tier 1, 15% in Tier 2, and 5% in Tier 3; Walker et al., 1996). Based on the overlap between these percentages, the grouping of students by ODR rate is frequently used for intervention placement decisions.

The current review found five studies that used ODRs to determine if students were in need of Tier 2 intervention. For example, Sherrod et al. (2009) used a web-based ODR tracking system, School Administrative Student Information, to identify a small group of four students who received three or more ODRs in one semester to participate in a counseling group led by the school guidance counselor. During this small group intervention, students

received eight weekly lessons aimed at increasing pro-social skills. These students demonstrated a significant reduction in ODRs following participation in the targeted intervention.

In another study, a combination of ODR data and teacher nomination was used to identify students to participate in a CICO intervention (Simonsen et al., 2011). Students in this study were selected if they had received two or more ODRs in the previous month, or if a teacher identified students as needing an intervention. The participating school had implemented SWPBIS for several years and used SWIS for tracking ODR data. When selecting participants, if any of the students' ODR was considered a "serious offense", the student was excluded from participation. According to the authors, a serious offense may suggest the student needed a more intensive intervention than CICO. No data were provided to indicate how many students were excluded based on this criterion. Outcome measures in this study included a direct measure of student behavior and the *Social Skills Rating Scale* (SSRS; Gresham & Elliot, 1990). Students who participated in CICO showed a significant decrease in off task behaviors measured by structured classroom observations ( $ES \Delta = -.90$ ). Although some gains were noted on the SSRS scales (i.e., social skills, problem behavior or academic competence), none reached the level of statistical significance. ODR data were not measured as an outcome variable because it was used for identification purposes only.

Hawken and colleagues (2007) used ODR data to select elementary school students to participate in BEP. The Behavior Education Program (BEP; Crone, Horner, & Hawken, 2004) was a pre-cursor to the CICO intervention and contained similar elements. Students

check-in with their teacher in the morning and receive a daily behavior expectation report card (e.g., Be respectful, Be responsible, and Be safe). Students check-out with their teacher at the end of the day and receive reinforcement based on the feedback given on the daily behavior expectations report card. Students were selected if they received three ODRs in a two month period or were nominated by a teacher. The school had implemented SWPBIS for three years and as a part of this initiative provided staff development on the appropriate use of ODRs. However, no web-based tracking system was reported as being used for ODR data. The students who participated in BEP showed a significant reduction in monthly ODR totals. No other types of behavioral data were examined in the study.

Alternatively, some research studies have utilized ODR data in the identification of students without specifying an exact number that warrants intervention. For example, Todd and colleagues (2008) identified students to participate in CICO based on teacher and administration nomination due to “frequency of office visits for disruptive behavior” (Todd et al., 2008, p. 47). Therefore, although ODR data were considered in identification, it was done so in a non-systematic manner. The participating schools had been implementing SWPBIS with fidelity for three years. However, there was no indication in the study that staff received training on ODRs or if a tracking system was in place for the data. The students who participated in the intervention demonstrated a reduction in ODR frequency.

In their description of a school’s attempt to implement SWPBIS interventions, Ervin, Schaughency, Matthews, Goodman, and McGlinchey (2007) detailed selection criteria for Tier 2 interventions, although no intervention was given. The model described paralleled the

Horner et al., (2002) criteria based on the number of ODRs. The authors separated students and found 76% of students fell at the Tier 1 (0 to 1 ODRs), 8.2% at Tier 2 (2 to 5 ODRs), and less than 5% (more than 5 ODRs) at Tier 3. The authors concluded Tier 1 interventions were successful at maintaining the appropriate number of students at the correct level of support based on ODR data.

Burke, Davis, Lee, Hagan-Burke, Kwok, and Sugai (2012) used ODRs to validate a screening measure that could also be used to identify students in need of a Tier 2 intervention. The authors of the study created a screener that was based on the school wide expectations developed as part of Tier 1 SWPBIS. The intent of the screener was to identify students who failed to meet the school-wide expectations in Tier 1 and therefore required more intensive intervention. To validate the school wide expectation screener, researchers examined the predictive validity of the screener with ODR data and behavior ratings from BASC-2 Behavioral and Emotional Screening System (BASC-2 BESS; Kamphaus & Reynolds, 2008) with students across three elementary schools. Confirmatory factor analysis examined the sensitivity to positively identify at-risk students and specificity to identify typically behaving students, using ODR risk status as the criterion variable. To be considered “at-risk” based on ODR data, the student had to have received two or more ODRs in accordance with the classification developed by Horner et al. (2002). The sensitivity of ODR risk status to identify students in comparison to SWPBIS screener was below chance levels for all of elementary schools (i.e., 18% to 42%). Therefore, a reliance on ODRs to identify students at-risk may lead to high rates of students in need who are unidentified.

Taken together, these studies documented the use of ODRs in SWPBIS programs to determine if a student is in need of more intensive intervention. While there are some guidelines (Horner et al., 2002) for possible ODR cut points, these recommendations have not been followed consistently, as some researchers do not specify the exact number of ODRs used in decision making. Additionally, as will be reviewed later, these cut points may not be adequate in separating students by behavioral and emotional functioning. As ODRs are being used to identify students for intervention and to monitor progress with interventions, the validity of the data must be established. The aim of the current study is to examine the validity of ODRs in relation to various predictors in the context of a Tier 2 intervention.

### **Validity**

**Construct Validity.** Given the importance of ODRs to SWPBIS an examination of the construct validity, or the ability of a tool to measure the construct it is intended to measure (Cronbach & Meehl, 1955), is necessary. Several approaches have been utilized in research to establish the validity of ODRs. The following is a review of three studies that have attempted to establish the validity ODRs through Messick's (1995) unified construct approach or a nomological network approach (Cronbach & Meehl, 1955). These studies are often cited as justification in other research for using ODRs.

***Messick's unified construct approach.*** According to Messick (1995), traditional validity concepts, such as criterion related validity and concurrent validity are part of a validation process that must examine the evidence and consequences for using and

interpreting a measure. Therefore, establishing validity is a process that must be grounded in classical psychometric analysis (i.e. criterion validity and concurrent validity) and also examine the following: evidential basis for test interpretation, evidential basis for test use, consequential basis for test interpretation, and consequential basis for test use. The evidential basis for test interpretation and use requires classical construct validity analysis that may be fortified with data examining specific incidents where a measure has been used successfully (Messick, 1995). The consequential basis for a measure examines any moral, ethical, and social implications that could result in using and interpreting the measure, such as placing a student in a more intensive intervention or considering the student for special education eligibility.

In an attempt to establish ODRs as a valid data based decision making tool, Irvin, Tobin, Sprague, Sugai, & Vincent (2004) examined the evidential basis for interpretation validity of ODRs by conducting a literature search to find empirical examples of ODR data being used as measures of school-wide climate, effectiveness of school-wide interventions, and behavior support within a school. For example, to justify the interpretation of ODR data as a measure of school-wide behavioral climate, the authors cited several studies that correlated ODR data to various elements of school-wide behavior climate, such as general student misbehavior and teacher perception of school safety. Similarly, several empirical studies reporting a reduction in ODR data as a result of universal SWPBIS implementation were cited as examples of evidential use. To examine the consequences of using and

interpreting ODRs, the authors cited studies in which ODR data were used for problem-solving or consultation focusing on disruptive behavior with resulting positive outcomes.

The authors concluded that enough empirical studies were found utilizing ODRs as an outcome measure to support the idea there is adequate evidential and consequential basis for using and interpreting ODRs. The authors failed, however, to discuss validity in terms of traditional statistical approaches despite the mandate that a unified approach to construct validity be grounded in psychometric statistics (Messick, 1995). Despite the lack of clear statistical support for the validity of ODRs, this research is frequently cited as the justification for the use of ODR data in other empirical studies (Bradshaw et al., 2010; Bradshaw, Mitchell, O'Brennan, & Leaf, 2010; Ervin et al., 2007; Hawken et al., 2007; McIntosh, Campbell, Carter, Dickey, 2009a; Sugai & Horner, 2006). As a caveat to their research, Irvin and colleagues cautioned interpreting the results as indicating that ODR data are valid for examining individual intervention need. Instead the authors suggested that the results indicated only that ODRs are valid for whole school decision making and using the data for the purpose of Tier 2 identification could lead to false-negatives and false-positives.

An additional study by Irvin et al. (2006) aimed to establish the validity of ODRs by looking at the evidential use of the data. Specifically, the authors argued ODRs are valid for data based decision making based on the rate of actual use, user's perception of usefulness, social validity, and sensitivity to program intervention. The authors, several of whom are developers of SWIS, surveyed elementary and middle school staff about their use of SWIS data including how frequently the data are accessed and how useful the data are in decision

making. Results indicated that over 50% of the schools were accessing ODR data with SWIS on a monthly basis for the purposes of early identification of problem behavior, identification of specific behavior problems, development and problem solving for interventions, and monitoring interventions. When asked about the usefulness of the data on a three point scale, 0 (not useful) to 3 (very useful) across the different purposes for decision making, the mean ratings were low and ranged from 1.12 to 1.71 for both middle and elementary schools. Therefore, on average, schools did not find ODR data useful.

However, even though the usefulness ratings by school personnel were low, Irvin and colleagues (2006) concluded that because SWIS ODR reports were being accessed monthly for data based decision making, it was a valid tool for that purpose. However, the authors concluded that this data can only be generalized to schools using the SWIS data based tracking system and may not even be generalized to all SWIS users, as those surveyed lived in close proximity to the developers of the program. Despite its flaws and limits on generalizability, this research is cited as evidence for validity of ODR use in SWPBIS systems (i.e., Horner et al., 2009; McIntosh, Campbell, Carter, & Dickey, 2009; McIntosh, Campbell, Carter, & Zumbo, 2009).

In conclusion, these two studies (Irvin et al., 2004; Irvin et al., 2006) failed to fully address Messick's (1995) unified approach to construct validity, as a consideration of basic psychometric properties is lacking. Messick stated numerical evidence (e.g., correlations) should form the basis for each of the four domains of unified construct validity. In fact, the concept of unified construct validity relies on traditional concepts of construct validity



including concurrent and content evidence, particularly for the evidential basis for a test (Messick, 1995). For example, to fully examine the evidential basis for test use there must be a consideration for both concurrent and content validity, as well as relevance/utility data. However, Irvin and colleagues (2004, 2006) did not address the basic concurrent and content validity of ODRs through statistical means in either study. Without the examination of convergent/divergent or criterion related validity, the evidential base for test interpretation is missing. Without this information, a unified construct validity of ODRs cannot be established. Therefore, despite its frequent citation as evidence for the construct validity of ODRs, Irvin and colleagues did not adequately address all aspects of the unified approach. The present study addresses some of the weakness of these studies by employing statistical techniques that numerally link ODR counts to predictor variables.

***Nomological network.*** The traditional psychometric model used to validate a construct is through the use of a nomological network (Cronbach & Meehl, 1955). In this approach, a series of relationships are hypothesized between a variable of interest and other related measures. Evidence for construct validity is ascertained by examining the statistical relationship among the various variables within the network of measurement. Within this network, various student predictors that are thought to be linked with ODRs in either a distal or proximal manner are assessed. Evidence for the construct validity of ODRs would be found by establishing that these predictors could explain the variance in ODR data. In an unpublished dissertation, Coleman (2008) used a nomological network to investigate the relationship between ODRs and other student predictor variables. Student variables

hypothesized to be correlated with ODRs included: access to learning (i.e., school engagement or school participation), academic skill (i.e., cumulative GPA or academic skills), student demographic variables (i.e., mobility, free and/or reduced lunch status, gender, and status as an English as a Second Language [ESL] speaker), special education status, and academically gifted status. ODR and student data were gathered for tenth grade students whose school was implementing a three tiered SWPBIS intervention model.

Initial regression analysis found a moderate relationship between ODR data and five of the student predictors including cumulative GPA, special education status, ESL status, free and reduced lunch status, and academically gifted status. Together, these variables explained approximately 26% of the variance in ODR data. An examination of the standardized regression coefficients indicated that cumulative GPA explained a large proportion of the variance. However, following this regression analysis, the author noted that a large percentage (i.e., > 80%) of students in the data had received none or only a small number of ODRs, skewing the data. Coleman hypothesized that the large number of students with a small number of ODRs were masking the relationship between ODRs and student predictor variables.

To address this issue of distribution of the ODR data, a second analysis added student attendance as a variable. To do this, a mean attendance score for all students was calculated. For every five percentage points a student was above that mean, an ODR was added to their existing ODR score. The logic behind this method was that in high school, students' attendance is a reflection of their problematic behavior and ODR data should capture the

students' misconduct. However, the author did not explore the reason for missed days (e.g., illness or family emergency). Instead, all missed days above the average were considered to be an act of disruptive behavior similar to those behaviors that receive an ODR as a consequence. The fallacy in extrapolating data in this way is that the notion that all students missed school for disruptive behavior may not be applicable to the entire population. Therefore, any results based on this extrapolation should be interpreted with caution.

Following the extrapolation of ODR data, only students who received an ODR with the new scoring system were used for analysis, thus eliminating students with zero ODRs and no absences from the analysis. With the addition of this variable, a larger percentage of variance in ODR data was explained by the student predictor variables (i.e.,  $R^2 = .68$ ). Significant student predictor variables included cumulative GPA, mobility, special education status, ESL status, academically gifted status, and gender. As seen with the previous analysis, the standardized coefficient for cumulative GPA was  $-.65$  and much larger than the other significant variables which ranged from  $.10$  to  $.15$ .

Therefore, results indicated that a moderate relationship exists between ODR and student predictor variables only when attendance is added to the ODR data and students with zero ODRs are removed from analysis. However, most of the variance in this relationship was explained by cumulative GPA, with higher GPAs predicting lower rates of ODRs. The remaining significant variables were all demographic variables of students, suggesting ODR rates vary across different types of students. Specifically, minority students, students in

special education, and students with lower socio-economic status were more likely to receive a higher frequency of ODRs.

The initial results from Coleman (2008) were generated from regression analysis with no regard to the assumptions of regression. For example, regression analysis should only be used with data that are normally distributed. Given that a majority of students receive zero to one ODR and a very few students receive many ODRs, the curve is best described as a *poisson* distribution. A *poisson* distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and appears very skewed. Coleman indicated a visual inspection of the data occurred and there was an overrepresentation of students with zero ODRs in the data. However, the author still chose to conduct a regression analysis and report coefficients. The issue of violating a regression assumption was not discussed within the manuscript.

The second analysis with the addition of the attendance variable employed an analysis of variance technique. However, again prior to the analysis, there was no indication that data were examined to ensure that the assumptions of normal distribution and homogeneity of variance were met. Given only a few students are represented in the “two or more ODR” category in comparison to the total student population, each ODR observation is not independent, as having at least one ODR increases the chances of having more. Therefore, because each ODR event may not be independent, the likelihood of heteroscedasticity among residuals increases, making an analysis of variance statistical approach inappropriate (Tabachnick & Fidell, 2007). Therefore, because ODRs do not meet the assumptions needed

for the statistical techniques used in this study, they are likely overestimated, thereby increasing the likelihood of making Type I errors. To address this issue, the current study employed a *Poisson* based hierarchical nonlinear modeling analysis. This method is more appropriate than ones utilized by Coleman (2008) because the distributions used in *Poisson* analysis do not need to be normally distributed.

**Concurrent Validity.** Concurrent validity is the extent to which measures evaluating similar constructs and administered at similar time points are related. To establish concurrent validity of ODRs, several studies have examined the relationship between ODRs and broadband behavior rating scales. Broadband behavior rating scales are an appropriate choice to validate ODRs, as these scales measure a range of behavior and emotional functioning and are validated for those purposes. Additionally, these scales are typically used in the evaluation of behavior and emotional disabilities (Merrell, 2010). When response to intervention models are used for special education eligibility decisions, outcome measures like ODRs are also typically used as part of evaluations for services. A recent study surveying school psychologists found that practitioners find both response to intervention data (e.g., ODRs) and rating scales useful components of emotional disability evaluations (Hanchon & Ryan, 2013). As both ODRs and rating scales claim to measure behavior and are utilized for decision making, the measures should correlate well with each other. A review of studies examining these correlations is discussed below.

***Child Behavior Checklist*** . Nelson, Benner, Reid, Epstein, & Currin (2002) conducted a study examining the use of ODR data in the identification of potential Tier 2 and

Tier 3 students. Comparisons were made between ODR data for students who were rated in the clinically significant or borderline range on the *Child Behavior Checklist - Teacher Rating Form* (i.e., CBCL-TRF; Achenbach, 2001). In this study, correspondence between the CBCL-TRF and ODR was low and led to a high percentage of false negatives when ODRs were used across several behavior domains. For example, the percentage of false negatives between internalizing problem behavior and ODRs was approximately 60% for the at-risk level of behavior and 84% for the clinically significant level of behavior. For externalizing problem behavior, the false negative percentage was 48% for at-risk and 45% for clinically significant. These findings implied that a reliance on ODR data for identification of Tier 2 or Tier 3 students led to the under identification of students in need intervention, particularly students with internalizing problem behaviors.

*Behavior Assessment for Children, Second Edition* . McIntosh, Campbell, Carter, and Zumbo (2009) offered a criticism of Nelson et al. (2002) by claiming that the ODR data examined in their study did not meet the true definition of ODR because the data were not collected or tracked in a systematic fashion that included a consistent operational definition of ODR. McIntosh and colleagues claimed to have such a tracking system in place that provided clear operational definitions of ODRs. The schools used in their study had been implementing SWPBIS for 10 years. As a part of the SWPBIS initiative, the schools used SWIS to track ODR data and teachers were trained on the definitions and uses of ODRs. To determine the relation between ODRs and *Behavior Assessment Scale for Children-2* (BASC-2; Reynolds & Kamphaus, 2004), classroom teachers completed the rating scale for

40 students and ODR data were gathered from the SWIS system. Bivariate correlations between these variables showed a significant correlation between the externalizing problem behavior composite of the BASC-2 and the rate of ODRs ( $r = .51$ ). However, no significant correlations were seen between the internalizing problem behaviors ( $r = -.05$ ) or adaptive behavior scales ( $r = -.22$ ) of the BASC-2 and the rate of ODRs. These findings led the authors to conclude that although ODR data may be a valid method for identifying students who need additional interventions for externalizing problem behaviors, it is not for those who exhibit internalizing problem behaviors. However, similar to Coleman (2008), the authors of this study failed to determine if the assumption of homogeneity of variance was met prior to correlating the variables.

McIntosh et al. (2009) also investigated if BASC-2 scores were different across students grouped based on ODR rate (e.g., 0-1, 2-5, 6 or more per academic year). If grouping students by ODR data is a valid method of differentiating intervention need, as indicated by Horner et al. (2002), then these groups of students should show significant differences in behavioral and emotional functioning on rating scales. In this study, students grouped together by ODR cut-points were significantly different on the BASC-2 externalizing problem behavior subscale but not on the internalizing problem behavior or adaptive behavior subscales. Specifically, there was a large effect on externalizing problem behaviors for students with 2-5 ODRs ( $d = .87$ ) and 6 or more ODRs ( $d = 2.08$ ) in comparison to the 0-1 ODR group. However, BASC-2 scores were not significantly different between students with 2-5 ODRs and students with 6 or more. Therefore, although ODRs

may distinguish between possible students that are not responsive to primary interventions in terms of externalizing problem behavior, the data may not be sensitive enough to distinguish between students needing a Tier 2 (2-5 ODRs) or Tier 3 (6 or more) intervention on a standardized measure of externalizing problem behavior. Additionally, the failure of results to demonstrate any differences between students based on internalizing problem behavior again supports the idea that ODR data is ineffective at identifying students with internalizing behavior problems.

***Social Skills Rating Scale.*** The *Social Skills Rating Scale* (SSRS; Gresham & Elliot, 1990) has also been used to determine if there are social skills differences between ODR groups. Using a sample of students whose schools had been implementing SWPBIS for at least three years, Walker, Cheney, Stage, and Blum (2005) separated students based on the number of ODRs received in two groups: one group with zero to one ODR and one with two or more. Data were not examined for the groups of students with six or more ODRs, as the sample size was too small. Scores between the two groups were examined on the SSRS Social Skills and the Problem Behavior subscales. Students with two or more ODRs had a significantly higher group mean ( $M = 122.3$ ) than students with zero to one ( $M = 113.6$ ) on the Problem Behavior subscale. There were no significant group differences on the Social Skills subscale means, suggesting that students with 0 to 1 ODR were not different from students with more than 2 ODRs in terms of teacher-rated social skills. These results indicated ODRs were more reflective of externalizing problem behavior than social skill



deficits. However, because students with social skill deficits are likely to benefit from SWPBIS interventions, identification methods should include these students.

*Systematic Screening for Behavior Disorders.* Walker et al (2005) studied whether ODR data could differentiate students in need of intervention in comparison to a multi-gating screening procedure, the *Systematic Screening for Behavior Disorders* (SSBD; Walker & Severson, 1992), which is used to identify at-risk students on externalizing and internalizing problem behavior. Across three participating schools, 124 students were identified as at-risk for either a behavior or an emotional disorder on the SSDB. Additional data were collected on 72 of the identified 124 participants including ODR data. An examination of ODR data, collected through SWIS, for the at-risk students showed that a majority (76%) had received 0 to 1 ODRs ( $n = 55$ ), 11 (16%) received 2 to 5 ODRs and 6 (8%) received 6 or more ODRs. Therefore, according to the data, a majority of the students identified as “at-risk” on the SSBD had not received a sufficient number of ODRs to warrant Tier 2 intervention based on the Horner et al. (2002) criteria. Additionally, of the 17 students who had received 2 or more ODRs, all had elevated externalizing problem behavior as measured by the SSBD. None of the students with 2 or more ODRs were identified as having an internalizing problem behavior. However, using the SSBD procedures, 54 of the 124 students identified as at risk were classified as having elevated internalizing problem behavior. Therefore, as seen with previous studies, schools relying on ODRs to identify students as at risk for behavioral and emotional disorders may miss a large percentage of students with internalizing problem behavior.

***Teacher Observation of Classroom Adaptation Checklist.*** Using a large sample ( $n = 8,645$ ) of elementary school students whose schools had been implementing SWPBIS with high fidelity for several years, Pas, Bradshaw, & Mitchell (2011) examined the relationship between ODRs and the Teacher Observation of Classroom Adaptation-Checklist (TOCA-A; Koth, Bradshaw, & Leaf, 2009). The TOCA-C is a checklist completed by teachers that measures a student's level of concentration problems, disruptive behavior, and pro-social behavior. Specifically, this study used a multivariate analysis of variance approach (MANOVA) to determine if students differed on TOCA-C mean subscale scores based on ODR grouping. ODR data were derived from SWIS and were coded in two different ways for analysis. First, data were categorized dichotomously as either 0 to 1 (Low) or 2 or more (Medium). SWIS ODR outcome data were also coded as an interval variable by grouping students by rate of ODRs: 0 – 1 ODRs (Low), 2-4 ODRs (Medium), and 5 or more ODRs (High). MANOVA results demonstrated TOCA-C subscales (concentration problems, disruptive behavior, and pro-social behavior) were significantly different between students' level of ODRs. Students with High rates of ODRs were significantly different from students who were in the Medium or Low group. According to the authors, these results supported the notion that ODR groupings can effectively distinguish between groups of students in terms of concentration problems, disruptive behavior, and pro-social behavior. But, the difference between the Medium (2-4 ODRs) and High group (> than 5 ODRs) was not significant. Therefore, ODRs may identify students who are not responsive to Tier 1 interventions (i.e.,

0-1 ODRs), but the data may not be able to differentiate between students needing Tier 2 (i.e., 2-4 ODRs) and Tier 3 (i.e., more than 5 ODRs) interventions.

In an additional analysis, a response operating curve (ROC) analysis examined if TOCA-A subscales could predict ODR groups at a better than chance level. For this analysis, three different groupings of ODR data were examined separately. One group included ODRs as reported by teachers. One group compared low rates of ODRs (0-1) versus medium/high rates of ODR (more than 2). A third grouping compared low/medium rates (0-4 ODRs) versus high (more than 5) ODRs. An examination of the area under the curve (AUC) statistic for teacher rated ODRs and TOCA-C subscales were as follows: .82 (Concentration), .92 (Disruptive), .85 (Pro-social). For low versus medium/high ODRs, AUC statistics were as follows: .80 (Concentration), .88 (Disruptive), and .82 (Pro-social). For low/medium versus high ODRs, AUC statistics were as follows: .84 (Concentration), .92 (Disruptive), .87 (Pro-social). AUC statistics above .90 are considered excellent and between .80-.89 are considered good. Therefore, these results indicated a moderate to high concordance between subscales and ODR, supporting the idea that ODR groups can identify students at an above chance level in terms of concentration problems, disruptive behavior, and pro-social skills.

However, the TOCA-C is a modified version of a longer, interview format measure. The reliability and validity of the checklist format has not been well documented. Rather, only a few studies have been conducted that used a small sample of participants (Koth et al., 2009). Additionally, unlike rating scales such as the CBCL, the TOCA-C provides scaled scores, but no cut points that would indicate if students are at a clinical or at-risk range of

functioning. Given that the TOCA-C does not distinguish among students in terms of behavioral and emotional psychopathological functioning, the checklist may not be a suitable assessment tool to determine if ODR data are capable of correctly identifying students by level of intervention.

In summary, several broadband scales including the BASC, CBCL, SSRS, SSBD, and TOCA-A have been previously used in literature to establish the concurrent validity of ODRs with mixed results. Overall findings indicated that when students are categorized by ODR risk status, significant differences between groups emerge for externalizing behavior, but not internalizing behavior. Mixed results were found with social skills measures. For example, students group together by ODR cut points do not differ on SSRS social skills ratings (Walker et al., 2005), whereas scores on the prosocial behavior scale of TOCA-A could differentiate ODR risk levels. Therefore, in schools that rely on ODRs as a measure to determine intervention need, students with internalizing behavior difficulties and social skill deficits may be missed. To further examine the concurrent validity of ODRs, the present study used standardized measures of externalizing behavior, internalizing behavior, and social skills using a more appropriate statistical analysis.

**Ethnicity and ODRs.** As stated earlier, validity is the precision of a tool to measure what it is intended to measure. In addition to the precision of measurement, the social consequences that can arise from using a measure should be considered when establishing validity (Messick, 1995). Previous research has documented links between ODRs and ethnicity. These links could lead to potentially negative social consequences, such as

inappropriate special education placement, when ODRs are used as a decision making tool. For example, Coleman (2008) demonstrated a significant relationship between ODRs and student demographics including ESL status. Additionally, a study by Vincent and colleagues (2011) found African-American students were overrepresented in ODR data in both schools using SWPBIS and schools that were not. Furthermore, neither type of school (SWPBIS and non-SWPBIS) showed a reduction in the “disciplinary gap” for minority students after SWPBIS implementation. Instead, results demonstrated African American students were still overrepresented in ODR data despite SWPBIS implementation and training in ODR uses which included clear operational definitions of behavior that warranted an ODR.

This overrepresentation of minority students persists even when classroom and teacher characteristics are controlled for in analysis. For example, Bradshaw, Mitchell, O’Brennan, and Leaf (2010) used a sample of 21 elementary schools that were implementing universal SWPBIS to examine the relationship between student demographics, teacher rating of student behavior, classroom variables (including teacher ethnicity, ratings of classroom disruption), and SWIS ODR data. Results showed African American students were more likely to receive an ODR, even after controlling for teacher ratings of student behavior and teacher ethnicity. A similar analysis was conducted by Pas et al. (2011) using hierarchical linear modeling to examine the effects of student, teacher, classroom, and school characteristics on the odds of a student receiving two or more ODRs. This study demonstrated African-American students were more likely to receive an ODR even after

controlling for teacher characteristics (e.g., gender and ethnicity) and school characteristics (e.g., faculty turn-over, enrollment, and percentage of students with free and reduced lunch).

Several factors could contribute to this overrepresentation. For example, Skiba, Michael, Nardo, and Peterson (2002) studied two commonly proposed mechanisms through which overrepresentation may occur, socioeconomic status (SES) and rates of disruptive behavior, by examining the disciplinary records of over 11,000 middle school students. Disciplinary practices coded during record reviews included ODRs and suspension. Specific behaviors leading to ODRs were examined also. In this study, white students were more likely to receive an ODR for smoking, vandalism, obscene language, and leaving school groups whereas African-American students were more likely to receive an ODR for disrespect, excessive noise, threats, and loitering. Additionally, an examination of suspension data showed that minority students were not as likely to be suspended as white students once they were referred to the office. However, African American students received more suspensions because they were more likely to be referred to the office in the first place. The authors concluded disproportionality for minority students in ODRs contributed to the higher rates of suspension for the ethnic group.

This study did not find support for SES and rates of problematic behavior as explanations for overrepresentation. Specifically, overrepresentation was still present after controlling for SES. Additionally, although not a direct measure of behavior, the authors reported record reviews did not suggest African-American students engaged in significantly more problematic behavior. However, the minority students engaged in different types of

problematic behavior (e.g., disrespect, excessive noise) than white students (e.g., vandalism, obscene language). The authors concluded that behavior engaged in by African-American students may actually be considered less severe but is certainly more subjective than the behavior of white students. For example, it is easier to judge whether or not a student used obscene language than it is to judge excessive noise or disrespect. Therefore, teacher bias may play a role in the interpretation of subjective behavior displayed by minority students, leading to overrepresentation in ODR data.

In conclusion, minority students are overrepresented in ODR data, even when controlling for problematic behavior, classroom characteristics, teacher demographics, and SES. Therefore, other factors, such as teacher bias may contribute to overrepresentation. The implications for overrepresentation in ODR data for SWPBIS programs are significant. For example, if SWPBIS programs are relying on ODR information for intervention selection, ethnically diverse students may be unfairly targeted. This over identification may lead to a form of assessment bias when intervention intensity decisions are based on ODR information. Additionally, when ODRs are utilized as an outcome measure for behavior change, minority students may not demonstrate adequate progress with interventions, as teacher bias may influence the number of ODRs received. If high-stakes decisions, such as special education placement, are being made with response to intervention data, minority students may not be given a fair chance at demonstrating progress when ODRs are used. The present study extended on previous work by examining ethnicity as a predictor variable for ODRs for students at-risk for a behavioral and emotional disorder.

## Conclusions and Hypotheses

In sum, SWPBIS is a widely used practice that often relies on ODRs as a data-based decision making tool. Several studies have claimed to find support for the construct and concurrent validity of ODRs; however, these studies have lacked strong statistical designs. Additionally, the validity of ODRs is questionable as it may not be an appropriate data tool when making decisions for specific populations, such as minority students or students with internalizing problem behavior. Therefore, despite its use in SWPBIS systems, the psychometric validity of ODRs has not been adequately supported in research. The present study examined the concurrent validity of ODRs in the context of a Tier 2 intervention called Check, Connect, and Expect. Specifically, the relationship between predictor variables (academic achievement, externalizing problem behavior, internalizing problem behavior, social skills, ethnicity, and graduate status from a Tier 2 intervention) and counts of ODRs received by students who were at-risk for a behavior and emotional disorder were examined. A *Poisson* based hierarchical nonlinear modeling analysis was used (described in more detail in the Methods section) that addresses some of the limitations of previous research. Specifically, by using a *Poisson* based distribution, ODR counts are preserved and not categorized (e.g., 0-2, 3-5). Additionally, assumptions of normality do not need to be met, as a *Poisson* based distribution is skewed in a manner similar to ODR data. Using the *Poisson* based hierarchical nonlinear modeling, the following hypothesis will be explored:

1. Given the strong correlations between externalizing problem behavior and ODR data, students whose scores on the CBCL Externalizing composite are one standard deviation



or above the mean will have a significant and positive event rate ratio, indicating that these students will receive ODRs at a higher rate in comparison to the overall sample rate (McIntosh et. al., 2009; Walker et al., 2005)

2. Given that ODR data tends to show no significant correlations with internalizing problem behavior, it is expected that students who score one standard deviation above the mean on the CBCL Internalizing composite will not show a significant higher event rate ratio (McIntosh, 2009b; Walker et al., 2005), indicating the acceptance of the null hypothesis for this variable, although it has not been previously tested with a statistical method that evaluated the full range of ODRs in the sample distribution.
3. It is hypothesized that students who score at least one standard deviation below the mean on the Social Skills Scale of the SSRS will have a significant and positive event rate ratio, indicating that these students will receive ODRs at a higher rate in comparison to the overall sample rate (Sherrod et al., 2009; Simonsen et al., 2011; Walker et al., 1995; Walker et al., 1996).
4. It is hypothesized that ethnically and racially diverse students (i.e., African-American and Hispanic students) will have a significant and positive event rate ratio, indicating that they will receive ODRs at a higher rate in comparison to the overall sample rate (Bradshaw et. al., 2011; Vincent et al., 2011).
5. It is hypothesized that students who score one standard deviation below average on the Letter-Word Identification or Math Calculation subtests of the Woodcock-Johnson Tests of Academic Achievement will have a significant and positive event rate ratio, indicating

that these students will receive ODRs at a higher rate in comparison to overall sample rate (Coleman, 2008).

6. Students at-risk on significant predictor variables from a main effects model will show a significantly lower ODR rate compared to the ODR rates of CCE non-graduates, as graduates from the CCE intervention showed improvements on measures of problem behavior (Cheney et al., 2009).

## Chapter Two: Method

### Participants

The sample for the current study included 204 elementary school aged children, grades first through third, who attended 1 of 18 schools in western Washington State and participated in the Check, Connect, and Expect (CCE), a Tier 2 positive behavior support intervention, during 2005-2006 and 2006-2007. Out of these 204 children, 119 (58%) participated in the CCE intervention while 85 (42%) served as comparisons. Students ranged in age from 6 to 10 years ( $M = 7.20$ ;  $SD = .96$ ) at the beginning of the study. Based on the *Systematic Screening for Behavioral Disorders SSBD*, 128 (62%) were identified as being at-risk for externalizing behavior disorders and 76 (37%) were at risk for internalizing behavior disorders. A majority of the sample was male (*male*  $n = 150$ ; 74%; *female*  $n = 54$ ; 26%) and Caucasian ( $n = 86$ ; 42%). The racial diversity of the sample was 15% African-American ( $n = 30$ ), 8% Hispanic ( $n = 17$ ), 6% Asian ( $n = 13$ ), and 0.5% American Indian ( $n = 1$ ). Race data were not available for 57 (28%) of the participants. The participants were fairly evenly distributed across free and regular lunch status with 82 participants (40%) receiving regular lunch and 79 (39%) participants receiving free lunch. Seventeen students (8%) received lunch at a reduced price and lunch status was unknown for 26 students (13%).

In the 2005-2006 school-year, 60 (29%) students had Individualized Education Plans (IEP) indicating that they received special education services. The distribution of disability categories was as follows: 18 students with a specific learning disability (9%), 15 students with other health impairment (7%), 9 students with a developmental delay (4%), 7 students

with an emotional disturbance (3%), 5 students with a speech-language impairment (3%), 3 students with autism (1%), 2 students with an intellectual disability (1%), and 1 student with a hearing impairment (< 1 %). In the second year of the study, 76 (37%) students were being served with IEPs. In the second year of the study, the distribution of primary disability category was as follows: 22 students with other health impairment (11%), 20 students with a specific learning disability (10%), 12 students with a speech-language impairment (6%), 8 students with an emotional disturbance (4%), 7 students with a developmental delay (3%), 3 students with autism (1%), and 2 students with an intellectual disability (1%). The primary disability information was missing for two participants.

Teacher characteristics were also examined. In the first year of the study, the 204 participating students were spread across 123 different classrooms with classroom data missing for 14 participants. Several of the teachers had multiple participating students in their classroom. In 2005, 1 teacher had 5 participating students in the classroom, 2 teachers had 4 students, 15 teachers had 3 students in their classroom, and 27 teachers had 2 participating students in their classroom. The remaining 78 teachers had 1 participating student in their class. In the 2006-2007 school-year, there were 125 teachers. Thirty-seven teachers from Year 1 remained in Year 2 with 9 students in Year 1 remaining with the same teacher in Year 2. Teacher data were missing for four students during the second study year. In addition, 4 teachers had 4 participating students in the classroom, 12 teachers had 3 students in their class; and, 39 teachers had 2 students in their class. The remaining 70 teachers had 1 participating student in the class.

Therefore, across the 2 year period, there were 211 participating teachers. A majority of the teachers were female ( $n = 169$ ; 68%) with 23 males participating (9%) and 19 participants (9%) whose gender was not reported. A majority of teacher participants were Caucasian ( $n = 176$ ; 83%). The ethnicities for the remaining teacher participants were African-American ( $n = 7$ ; 3%), Asian/Pacific Islander ( $n = 5$ ; 2%), Hispanic ( $n = 4$ ; 2%), American Indian/Alaskan ( $n = 2$ ; .5%), and Irish American ( $n = 1$ ; .5%). Nineteen participants (9%) did not provide ethnicity data and several teachers selected more than one ethnicity.

On average, the teachers had 14 years of teaching experience ( $M = 13.9$   $SD = 10.35$  range 0-40), with an average of 7 years of experience at their current school ( $M = 6.75$ ,  $SD = 6.65$ , range 0-29). The most common educational degree reported by teachers was a Master's degree ( $n = 109$ ; 52%). Forty-five teachers (21%) reported having at least one year of course work beyond their bachelor degree. Seventeen teachers (8%) reported having a bachelor's degree, 8 (4%) reported having an education specialist degree, and 1 reported having a doctoral degree (.5%). Nine (4%) reported having a different degree than what has been described and 22 teachers degree status was missing (10%).

When asked about the type of teaching license held, 165 (78%) teachers responded that they had a standard teaching certificate, 18 (9%) had a probationary or provisional license, 1 had an emergency waiver (.5%), 5 indicated "other" (2%), and 22 (10%) did not report their certification status.

The teachers report of the grade level(s) taught were as follows: 66 taught second grade (31%), 65 taught third grade (30%), 45 taught first grade (21%), 43 taught fourth (20%), 8 taught fifth (3%), 1 taught sixth (.5%), and 2 taught kindergarten (.5%). Several teachers reported multiple grade levels and 22 did not report a grade taught. The average number of English Language Learner students in the classroom was approximately 2 ( $M = 1.84$ ,  $SD = 3.08$ ). The average number of special education students was approximately 2 ( $M = 1.60$ ,  $SD = 3.17$ ), and the average the number of general education students was 20 ( $M = 20.04$ ,  $SD = 7.30$ ).

### **Procedure**

The University of Washington at Seattle's Institutional Review Board approved the study. Eighteen schools in three school districts throughout western Washington State participated in the current study. Two schools from the same district were matched on demographic variables prior to the study implementation including school size, percentage of students with Individualized Education Plans (IEPs), percentage of students receiving free and reduced lunch, and race. Within each matched school, one school served as an intervention school and one school served as a comparison. Each district had three intervention and three comparison schools.

Both intervention and comparison schools identified students as at-risk for a behavioral and emotional disorder based on the SSBD, a multi-gating screening procedure. Students were screened in spring and fall of 2005. During the first step of the SSBD, teachers were provided descriptions of externalizing and internalizing problem behaviors and asked to

rank order all their students based on presentation of these behaviors. Two separate lists were generated, one for students who displayed internalizing problem behaviors and a separate one for students who displayed externalizing problem behaviors. From this list, teachers completed the SSBD scale, which measured critical events, adaptive behavior, and maladaptive behavior. Students whose scores on the SSBD scale met the cut-off criterion and for whom parental consent was obtained participated in CCE study. Parental informed consent was sought for all participating students through obtained letters, school conferences, and phone calls. Participating teachers also signed informed consent.

All intervention students started the CCE program at the beginning of fall 2006 in the Basic intervention phase. During this intervention, students were assigned a coach who worked in their school who checked-in with students in the morning and checked-out the student in the afternoon. During the check-in process, coaches and students reviewed a daily progress report (DPR). On the DPR, teachers indicated if students were able to follow school-wide behavioral expectations throughout the day by rating student performance on a three to four point Likert scale. If a student's performance on the DPR was low, coaches problem-solved behavioral incidents during the check-out process. Students were considered successful at the Basic intervention level if they earned more than 75% of possible points on the DPR for more than 80% of days across an 8 week period.

If students were successful at the Basic level, they moved to the Self-Monitoring phase of the CCE intervention. During this phase, students rated their own behavior on the DPR and compared their rating with the teachers' ratings, and if the behavior ratings were

within 1 point on each of the behavior expectations for 8 out of 10 days, the student independently rated their own behavior on the DPR for the next 2 weeks. If the students successfully maintained their DPR average without a negative behavior incident, they graduated from the program. Graduates of the program received monthly behavior feedback, with the coach who received information from their teacher. If students were not successful at this level, they returned to the Basic intervention. Students who were not successful at the Basic level of the program after eight weeks were entered into the Basic Plus intervention. During this intervention, the DPR 75% criterion was reduced by 5% and students were given extra incentives for meeting DPR goals, such as additional time with the coach. Students were also provided with social skill instruction (i.e., *The Stop and Think Social Skills Program*; Knoff, 2001) for work completion if they received low progress reports as determined by their DPR. The Basic Plus intervention lasted four weeks. At the end of this time (i.e., 8 weeks), students who successfully met the 75% DPR criterion returned to the Basic intervention.

The students who were not successful in the Basic Plus intervention after 4 weeks (12 weeks total) entered an Intensive intervention of the CCE. During this phase, functional behavioral interventions were designed to intervene with the purported function of their inappropriate behavior (i.e., Stage et al., 2006). These interventions included differential reinforcement using teacher attention, free time, or group contingencies based on the purported function of the student's inappropriate behavior.



At the end Year 2, project staff, under the supervision of a licensed psychologist, administered standardized academic outcome measures in the spring (late April through May). At this time, teachers completed the CBCL and the SSRS.

Teachers from both intervention and comparison schools were trained on the use of ODRs by a SWIS facilitator using training manual *Becoming Swift with SWIS* (May et. al., 2000). This manual provided information about both the use of the SWIS computer software (e.g., reporting options, data entry) and about general ODR practices. For example, as part of this training, staff learned the differences between behaviors that constitute major and minor disciplinary referrals. Minor ODRs were given for behaviors that are low intensity and are handled at the classroom level (e.g., violation of classroom expectations, tardies). Major ODRs were given for behaviors that are high intensity and involve the violation of other's rights or aggressive acts (e.g., vandalism, insubordination). Students were sent to administration for consequences following a major ODR. The present study recorded major ODRs only. Following the training, it was expected that teachers would be able to accurately determine if a behavior warranted an ODR and complete the appropriate documentation. All ODR data were tracked and maintained in the SWIS data system. Previous research has supported the use of ODRs for identification of students with problematic behavior if procedural routines such as staff training and the use of a data tracking system are in place (McIntosh et al., 2009).

## Measures

### Screening Measure

*Systematic Screening for Behavioral Disorders.* The SSBD (Walker & Severson, 1992) is a multistage screening procedure used to identify students in grades kindergarten through sixth who are at risk for developing an internalizing or externalizing behavioral disorder. The SSBD uses a combination of teacher nomination and teacher ratings of critical events (e.g., fire setting), adaptive behavior, and maladaptive behavior. During the first step of the SSBD, teachers are provided descriptions of externalizing and internalizing problem behaviors and asked to rank order all of their students based on their presentation of these behaviors. For the second step, teachers complete ratings scales on the three highest students in each category (i.e., internalizing and externalizing). This rating scale evaluates critical behavioral events, maladaptive behavior, and adaptive behavior. During the third step, classroom and playground observations are completed for students whose scores are above a cut-off point on the rating scale based on a nationally normed sample. The present study omitted the third step and used only Stages 1 and 2 of the SSBD to identify the students who are at-risk for a behavioral and emotional disorder. Previous research has demonstrated that students who pass Stage 2 are at risk for developing behavioral and emotional disorders at a similar rate compared to using all three stages in screening (McKinney, Montague, & Horcutt, 1998). All students in the study were considered at-risk using the SSBD prior to entry into the study.

Psychometric properties of the SSBD have been well established. Test-retest correlations using Spearman's rho have been established at .88 for externalizing problem behavior classification and .74 for internalizing problem behavior. Discriminant validity analysis has demonstrated that the SSBD can correctly classify students as having internalizing problem behavior, externalizing problem behavior, or neither. Concurrent validity studies using the Walker-McConnell Test and School Archival Records have yielded coefficients above .70 (Walker et al., 1990).

### **Predictor Variables Measures**

*Child Behavior Checklist Teacher Report Form.* The CBCL-TRF (Achenbach, 2001) is a nationally normed, standardized rating scale consisting of 118 items completed by teachers to evaluate behavioral and emotional functioning across several domains for children ages 5 to 18. CBCL-TRF contains eight clinical syndrome scales; however, only the internalizing and externalizing problem behavior composites were used in the present study. The CBCL-TRF classifies behavior composites with T-scores between 60-69 as "at-risk" and "clinically significant" for T-scores 70 and above. T-scores are standardized scores with a mean of 50 and a standard deviation of 10. A student's T-score is calculated with the following formulae:  $T \text{ score} = 10(z \text{ score}) + 50$ . A student's z-score is calculated with the raw score minus the mean of the distribution and then divided by the standard deviation. The present study dichotomized students' T-scores on the internalizing problem behavior and externalizing problem behavior composites with 1 for T-scores at or above 60 and 0 for T-scores below 60.

Test-retest reliabilities for the internalizing ( $r = .90$ ) and externalizing composites ( $r = .95$ ) are high. Moderate to high correlations have been found between the internalizing and externalizing composites and other diagnostic tools including the DSM-IV Check list (internalizing  $r = .59$ ; externalizing  $r = .62$ ) and the BASC (internalizing  $r = .83$ ; externalizing  $r = .88$ ). Several validity studies have shown that the CBCL can distinguish between clinically-referred and non-referred populations (Achenbach, 2009; Ebseutani et al., 2010; Nakamura, Ebseutani, Bernstein, & Chorpita, 2009)

***Daily Progress Report Card (DPR)***. Each intervention school created a DPR to reflect the three to five PBIS behavioral expectations of the individual school. Every students' performance at meeting each expectation was evaluated on a Likert scale three to four times a day. The Likert scale ranged on a scale from 1 (low/try again) to 4 (excellent). Ratings on the DPR determined whether or not a participant graduated from the CCE program. The present study used the students' status either graduate or non-graduate of the CCE program as a predictor variable. Graduate status was determined by a student maintaining 75% of their possible DPR points for 12 weeks. Specifically, the student earned 75% or higher of their DPR points for 80% of the days for 8 weeks followed by 2 weeks of matching their DPR to within 1 point of their teacher's evaluation; followed by 2 weeks of self-monitoring, while maintaining an appropriate level of behavior in accordance with their DPR. Table 2 shows participants' progression and regression through the years as determined by DPR percentage points.

Research studies have examined the psychometric properties of the DPR. Stage, Cheney, Mielenz, Lynass, and Flower (2012) found that level and change of DPR is predictive of change and level of externalizing problem behavior, with CBCL-TRF externalizing behavior ratings decreasing as points on DPRs increase. Inter-rater reliability between teachers and research assistants yielded an overall reliability coefficient of .68 (Lynass, 2010). In addition, CCE graduates whose DPR average was over 75% for 8 weeks showed reductions in internalizing and externalizing problem behaviors (Cheney et al., 2009). Appendix B is an example of one school's DPR.

***Ethnicity.*** The child's reported ethnicity was used as a predictor variable. For the current study, students who were African-American or Hispanic were grouped together to form an "ethnically diverse" variable. These two ethnicity groups were chosen because of their frequency of over representation in disciplinary practices including ODRs (Bradshaw et al., 2010; Vincent et al., 2011). All other ethnicities were combined with the Caucasian students because of limited sample size and a lack of empirical support of their ethnicity as at-risk for increased ODRs.

***Social Skills Rating System-Teacher Report Form.*** The SSRS (Gresham & Elliot, 1990) is a rating scale designed to provide a comprehensive picture of social behaviors. The SSRS teacher form is a 52 item instrument completed by teachers to measures social skills, problem behavior, and academic competency. The Social Skills Scale reflects behaviors of cooperation, empathy, assertion, self-control, and responsibility with a higher score reflecting higher social skills functioning. The SSRS social skills at-risk variable was created for data

analysis by identifying students whose social skills standard score was one standard deviation ( $SD = 15$ ) below the mean ( $M = 100$ ). Therefore, all students with a standard score at 85 or below were considered “at-risk” for social skills functioning.

The SSRS was standardized using a large sample of students ( $n = 4,170$ ), parents ( $n = 1,027$ ), and teachers ( $n = 259$ ) representing 1980 census data demographics. Reliability statistics for the elementary school age form are strong including a measure of internal consistency of .94 for the total score and a test-retest reliability of .85. Comparisons between the SSRS social skills scale and rating scales that capture problematic behavior have been found to be negative and in the moderate range (Social Behavior Assessment Total  $r = -.68$ ; CBCL Total Behavior Problems =  $-.64$ ).

***Woodcock-Johnson Test of Academic Achievement.*** (WJ-III; Woodcock, McGrew, & Mather, 2001). The WJ-III Achievement battery is a set of individually administered measures of academic skills in five curricular areas including reading, mathematics, written language, oral language, and academic knowledge. The present study used the Letter-Word Identification and Math Calculation as measures of academic achievement. During the Letter-Word Identification subtest, individuals are required to read from progressively harder lists of words as a measure of basic reading skills. The Math Calculation subtest measures the ability to perform mathematical computations including addition, subtraction, multiplication, and division. “At-risk” variables for both the Letter-Word Identification and the Math Calculation subtest were created by identifying those students whose standard score was one standard deviation ( $SD = 15$ ) at or below the mean ( $M = 100$ ). Therefore, all students whose

standard scores were 85 or below were considered “at-risk” on the Letter-Word Identification and Math Calculation subtest.

The WJIII was standardized on a large, nationally representative sample of over 8,000 children. Internal consistency reliability statistics for children ages 6 to 9 range from .94 to .98. Validity statistics have demonstrated that the Letter-Word subtest has high correlations ( $r = .88$ ) with cognitive tests that measure the construct of reading and writing ability. Reliability indices for children ages 6 to 9 range from .80 to .96. The math calculation subtest has shown a high correlation ( $r = .92$ ) with cognitive tests that measure the Cattell-Horn-Carroll cognitive ability of quantitative reasoning for children ages 6 to 8.

### **Dependent Measure**

**ODR.** Office disciplinary referral data for the current study were entered and maintained in each schools’ SWIS data system (SWIS; May et al., 2000). SWIS is a web based computer program that is used to collect and report ODR data. The developers of SWIS report the system designed for decision making at both the individual student and school-wide behavioral support level (May et al., 2000). While schools make the distinction between major and minor ODRs, only major ODRs were used in the study because they represent serious problematic behavior reports that many schools use to identify students in need of more intervention and to identify schools overall level of problem behavior. The current study examined the relationship between reviewed predictor variables and the students’ total number of ODRs after two years of intervention by calculating the event rate ratio of receiving an ODR in relation to “at-risk” status on predictor variables.

## Statistical Analyses

The present study used a *Poisson* distribution-based nonlinear modeling analyses (Raudenbush, Bryk, Cheong, & Congdon, 2004) to examine the concurrent validity of ODR data with other predictor variables. A *Poisson* based analysis is appropriate for ODR data for several reasons. First, the *Poisson* distribution is a discrete probability distribution that shows the probability of a given number of events occurring in a fixed interval of time. The distribution describes the likelihood that an event will occur given the mean number of events that have occurred and assumes that the mean value of an event will be non-negative.

*Poisson* distribution based analyses have been used with crime rate data (Osgood, 2000) because these events, like ODRs, occur infrequently by most of the population but a large number of crimes can be committed by a few, which results in a non-normal distribution. Also like ODRs, crime rates tend to not be independent acts as recidivism increases the likelihood that one person may be responsible for several crimes. Given the similarities between ODR and crime rate data, a *Poisson* based distribution of data is a good fit for statistical analysis with ODR data. The analysis yields an event rate ratio for each predictor variable that is the number of events that occurred in relation to the predictor variable entered. To calculate the event rate ratio, a *Poisson* statistical based distribution use a logarithm of ODR count data to link the distribution to the lognormal distribution. The loglink gives the expected value of an ODR counts in relationship to specified predictor variables. Figure 1 displays the ODR frequency count in this sample and shows the distribution to be non-normal and similar to the lognormal distribution showed in Figure 2.



### Chapter Three: Results

The percentage of at-risk students for each predictor variable by each demographic variable was analyzed with Fisher's Exact Test to determine the statistical significance of each relationship. See Table 2 for the percentages and significance. For the CBCL Externalizing, CBCL Internalizing, and WJIII Math Calculation variables data were available for 203 participants. On WJIII Letter-Word Identification, data were available for 204 participants. For SSRS, data were available for 201 participants. Ethnicity data were available for 204 participants. Participants whose ethnicity was unknown were grouped with the non-ethnically diverse group for analysis.

#### Statistical Analyses Design

**Null model.** To test hypotheses one through six, ODR data were collected from the each school's SWIS data system at the end of each school year from 2005-2006 to 2006-2007 and summed to calculate a total ODR score. A null model was conducted to determine the event ratio for receiving an ODR in the sample without taking any of the predictor variables into account, based on the mean number of ODR events in the sample. The two-level model below shows the equation for that model:

Level-1 Model

$$E(Y|B) = L$$

$$V(Y|B) = L$$

$$\log[L] = B_0$$

Level-2 Model

$$B_0 = G_{00} + U_0$$

The model shows that E (the event rate) is the sum of individual student's ODRs (Y) by total rate (B) which equals the "true" rate of ODRs (L). The Level-1 equation below the true rate equation represents the variability about the true rate (V). These equations are converted to the [L] log of the event ratio predicted by the intercept (B0) of the other predictor variables represented by the letters G<sub>i</sub>, although in the null model no other variables were entered to change the event ratio dependent on their influence. Therefore, the null model provided the overall average number ODRs per student and the confidence interval about the average number of ODRs assigned for the sample in the study across two years of exposure. This model yielded a significant event ratio of 7.22 ( $p = .008$ , 95% CI [3.20 – 16.12]), indicating that students in this sample averaged 7.22 ODRs over the two years.

**Main Effects Model.** The model below shows the formula and two level model used to test the main effects of the predictor variables:

Level-1 Model

$$E(Y|B) = L$$

$$V(Y|B) = L$$

$$\log[L] = B0 + B1*(ETHR) + B2*(SS) + B3*(CAL) + B4*(LW) + B5*(EXT) + B6*(INT)$$

Level-2 Model

$$B0 = G00 + U0$$

$$B1 = G10$$

$$B2 = G20$$

$$B3 = G30$$

$$B4 = G40$$

$$B5 = G50$$

$$B6 = G60$$

The Log(L) represents the log of the number of ODRs that each student receives. B0 is the intercept coefficient or the logarithm-transformed ODR values which are converted into an event rate ratio for a student in this sample receiving an ODR across the two years that the data were collected. Coefficient B1 (ETHR) represents students who are identified as either African-American or Hispanic. Coefficients B2- B6 represent event rate ratios for at-risk levels of social skills (i.e., SS; SSRS Social Skills Scale), math calculation (i.e., Cal; WJ-III Math Calculation), letter-word identification (i.e., LW; WJ-III Letter Word Identification) externalizing problem behavior (i.e., EXT; CBCL Externalizing Problem Behavior Scale), internalizing problem behavior (i.e., INT; CBCL Internalizing Problem Behavior Scale). CBCL Externalizing and CBCL Internalizing composite scores gathered at the end of Year 2 were transformed into a dichotomous variable of either “at-risk” (T-score  $\geq 60$ ) or “no risk” (T-score  $< 60$ ). For the WJIII and SSRS, subtest scores gathered at the end of Year 2 were transformed into a dichotomous variable of either “at-risk” (standard score  $\leq 85$ ) or “no risk” (standard score  $> 85$ ). For the ethnicity variable, African-American and Hispanic students were considered “at-risk” and all other ethnicities were considered “no risk”. This model tested the relative event ratios defined by the main predictor variables explained above.

The intercept in this analysis was not significant indicating the predictor variables in the equation better accounted for the change in event rate ratio than the overall average change. This model reliably estimated 98% of the variance associated with the variables in the model, with significant event ratios for diverse ethnicity/race, externalizing problem behavior, and internalizing problem behavior. For diverse ethnicity/race, the event rate ratio

was 1.72, with a 95% confidence interval of between 1.54 and 1.93, indicating that students who were African-American or Hispanic received on average 1.72 more ODRs by virtue of their race or ethnicity than the sample average of 1.58 ODRs (i.e., the fixed effect intercept). Therefore, the average ODR count for these students is 3.3 ODRs across the two year period.

Students at-risk for externalizing problem behavior received an event ratio of 6.90 with a 95% confidence interval of 5.73 to 8.31, indicating these students received 6.90 more ODRs on average in comparison to the sample average of 1.58. Therefore, across the two year period these students received on average 8.48 ODRs. For at-risk internalizing students, the event ratio was less than one (event rate ratio = .86, 95% CI [.77, .96]), indicating that these students had less than one ODR when compared to those without elevated levels of internalizing problem behavior in the sample average. Non-significant event rate ratios were found for social skills (event rate ratio = .90, 95% CI [.80, 1.01]), math calculation (event rate ratio = 1.14, 95% [.96, 1.35]), and letter word identification (event rate ratio = .96, 95% [.82, 1.13]), Table 3 displays complete statistics for the main effects model. In addition, the random effect was significant indicating that there was significant variability in individual students' event rate ratios within the second level of model tested, or graduate status.

**Main Effects Nested within Graduate Status.** To answer hypothesis six, another hierarchical nonlinear model was conducted using the significant main effect predictors identified in the previous analysis nested within graduate status. Only significant main effects were entered into the model because when all predictor variables were entered into the model

with nested effects, the model was overidentified and sphericity was detected. The following equation shows the formula for this data analysis:

#### Level-1 Model

$$E(Y|B) = L$$

$$V(Y|B) = L$$

$$\log[L] = B0 + B1*(ETHR) + B2*(EXT) + B3*(INT)$$

#### Level-2 Model

$$B0 = G00 + U0$$

$$B1 = G10 + G11*(GRAD)$$

$$B2 = G20 + G21*(GRAD)$$

$$B3 = G30 + G31*(GRAD)$$

The Log(L) represents the log of the number of ODRs that each student receives. B0 is the intercept coefficient or the logarithm-transformed ODR values which are converted into an event rate ratio for a student in this sample receiving an ODR across the two years that the data were collected. Coefficient B1 represents students who were identified as either African-American or Hispanic. The coefficients B2 and B3 represent students who are at-risk for externalizing and internalizing problem behaviors, respectively. G11 is the slope or test of the change in the event rate ratio of being in the CCE graduate group within the diverse ethnicity/race (ETHR) variable. Likewise, G21 and G31 are the slopes or tests of the changes in the event rate ratios of being in the CCE graduate group on externalizing problem behavior (EXT) and internalizing problem behavior (INT), respectively. This model tested the relative event ratios defined by the significant predictor variables nested within graduation status.

The main effects of ethnicity and externalizing problem behavior continued to be significant, whereas internalizing problem behavior was not. The only predictor variable that continued to be significant when nested in graduate status was the ethnicity/race variable. For ethnicity and race, African American and Hispanic graduates of the CCE program had a significant main effect event rate ratio of 1.52 (95% CI [1.34, 1.73]) and a nested event rate ratio of 2.07 (95% [CI 1.58, 2.72]). The nested effect is additive to the main effect so the count was increased from the intercept of 1.39 by 2.07 ODRs, indicating African-American and Hispanic students who graduated from CCE received on average 4.98 ODRs across the two year period. The nested effect of students who graduated and also had externalizing and internalizing problem behavior was negligible. In addition, the test of random effects was significant indicating that the students in the nested model showed a significant amount of variability in their event rate ratios about the predictor variables used in this model. Table 4 shows the results of the significant main effects with nested graduate effects.

## Chapter Four: Discussion

The aim of the present study was to examine the concurrent validity of ODRs as a data based decision making tool in relation to other measures taken from the literature predicting student performance within a sample of students who received a complex Tier 2 intervention and those who did not. Establishing the validity of ODRs is necessary as the data are often used to make decisions in SWPBIS systems that also have important social consequences (Messick, 1995). The present study employed hierarchical nonlinear modeling with a *Poisson* based distribution analyses to examine the main effects of internalizing problem behavior, externalizing problem behavior, social skills deficits, and academic skills deficits. The effects of ethnicity and race on ODR rates were examined as well, as previous research has demonstrated African-American and Hispanic students are often overrepresented in a variety of disciplinary practices in comparison to Caucasians (Bradshaw et al., 2011; Skiba et al., 2002; Vincent et al., 2011; Pas et al., 2011). Additionally, an analysis was conducted using significant main effect predictors to determine if students who graduated from the CCE intervention changed in their event rate ratio of ODRs received within the at-risk problematic behaviors tested. Graduates of the intervention have demonstrated improved performance on other measures of behavior (e.g., on externalizing problem behaviors and internalizing problem behaviors, Cheney et al., 2009) and therefore should have a significantly different number of ODRs than non-graduates. Overall, the concurrent validity of the predictor variables showed that the only variable that was robustly associated with ODRs was at-risk status on externalizing problem behavior. However, there

were other important predictive variables that contributed to a reliable association to ODRs that are discussed below.

### **The Null Model**

An initial statistical model, the null model, was conducted to determine if the rate of ODRs in this sample was significantly greater than zero without specifying any predictor variables. This analysis yielded a significant event rate ratio of 7.22, indicating students received an average of 7.22 ODRs across the two year period, or 3.61 ODRs on average per year.

This average aligns with the ODR cut point recommendations proposed by Horner et al.'s (2002) research. Specifically, SWPBIS is based on a triangular framework that organizes prevention and intervention efforts by level of need, with most students' needs being met with universal prevention at the bottom of the triangle and fewer students needing intensive intervention at the top. According to this model, SWPBIS systems should meet the needs of 80% of the school population in Tier 1, 10-15% in Tier 2, and 5-10% in Tier 3 (Walker et al., 1996). As previously reviewed, a common model for identifying students in need of intervention using ODRs is based on this triangle (Horner et al., 2002). In this model, students with 0-1 ODRs are considered responsive to primary SWPBIS interventions, students with 2-5 ODRs are in need of Tier 2 intervention, and students with 6 or more are in need of Tier 3 intervention. Results from the null model align with Horner and colleagues' suggestion (2002) that, on average, these students would meet that criterion of being at-risk and in need of Tier 2 intervention. However, as discussed below, when the main effects of



predictor variables were added to the null model, the concurrent validity of using ODRs to identify at-risk students is questionable in certain behavioral domains.

Several of the previously reviewed studies used the Horner et al., (2002) model to form groups of students based on ODR risk status for data analysis (McIntosh et al., 2009; Pas et al., 2011; Walker et al., 2005). However, the current study used a *Poisson* based hierarchical nonlinear model to predict the actual number of ODRs received without categorizing the data into presupposed levels of problem behavior. Preserving rather than categorizing the raw ODR data makes the statistical analyses in the present study a more precise measure of ODRs than the previous work.

### **The Main Effects Model**

The second statistical analysis conducted compared the rates of ODRs received by students who were at-risk on predictor variables measuring externalizing problem behavior, internalizing problem behavior, social skills deficits, academic skills deficits, and ethnic and racial diversity. Results from this model indicated that the ODR event rate ratios varied by risk status for some variables, but not all. Specifically, measures of academic and social skills did not have significant event rate ratios, indicating these students did not demonstrate a relative rate change in ODRs compared to the sample based on risk-status for these variables.

Failure of the present study to find an increase rate of ODRs based on at-risk social skills behavior is consistent with some previous research in which students grouped together by ODR risk cut point did not differ significantly on the Social Skills Rating System (i.e., Nelson et al., 2002; Walker et al., 2005). Therefore, schools that rely only on ODRs as a

method for identifying students in need of social skill instruction may miss students who do not present with accompanying externalizing problem behavior. Additionally, if the aim of Tier 2 intervention is to improve social skills, ODRs will not be a valid outcome measure of behavior change, as ODRs may not be reflective of social skills deficits. The Office of Special Education Program's Center for Positive Behavior Interventions and Support indicates that social skills instruction should be an important part of Tier 2 interventions and recommends measures such as teacher rating scales and direct observation when identifying students at-risk (Lewis, 2008). Additionally, the center recommends ODRs as a data based measure when working with students with at-risk social skills behavior (Lewis, 2008). However, results from this study would suggest alternative methods (e.g., rating scales, direct observations) should be used.

Another variable tested was at-risk performance on measures of academic achievement, as the literature would suggest these students would have a higher average of ODRs in comparison to the sample average. Academic achievement is an essential component of special education eligibility evaluations as students must demonstrate an academic deficiency related to their disability before they can qualify for services (IDEIA; Individuals with Disabilities Education Improvement Act, 20 U.S.C. § 1412, 2004). Additionally, when students receive an ODR, they usually leave the classroom to go to the office and this loss of instructional time is suggested to lead to academic skill deficits (Scott & Barrett, 2004). The literature shows an association between academic achievement and behavior. For example, a meta-analysis conducted by Reid, Gonzalez, Nordenness, Trout, and

Epstein (2004) examining academic performance and students with emotional disabilities found that across the 26 studies, the effect size of emotional disability on academics was  $-.69$ . Effect sizes in the area of math and reading were  $-.81$  and  $-.61$ , respectively. Academic achievement in these studies was measured using a variety of standardized assessments, including the Woodcock-Johnson Tests of Achievement. The failure of the current study to establish a relationship between ODRs and academic achievement may suggest that rudimentary academic skills such as word reading and math computation are not necessarily related to school misbehavior resulting in ODRs.

The failure to find a relationship between ODRs and academic skills contradicts previous research (Coleman, 2008; Lane et al., 2007). However, this inconsistency may be a result of the operational definition of academic performance and sample populations. The present study used performance on a standardized measure of reading and math achievement, whereas previous research has utilized GPA as an outcome measure (Coleman, 2008; Lane et al., 2007). Previous studies examining GPA in the context of a SWPBIS system also used secondary education populations, making the calculation of GPA more appropriate than in primary grades. Together, these findings suggest that the relationship between ODRs and academic achievement may be salient only in secondary education settings.

However, significant event rate ratios were found in the main effects model for students who are at-risk in the areas of externalizing and internalizing problem behavior, as well as for students who are ethnically and racially diverse. These results showed the

students who are at-risk in these areas received a significantly different number of ODRs than the sample average.

Students at-risk for externalizing problem behavior had an event rate ratio of 6.90. Therefore, students who were at-risk for externalizing problem behavior ( $T \geq 60$ ) were likely to receive 6.90 more ODRs (95% CI [5.73 -8.31]), on average, compared to the sample rate average of 1.58 (i.e., fixed effects intercept) without specification of any other at-risk status. The results of this study supported previously established links between externalizing behavior and ODRs (McIntosh et. al., 2009; Walker et al., 2005). Given these relationships between ODRs and externalizing problem behavior, results from the present study supported the concurrent validity of ODR and externalizing problem behavior. Therefore, ODRs may be an appropriate decision making tool for students in this behavior domain. However, as discussed below, the ethnicity and race of a student should be taken into consideration when using ODRs to make decisions.

Specifically, the main effects model found a significant event rate ratio for students who were ethnically and racially diverse (i.e., African-American or Hispanic). Ethnically and racially diverse students were likely to receive 1.72 more ODRs (95% CI [1.54-1.92]) on average compared to sample rate average of 1.58 (i.e., fixed effects intercept). However, it is important to note the magnitude of the difference is not large (i.e., 3.3 versus 1.58). This finding is similar to other research documenting the overrepresentation of minority students in ODRs (Bradshaw et al., 2010; Coleman, 2008; Vincent et al., 2011; Pas et al., 2011). Therefore, schools that rely solely on ODRs to identify students at-risk for behavioral and

emotional disorders may over identify minority students. The implications of the use of ODRs with ethnically and racially diverse students are discussed below.

In addition, the main effects model yielded a significant event rate ratio for students with elevated internalizing problem behavior. Specifically, this group had an event rate ratio that was less than one, indicating that these students were significantly less likely to receive an ODR in comparison to the sample average. The findings from this study are consistent with previous research that has demonstrated that students with internalizing behavior disorders do not frequently receive ODRs (McIntosh et al., 2009; Nelson et al., 2002; Walker et al., 2005). Therefore, if ODR data are used as the only means of identifying students at-risk for a behavioral and emotional disorder, those with internalizing disorders may be missed. Identification of these students is crucial as children and adolescents with internalizing disorders are at-risk for a variety of negative outcomes. For example, children and adolescents with depression have demonstrated poor social well-being and academic performance (Verboom, Verhurst, Sijtsema, & Penninx, 2014); in addition, childhood depression has been associated with adult outcomes including suicidal ideation and marital dissatisfaction (Lewinsohn, Rohde, Klein, & Seely, 1993). Similarly, young children with anxiety problems demonstrate persistent peer difficulties (Danzing et al., 2013) and lower academic performance (Grills-Taquechel, Fletcher, Vaughn, Denton, & Taylor, 2014). As an identification tool, ODRs may not accurately identify students with internalizing disorders. Schools should consider alternative methods to identify these students, such as standardized universal behavioral screenings.

### **The Nested Model**

The nested model tested significant predictors from the main effects model nested within graduate status to determine if the rates of ODRs varied for students who graduated from the CCE program and were still at-risk and those who did not graduate. For internalizing problem behavior, neither the main effect nor the nested effect was significant in this model. Therefore, these students did not receive a significantly different number of ODRs when they graduated from CCE. Given that ODR rates for students at-risk for internalizing problem behaviors were below an average event rate ratio of one across the two year period for both the CCE program graduates and for all the students in the sample, it seems apparent that behaviors indicative of internalizing problem behaviors (e.g., socially withdrawn, cries easily, and makes self-deprecating comments) are not the type of behaviors for which teachers assign ODRs.

In the nested model, the main effects of ethnicity and racial diversity and externalizing problem behavior remained significant. For externalizing problem behavior, the main effect event rate ratio in this sample was 7.29 (95% CI [5.55-9.58]), indicating these students received a significantly higher number of ODRs in comparison to the sample average of 1.39 (i.e., fixed effects intercept). When the nested within graduate status was tested, the event rate ratio was 0.93 (95% CI [.65, 1.33]). The nested effect was negative and therefore represents a subtraction in the event rate ratio. However, it did not reach statistical

significance. One potential issue in this result is that only 21 of the 109 students or 19% of the students at-risk for externalizing problem behavior were also graduates of the CCE program, resulting in a lack of statistical power to find the difference. Therefore, although students who graduated from the CCE program but were still at-risk for externalizing problem behavior received less ODRs than those who did not graduate the reduction is not significant.

Post hoc crosstab analyses were conducted to examine ODR frequencies for graduates and non-graduates by risk status for externalizing problem behavior. Figure 3 displays ODR frequencies for non-graduates of CCE separated by externalizing problem behavior risk status. Figure 4 displays ODR frequencies for graduates of CCE by externalizing problem behavior risk status. An examination of these figures shows fewer ODRs for students who are at-risk for externalizing problem behavior and for those who graduated from the CCE program compared to the non-graduates.

For ethnically and racially diverse students, the nested main effect yielded an event rate ratio of 1.52 (95% CI [1.34, 1.72]). When ethnicity and race are nested in graduate status, the event rate ratio raises to 2.07 (95% CI [1.58, 2.72]) ODRs, indicating African-American and Hispanic students who graduated from CCE had more ODRs on average than those who did not. Specifically, when the nested effect of graduate status is added to the main effect, ethnically diverse students received 3.59 more ODRs across the two year period, for a total of 4.98 ODRs. This number is more than double the sample average of 1.39. Figures 5 shows ODR counts for non-graduates by ethnicity and Figure 6 shows ODR counts for

graduates by ethnicity. As with the data for externalizing problem behavior, these figures show that only a small number of ethnically and racial diverse students graduated from the program ( $n = 10$ ; 5%).

An examination of ODR counts for non-graduates on externalizing problem behavior (Figures 3) and ODR rates by ethnicity for graduates (Figure 6) suggests this sample contained a smaller number of students that received a large number of ODRS across the two year period. These students are colloquially referred to in SWPBIS systems as “high flyers” (George, 2012; George & Barrett, 2011). For example, the ethnically diverse graduate group contained one student who received 68 ODRs across the two year period (Figure 6). In comparison, for the “other” ethnicity group of graduates the highest number of ODRs received by one student was 28. The large addition to the event rate ratio of ethnically diverse graduates is partially attributable to this high flyer student. Figure 3 shows that the non-graduate ethnically diverse sample had one high-flyer student who received 85 ODRs. However, the non-graduate “other” ethnicities and races sample had one high flyer with a somewhat comparable number of 67 ODRs across a two year period. Among the top three students with the highest number of ODRs (85, 67, 68), two were ethnically or racially diverse.

### **The Concurrent Validity of ODRs**

In sum, results showed students at-risk for externalizing problem behavior received a higher number of ODRs, suggesting concurrent validity with a measure of externalizing problem behavior. Therefore, ODRs appear to be an appropriate data based decision making



tool for these students. However, concurrent validity was not established for measures of internalizing problem behavior, social skills deficits, and poor academic performance. In fact, students at-risk for internalizing problem behavior received significantly fewer ODRs. Therefore, students presenting with behavioral and emotional needs indicative of internalizing problem behaviors (e.g., socially withdrawn, cries easily, and makes self-deprecating comments) will not be identified if ODRs are the sole measure of behavior change.

Additionally, the consequential validity (Messick, 1995) of ODRs is questionable given the ethnic and racial disparity seen in these results. Specifically, the main effect of ethnicity and racial diversity suggested that African-American and Hispanic students received more ODRs than the sample average. Therefore, the sole use of ODRs in decision making may lead to an overrepresentation of ethnically diverse students in Tier 2 or Tier 3 interventions. This problem is further compounded when these students received a higher number of ODRs despite successful completion of a Tier 2 intervention, as the nested model illustrated. Therefore, if ODRs are used as an outcome measure, these students may be targeted for more intensive intervention and perhaps may be considered for special education eligibility if other information is not collected.

As response to intervention (RTI) systems continue to grow in usage, schools will rely on data generated from this framework to make special education placement decisions. In a recent survey of school psychologists the usefulness of response to intervention data during emotional disturbance (ED) evaluations was rated on average as 4 (very useful) on a

5-point Likert scale (Hanchon & Allen, 2013). There are some benefits to using this model in eligibility determination. For example, Harris-Murri, King, and Rostenberg (2006) assert that the use of an RTI framework for the identification of students with emotional disturbances (ED) will make minority students less susceptible to being classified, as this approach would not rely on an interpretation of subjective criteria for ED found in IDEIA. However, these authors discuss that this goal will only be accomplished if culturally-sensitive assessments and clear operational definitions of what qualifies as an adequate “response to intervention” are used within the process. Using a reduction in ODRs as a way to define “response to intervention” may make minority students more vulnerable to special education classification. These students are already over-represented in special education classrooms and recent data suggests that this trend continues to be a problem (Zhang, Katsiyannis, Ju, & Roberts, 2014). Given these historic trends, the federal government has created safeguards to protect the rights of these students. For example, the IDEA re-authorization in 2004 included provisions that states must ensure eligibility practices prevent over-identification and disproportionality of minority students in special education programs. Additionally, states must examine data on disciplinary practices and classroom placements of minority students (U.S. Department of Education, 2007). The use of ODRs as a tool for identifying students in need of intervention and as a measure of behavioral change may place minority students at-risk for special education placement and violate IDEIA provisions that practices should prevent disproportionality. Therefore, the data should be used with caution with this population.

Overall, the results from this study suggested ODRs do not pass Messick's (1995) notion of social consequential validity and have poor concurrent validity with certain behavioral domains. Therefore, ODRs should be not solely be used for making decisions in SWPBIS programs for students, particularly those from diverse backgrounds.

### **Functional Behavior Assessments (FBAs)**

One partial solution for the over-representation of diverse students may involve the use of FBAs. As noted earlier, a small number of students in this sample received a large number of ODRs across the two years. These students also tended to be ethnically diverse. Schools making intervention decisions about these "high flyer" students might consider using additional assessment procedures. For example, the use of individualized FBAs has been advocated for in a pre-referral process for minority students to reduce disproportionality in special education. Lo and Cartledge (2006) utilized FBAs to develop behavioral intervention plans (BIPs) as prevention measures for African-American students who were at-risk or had a behavior and/or emotional disability. In their study, BIPs were designed based on the results of FBAs, which suggested that the identified function maintaining the problem behavior was attention seeking behavior. The effectiveness of the BIPs was measured with a multiple baseline design that compared target students to comparison students on observations of off-task and attention-seeking behavior. Results showed that through the use of BIPs, students demonstrated a rate of off-task behavior comparable to their peers, representing an improvement over baseline levels. These students also showed an increase in appropriate attention seeking behavior and a decrease in inappropriate attention seeking

behavior. The researchers concluded that the use of individualized FBAs with minority children may prevent these students from participating in ineffective interventions that use invalid measures of progress which could lead to special education placement. The use of direct observational data gathered as part of an FBA/BIP with minority children, in addition to ODRs, may allow these students to demonstrate adequate progress with interventions. Therefore, school personnel should consider the use of data generated from individualized FBAs and BIPs when intervening with ethnically diverse students, rather than relying exclusively on ODRs.

In the first year of the CCE intervention, Stage et al. (2012) showed that 3 out of 6 students who were not able to reliably maintain 75% of their DPR points and also did not show an increase in their DPR percentage after individualized social skills instruction (i.e., Basic Plus), did benefit from a BIP based on an FBA. This resulted in the students scoring in the normal range on externalizing problem behaviors. In extending FBA interventions to ODRs, the SWIS manual (May et al., 2002) describes how to analyze ODR data. Obviously, by localizing students who are “high flyers”, personal characteristics such as ethnicity and race would be evident. It is important to note that the intention of this type of ODR analysis would not be to profile diverse students but to recognize individual “high flyers” and progress with an individualized FBA sensitive to the settings and context of where behavior that resulted in ODRs occurred. Lo and Cartledge’s (2006) study shows that individualizing assessment of behaviors with FBA also allows interventionists to determine potential context variables of any sort, such as ethnic or racial diversity, that should be considered. A

cautionary note to these suggestions in the context of the “high flyer” students in this study is that waiting until a student has 68 ODRs or 85 ODRs in one year is too late.

In sum, schools using SWPBIS systems to make decisions about student intervention should use alternative methods rather than relying solely on ODRs, particularly when intervening with students who are ethnically diverse or unresponsive to other levels of treatment. Data collected from an FBA process might be particularly helpful with these populations, as research has shown these students have demonstrated success in response to FBA based interventions.

### **Poisson Hierarchical Nonlinear Modeling**

This study makes a unique contribution to the field with its use of *Poisson* based hierarchical nonlinear modeling analyses to statistically evaluate the validity of ODRs as a data based decision making tool. A review of the literature did not find any other study which used the *Poisson* distribution to analyze ODRs. Previous validity studies have employed nomological networks (Coleman, 2008) and Messick’s unified approach (Irvin et al., 2004, 2006) in attempts to establish validity. However, the nomological network approach used by Coleman (2008) failed to take into account the skewed nature of ODR data and therefore relied on traditional regression techniques that are not appropriate. Irving and colleagues’ (2004; 2006) attempts to use Messick’s unified approach did not apply a statistical technique that would numerically link research based variables to the actual number of ODR events, such as a statistical test of whether there were social consequential validity issues with over-representation. Instead they relied on school personnel’s opinions which presents a mono-

method bias (i.e., just because someone says it has social validity does not mean it's so). Therefore, both approaches had weakness that limited their internal validity to establish either concurrent or unified validity, despite the authors' claims. The *Poisson* distribution addresses both of these weaknesses, as models with this distribution are able to statistically link event counts of ODRs to predictor variables without the need for normally distributed data.

### **Limitations**

As with any research, the present study has some limitations. One possible limitation is the quality of the ODRs. Although teachers were trained using the SWIS manual, it is difficult to ascertain the quality of ODRs utilized, as no information about integrity of the data was collected as part of the study. Therefore, the ODRs may represent varying levels of accuracy, depending on the teacher's tolerance of certain behaviors and the teacher's willingness to complete the SWIS ODR form. It is assumed that the ODRs gathered in the SWIS system met true criteria as a major ODR. Additionally, it should be mentioned that none of the other studies cited in this literature have established the reliability of the major ODRs assigned.

Also, although the present study established a relationship between the rate of ODRs and ethnicity and race, ethnicity and race data was missing for 27% of the sample. For analyses involving ethnicity and race, participants with missing ethnicity or race data were grouped with the "other" ethnicities or races (e.g., Caucasian, Asian). It is possible that some of these participants were African-American or Hispanic. Having more participants in the

ethnically and racially diverse group may have altered the relationship between ethnicity and race and ODRs.

Additionally, this study was unable to ascertain the precise timing of ODRs and how this may relate to intervention level. Specifically, only total counts of ODRs by year were available for analysis. Therefore, it is impossible to determine if a student's rate of ODR varied as the intervention level changed.

### **Future Research Directions**

Although the SWIS form collects specific information about the ODR such as the location of the incidence and the person completing the form, the present study only analyzed the frequency of ODR data obtained by a student. While the present study had teacher demographic data (e.g., education level, ethnicity), the relationship between these variables and ODRs cannot be examined because it is unknown if the classroom teacher wrote the ODR or if it was given by another staff member. Previous research has examined the relationship between teacher ethnicity and ODRs (Bradshaw et al., 2010). However, this research could be extended to use *Poisson* based analysis to examine more teacher characteristics such as education level and licensure, as well as classroom characteristics such as number of special education students. If differential relationships exist, it might suggest that the consequential validity of assignment of ODRs is similar to other findings of over-representation of diverse populations to excessive disciplinary practices (Bradshaw et al., 2010; Coleman, 2008; Vincent et al., 2011; Pas et al., 2011).

Research in the area of SWPBIS relies on ODR data. The previously reviewed literature included more than 20 studies that examined ODRs in relation to SWPBIS. As SWPBIS practices continue to grow, more research is needed to validate other data sources, like FBAs, within the context of the framework. Once a greater emphasis is placed in research on tools other than ODRs, schools may begin to shift data-based decision making away from a reliance on ODRs to other data sources. The fact that more than 8,000 schools currently use SWIS (SWIS, 2013) to track ODRs shows that a large number of schools implementing SWPBIS rely on the data for decision making. Therefore, these schools need more guidance and training from researchers on alternatives to ODRs given the poor validity of the data.



## REFERENCES

- Achenbach, T. M., & Rescorla, L. A. (2001). Manual for the ASEBA school-age forms and profile. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Bradshaw, C. P., Koth, C.W., Thornton, L. A., & Leaf, P. J. (2009). Altering school climate through school-wide positive behavioral interventions and supports: Findings from a group-randomized effectiveness trial. *Prevention Science, 10*, 100-115.
- Bradshaw, C. P., Mitchell, M. M., & Leaf, P. J. (2010). Examining the effects of schoolwide positive behavioral interventions and supports on student outcomes: Results from a randomized controlled effectiveness trial in elementary schools. *Journal of Positive Behavior Interventions, 12*, 133-148.
- Bradshaw, C. P., Mitchell, M. M., O'Brennan, L. M., & Leaf, P. J. (2010). Multilevel exploration of factors contributing to the overrepresentation of black students in office disciplinary referrals. *Journal of Educational Psychology, 102*, 508-520.
- Bradshaw, C. P., Reinke, W. M., Brown, W. M., Bevans, K. B., & Leaf, P. J. (2008). Implementation of schoolwide positive behavioral interventions and supports in elementary schools: Observations from a randomized trial. *Education and Treatment, 31*, 1-26.

- Burke, M. D., Davis, J. L., Lee, Y. H., Hagan-Burke, S., Kwok, O., & Sugai, G. (2012). Universal screening for behavioral risk in elementary school using SWPBS expectations. *Journal of Emotional and Behavioral Disorders*, 38-54.
- Cheney, D. A., Stage, S. A., Hawken, L. A., Lynass, L. S., Mielenz, C. M., & Waugh, M. (2009). A 2-year outcome study of check, connect, and expect intervention for students at risk of severe behavior problems. *Journal of Emotional and Behavioral Disorders*, 20, 1-17.
- Coleman, M. E. (2008). *Investigating the construct validity of office disciplinary referrals as a measure of school-wide positive behavior support* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Crone, D. A., Horner, R. H., & Hawken, L. S. (2004). *Responding to problem behavior in schools: The Behavior Education Program*. New York: Guilford Press.
- Eber, L., Sugai, G., Smith, C. R., & Scott, T. (2002). Wraparound and positive behavioral interventions and supports in schools. *Journal of Emotional and Behavioral Disorders*, 10, 171-180.
- Ebseutani, C., Bernstein, A., Nakamura, B. J., Chorpita, B. F., Higa-McMillian, & C. K., Weisz, J.R. (2010). Concurrent validity of the Child Behavior Checklist DSM-

- oriented scales: Correspondence with DSM diagnoses and comparison to syndrome scales. *Journal of Psychopathology and Behavioral Assessment*, 32, 373-384.
- Ervin, R. A., Schaughency, E., Matthews, A., Goodman, S. D., & McGlinchey, M. T. (2007). Primary and secondary prevention of behavior difficulties: Developing a data-informed problem-solving model to guide decision making at a school-wide level. *Psychology in the Schools*, 44, 7-18.
- Fairbanks, S., Sugai, G., Guardino, D., & Lathrop, M. (2007). Response to Intervention: Examining classroom behavior support in second grade. *Exceptional Children*, 73, 288-310.
- Filter, K. J., McKenna, M. K., Benedict, E. A., Horner, R. H., Todd, A.W., & Watson, J. (2007). Check in/check out: A post-hoc evaluation of an efficient, secondary-level targeted intervention for reducing problem behaviors in schools. *Education and Treatment of Children*, 30, 69-84.
- George, H. (2012). *National PBIS Landscape*. Paper presentation at Texans Care for Children: Austin, TX.
- George, H.P. (2012). *Statewide PBIS leadership conference*: Wisconsin Dells, WI.
- George, H. P., & Barrett, S. (2011). *Advanced PBIS coaching: Building solutions that promote and sustain evidence based practice*. Paper presentation at the 6th Annual

- Summit on Substance Abuse, Mental Health and School Safety/Discipline: Atlanta, GA
- Gresham, F., & Elliot, S. N. (1990). *Social Skills Rating System*. Circle Pines, MN: American Guidance Service.
- Grills-Taquechel, A. E., Fletcher, J. M., Vaughn, S. R., Denton, C. A., & Taylor, P. (2013). Anxiety and inattention as predictors of achievement in early elementary school children. *Anxiety, Stress, and Coping: An International Journal*, 26, 391-410.
- Hachon, T. A., & Allen, R. A. (2013). Identifying students with emotional disturbance: School psychologist practices and perceptions. *Psychology in the Schools*, 193-208.
- Harris-Murri, N., & King, K., Rostenberg, D. (2006). Reducing disproportionate minority representation in special education programs for students with emotional disturbances: Toward a culturally responsive response to intervention model. *Education and Treatment of Children*, 29, 779-799.
- Hawken, L. S., MacLeod, K. S., & Rawlings, L. (2007). Effects of the *Behavior Education Program* on office discipline referrals of elementary school students. *Journal of Positive Behavior Interventions*, 9, 91-101.
- Horner, R. H., Sugai, G., Lewis-Palmer, T., Irvin, L. K., Sugai, G., & Boland, J. (2004). The School-wide Evaluation Tool (SET): A research instrument for assessing school-wide positive behavior supports. *Journal of Positive Behavior Interventions*, 6, 131-147.

- Horner, R. H., Sugai, G., Smolkowski, K., Eber, L., Nakasato, J., Todd, A.W., & Esperanza, J. (2009). A randomized, wait-list controlled effectiveness trial assessing school-wide positive behavior support in elementary schools. *Journal of Positive Behavior Interventions, 11*, 133-144.
- Horner, R. H., Sugai, G., Todd, A. W., & Lewis-Palmer, T. (2002). Schoolwide positive behavior support. In L.M. Bambara & L. Kern (Eds), *Individualized supports for students with behavior problems*, 359-390. New York, NY : The Guilford Press.
- Individuals with Disabilities Education Act. (2004). Pub. L. 108 – 446.
- Irvin, L. K., Horner, R. H., Ingram, K., Todd, A. W., Sugai, G., Sampson, N. K., & Boland, J. B. (2006). Using office discipline referral data for decision making about student behavior in elementary and middle schools: An empirical evaluation of validity. *Journal of Positive Behavior Interventions, 8*, 10-23.
- Irvin, L. K., Tobin, T. J., Sprague, J. R., Sugai, G. & Vincent, C. G. (2004). Validity of office discipline referral measures as indices of school-wide behavioral status and effects of schoolwide behavioral interventions. *Journal of Positive Behavioral Interventions, 6*, 131-144.
- Kamphaus, R., & Reynolds, C. (2008). *BASC-2 Behavioral and Emotional Screening System*. Bloomington, MN: Pearson Assessments.

- Kern, L., & Manz, P. (2004). A look at current validity issues of school-wide behavior support. *Behavioral Disorders, 30*, 47-59.
- Knoff, H. M. (2001). *The Stop and Think Social Skills Program*. Longmont, CO: Sopris West
- Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2009). Teacher Observation of Classroom Adaptation-Checklist (TOCA-C); Development and factor structure. *Measurement and Evaluation in Counseling and Development, 42*, 15-30.
- Lane, K. L. (2007). Identifying and supporting students at risk for emotional and behavioral disorders within multi-level models: Data driven approaches to conducting secondary interventions with an academic emphasis. *Education and Treatment of Children, 30*, 135-164.
- Lane, K. L., Gresham, F. M., & O'Shaughnessy, T. E. (2002). Serving students with or at-risk for emotional and behavior disorders: Future challenges. *Education and Treatment of Children, 25*, 507-521.
- Lane, K. L., Wehby, J. H., Robertson, E. J., & Rogers, L. A. (2007). How do different types of high school students respond to schoolwide positive behavior support programs? Characteristics and responsiveness of teacher-identified students. *Journal of Emotional and Behavioral Disorders, 15*, 3-20.

- Lassen, S. R., Steele, M. M., & Sailor, W. (2006). The relationship of school-wide positive behavior support to academic achievement in an urban middle school. *Psychology in the Schools, 43*, 701-711.
- Lewis, T. (2008). *Teaching social skills: The cornerstone of SW-PBS*. The Office of Special Education Programs' Center for Positive Behavior and Intervention Support: Author.
- Lewinsohn, P. M., Rohde, P., Klein, D. N., & Seeley, J. R., (1999). Natural course of adolescent major depressive disorder: I: Continuity into young adulthood. *Journal of the American Academy of Child and Adolescent Psychiatry, 36*, 56-63.
- Lo, Y., & Cartledge, G. (2006). FBA and BIP: Increasing the behavior adjustment of African-American boys in schools. *Behavioral Disorders, 31*, 147-161.
- Lynass, L. (2010). *An analysis of the reliability of teacher's ratings on daily behavior progress report card in a tier 2 intervention*. Unpublished manuscript.
- May, S., Ard, B., Todd, A., Horner, R., Glasgow, A., Sugai, G., & Sprague, J. (2000). *Becoming Swift at SWIS: Learning to use the school-wide information system*. University of Oregon.
- Messick, S. (1995). Validity of psychological assessment: Validations of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist, 741-749*.

- Merrill, K. M. (2010). *Behavioral, Social, and Emotional Assessment of Children and Adolescents*. Mahwah, New Jersey: Lawrence Erlbaum Association.
- McIntosh, K., Campbell, A. L., Carter, D. R., & Dickey, C. R. (2009a). Differential effects of a tier two behavior intervention based on function of problem behavior. *Journal of Positive Behavior Interventions, 11*, 82-93.
- McIntosh, K., Campbell, A. L., Carter, D. R., & Zumbo, B.D. (2009b). Concurrent validity of office disciplinary referrals and cut points used in schoolwide positive behavior support. *Behavioral Disorders, 34*, 100-113.
- McKinney, J., Montague, M., & Hocutt, A. (1998). A two year follow up study of children at-risk for developing SED: Initial results from a prevention project. In C. Liberton, K. Kutash, & R. Friedman (Eds.), *A system of care for children's mental health: Expanding the research base, tenth annual proceedings* (pp. 271–277). Tampa: University of South Florida, Research and Training Center for Children's Mental Health.
- Nakasto, J., (2000). Data-based decision making in Hawaii's behavior support effort. *Journal of Positive Behavior Interventions, 2*, 247-251.
- Nakamura, B. J., Ebseutani, C., Bernstein, A., & Chorpita, B. E. (2009). A psychometric analysis of the DSM-oriented scales. *Journal of Psychopathology and Behavior Assessment, 31*, 178-189.



- National Association of School Psychologist. (2010). Model for comprehensive and integrated school psychological services. Author.
- Nelson, J. R., Benner, G. J., Reid, R. C., Epstein, M. H., & Currin, D. (2002). The convergent validity of office discipline referrals with the CBCL-TRF. *Journal of Emotional and Behavioral Disorders, 10*, 181-188.
- Nelson, J. R., Hurley, K. D., Synhorst, L., Epstein, M. H., Stage, S., & Buckley, J. (2009). The child outcomes of a behavioral model. *Exceptional Children, 76*, 7-30.
- Nelson, J. R., Martella, R., & Galand, B. (1998). The effects of teaching school expectations and establishing consistent consequences on formal office disciplinary actions. *Journal of Emotional and Behavioral Disorders, 6*, 153-161.
- Osgood, D. W. (2000). Poisson-based regression analysis of aggregate crime rate. *Journal of Quantitative Criminology, 21*, 21-43.
- Pas, E. T., Bradshaw, C., & Mitchell, M. M. (2011). Examining the validity of office disciplinary referrals as an indicator of student behavior problems. *Psychology in the Schools, 48*, 541-555.
- Raudenbush, S., Bryk, A., Cheong, Y.F., & Congdon, R. (2004). *Hierarchical linear and nonlinear modeling*. Lincolnwood, IL: Scientific Software International, Inc.

Reid, R., Gonzalez, J. E., Nordness, P. D., Trout, A., & Epstein, M. H. (2004). A meta-analysis of the academic status of students with emotional/behavioral disturbance.

*The Journal of Special Education, 38*, 130-143.

Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment Scale for Children* (2<sup>nd</sup> ed.). Circle Pines, MN: AGS Publishing

School-wide information systems. (2013). Overview of SWIS. Retrived from

<http://www.swis.org/index.php?page=overview>.

Simonsen, B., Myers, D., & Briere, D. E. (2011). Comparing a behavioral check-in/check-out (CICO) intervention to standard practice in an urban middle school setting using an experimental group design. *Journal of Positive Behavior Interventions, 13*, 31-48.

Sherrod, M. D., Getch, Y. Q., & Ziomek-Daigle, J. (2009). The impact of positive behavior support to decrease discipline referrals with elementary students. *Professional School Counseling, 12*, 421-427.

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

Stage, S. A., Cheney, D., Mielenz, C., Lynass, L., & Flower, A. (2012). Three validity studies of the daily progress report in relationship to the Check, Connect an Expect intervention. *Journal of Positive Behavior Interventions, 14*, 181-191.

- Stage, S. A., Jackson, H. G., Moscovitz, K., Erickson, M. J., Thurman, S. O., & Jessee, W., (2006). Using multi-method, multi-source functional behavioral assessment with students with behavioral disabilities. *School Psychology Review, 35*, 451-471.
- Sugai, G. & Horner, R. H. (2002). The evolution of discipline practices: School-wide positive behavior supports. *Behavioral Psychology in the Schools, 24*, 23-50.
- Sugai, G. & Horner, R. H. (2006). A promising approach for expanding and sustaining school-wide positive behavior support. *School Psychology Review, 35*, 245-259.
- Sugai, G. & Horner, R. H. (2009). Responsiveness-to-intervention and school-wide positive behavior supports: Integration of multi-tiered systems approaches. *Exceptionality, 17*, 222-237.
- Sugai, G., Sprague, J. R., Horner, R. H., & Walker, H. M. (2000). Preventing school violence: The use of office discipline referrals to assess and monitor school-wide discipline interventions. *Journal of Emotional and Behavioral Disorders, 8*, 94-101.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics , 5th ed.* Boston : Allyn and Bacon.
- Taylor-Greene, S., Brown, D., Nelson, L., Longton, J., Gassman, T., Cohen, J., Swartz, J., Horner, R. H., Sugai, G., Hall, S. (1997). School-wide behavioral support: Starting the year off right. *Journal of Behavioral Education, 7*, 99-112.

- Taylor-Greene, S. J., & Kartub, D. T. (2000). Durable implementation of school-wide positive behavior support: The High Five Program. *Journal of Positive Behavior Interventions, 2*, 233-235.
- Todd, A. W., Campbell, A. L., Meyer, G. G., & Horner, R. H. (2008). The effects of a targeted intervention to reduce problem behaviors. *Journal of Positive Behavior Interventions, 10*, 46-55.
- U.S. Department of Education. (2007). *Disproportionality and over identification*. Author.
- Verboom, C. E., Sijtsema, J. J., Verhulst, F. C., Penninx, B., & Ormel, J. (2014). Longitudinal associations between depressive problems, academic performance, and social functioning in adolescent boys and girls. *Developmental Psychology, 50*, 247-257.
- Vincent, C. G., Swain-Bradway, J., Tobin, T. J., & May, S. (2011). Disciplinary referrals for culturally and linguistically diverse students with and without disabilities: Patterns resulting from school-wide positive behavior support. *Exceptionality, 19*, 175-190.
- Walker, B., Cheney, D., Stage, S., & Blum, C. (2005). Schoolwide screening and positive behavior supports: Identifying and supporting students at risk for school failure. *Journal of Positive Behavior Interventions, 7*, 194-204.
- Walker, H. M, Horner, R. H., Sugai G., Bullis, M., Sprague, J. R., Bricker, D., & Kaufman, M. J. (1996). Integrated approaches to preventing antisocial behavior patterns among

- school-age children and youth. *Journal of Emotional and Behavioral Disorders*, 4, 194-220.
- Walker, H. M., Kavanagh, K., Stiller, B., Golly, A., Severson, H., & Feil, E. (1998). First Steps to Success: An early intervention approach for preventing school antisocial behavior. *Journal of Emotional and Behavioral Disorders*, 6, 66-80.
- Walker, H. M., & Severson, H. H. (1992). *Systematic Screening for Behavior Disorders: User's guide and technical manual*. Longmont, CO: Sopris West.
- Walker, H. M., Severson, H. H., Todis, B. J., Block-Pedgo, A. E., Williams, G. J., Haring, N. G., & Barckley, M. (1990). Systematic Screening for Behavior Disorders (SSBD): Further validation, replication, and normative data. *Remedial and Special Education*, 11, 32-46.
- Woodcock, R.W., McGrew, K.S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. IL: Riverside Publishing.
- Zhang, D., Katsiyannis, A., Ju, S., & Roberts, E. (2014). Minority representation in special education: 5 year trends. *Journal of Child and Family Studies*, 23, 118-127.

Table 1

*Frequency and Percentage for Treatment Level across the Two Years*

Student status at end of Year 1	Student status at end of Year 2									
	Basic Plus		Basic		Self-Monitoring		Graduate		Total	
	N	%	N	%	N	%	N	%	N	%
Intensive	4	22%	0	0%	2	8%	3	6%	9	9%
Basic Plus	9	50%	2	33%	7	29%	7	13%	25	25%
Basic	3	17%	3	50%	10	41%	17	32%	33	32%
Self-Monitoring	2	11%	1	17%	5	20%	20	37%	28	28%
Graduate	0	0%	0	0%	0	0%	7	13%	7	7%
Total	18	18%	6	6%	24	23%	54	53%	102	

Note. Treatment level data available for 102 participants of the 121 that were in the CCE intervention.

Table 2

*Demographic Variable Percentages for Participants At-risk by Predictor Variables*

	At-risk Externalizing		At-risk Internalizing		At-risk Social Skills		At-risk Letter-Word ID		At-risk Math Calculation		At-risk Ethnicity	
	N	%	N	%	N	%	N	%	N	%	N	%
<i>Gender</i>												
Male	84	41%	59	29%	41	20%	26	13%	16	8%	34	17%
Female	25	12%	22	11%	18	9%	7	4%	11	5%	13	6%
<i>Ethnicity</i>												
Caucasian	38	19% <sup>a</sup>	41	20%	22	11%	14	7%	10	5%		
African-American	18	9%	9	4%	9	5%	6	3%	6	3%	30	15%
Hispanic	11	5%	8	4%	5	2%	2	1%	3	2%	17	8%
Asian	3	1% <sup>a</sup>	7	3%	1	.5%	1	.5%	2	1%		
American Indian	1	.5%	1	.5%	0	0%	0	0%	0	0%		
Unknown	38	19% <sup>b</sup>	15	7% <sup>a</sup>	22	11% <sup>b</sup>	10	5%	6	3%		

Note. Percentages are out of the total *n* (both at-risk and not at-risk) for predictor variable. On the CBCL Externalizing, CBCL Internalizing, and WJIII Math Calculation data were available for 203 participants. On WJIII Letter-Word Identification, data were available for 204 participants. For SSRS, data were available for 201 participants. <sup>a</sup> denotes percentages lower than expected <sup>b</sup> percentages higher than expected

Table 2, continued

	At-risk Externalizing		At-risk Internalizing		At-risk Social Skills		At-risk Letter-Word ID		At-risk Math Calculation		At-risk Ethnicity	
	N	%	N	%	N	%	N	%	N	%	N	%
<i>Lunch Status</i>												
Regular	37	18%	31	15%	27	14% <sup>b</sup>	12	6%	10	5%	17	8%
Reduced	12	6%	9	4%	5	2%	1	.5%	3	2%	6	3%
Free	45	22%	31	15%	15	7% <sup>a</sup>	12	6%	9	4%	23	11%
Unknown	15	7%	10	5%	12	6%	8	4%	5	3%	1	.5%
<i>Graduate Status</i>												
Graduate	21	10% <sup>a</sup>	18	9% <sup>a</sup>	9	5% <sup>a</sup>	8	4%	3	2% <sup>a</sup>	10	5% <sup>a</sup>
Non-Graduate	88	43%	63	31%	50	25%	25	12%	24	12%	37	18%
<i>IEP Status</i>												
Yes	44	22%	43	21% <sup>b</sup>	32	16% <sup>b</sup>	26	13% <sup>b</sup>	20	10% <sup>b</sup>	20	10%
No	64	32%	37	18%	27	14%	7	3%	7	3%	27	13%
Missing	2	.5%	1	.5%	0	0%	0	0%	0	0%	0	0%

Note. Percentages are out of the total *n* (both at-risk and not at-risk) for predictor variable. On the CBCL Externalizing, CBCL Internalizing, and WJIII Math Calculation data were available for 203 participants. On WJIII Letter-Word Identification, data were available for 204 participants. For SSRS, data were available for 201 participants. <sup>a</sup> denotes significant Fisher's Exact Test with percentage lower than expected <sup>b</sup> denotes significant Fisher's Exact Test with percentage higher than expected



Table 2, continued

	At-risk Externalizing		At-risk Internalizing		At-risk Social Skills		At-risk Letter-Word ID		At-risk Math Calculation		At-risk Ethnicity	
	N	%	N	%	N	%	N	%	N	%	N	%
<i>Primary Disability Category</i>												
Emotional Disturbance	8	4% <sup>b</sup>	5	2%	3	1%	3	1%	2	.5%	2	.5%
Other Health Impaired	12	6%	15	7%	9	4%	7	3%	5	2%	6	3%
Specific Learning Disabled	11	5%	8	4%	13	6%	10	5%	6	2%	7	3%
Speech-Language Impaired	4	2% <sup>a</sup>	7	3%	4	2%	1	.5%	3	1%	2	.5%
Developmental Delayed	6	2% <sup>b</sup>	3	1%	1	.5%	3	1%	2	.5%	3	1%
Hearing Impaired	0	0%	1	.5%	0	0%	1	.5%	1	.5%	0	0%
Intellectual Disability	2	.5%	1	.5%	1	.5%	1	.5%	1	.5%	0	0%
Autism	0	0%	2	.5%	0	0%	0	0%	0	0%		
Missing	2	.5%	2	.5%	1	.5%	0	0%	0	0%	0	0%

Note. Percentages are out of the total  $n$  (both at-risk and not at-risk) for predictor variable. On the CBCL Externalizing, CBCL Internalizing, and WJIII Math Calculation data were available for 203 participants. On WJIII Letter-Word Identification, data were available for 204 participants. For SSRS, data were available for 201 participants. <sup>a</sup> denotes percentages lower than expected <sup>b</sup> percentages higher than expected

Table 3

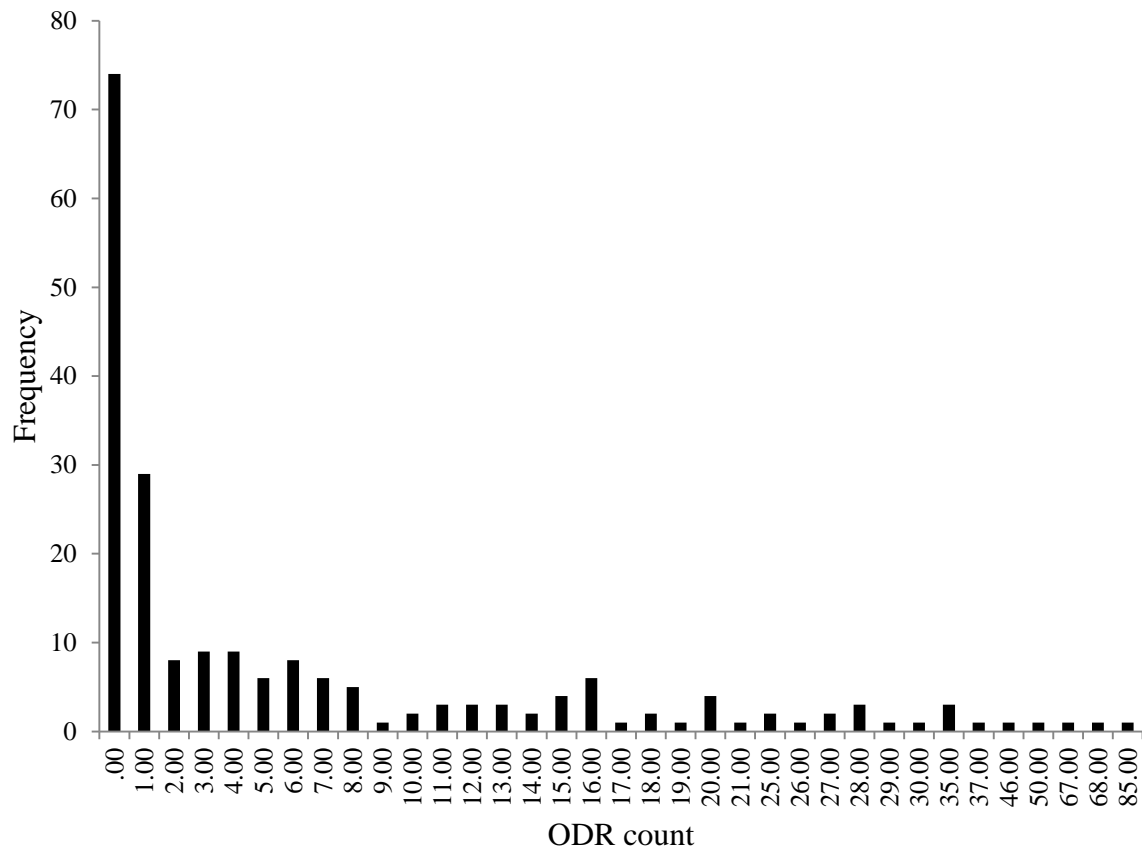
*Main Effects Hierarchical Nonlinear Modeling Estimates Predicting ODR rate*

<i>Fixed Effects</i>						
	Coefficient	SE	T-ratio	<i>df</i>	<i>p</i> -value	Event Rate Ratio [95% CI]
Intercept	0.459325	0.282924	1.623	2	0.238	1.583005 [0.847, 2.957]
Ethnicity Intercept	0.545443	0.056464	9.660	190	0.000	1.725372 [1.544, 1.928]
Social Skills Intercept	-0.106856	0.060517	-1.766	190	0.079	0.898655 [0.798, 1.012]
Calculation Intercept	0.131392	0.085846	1.531	190	0.127	1.140415 [0.963, 1.351]
Letter-Word Identification Intercept	-0.036778	0.080822	-0.455	190	0.649	0.963890 [0.822, 1.130]
Externalizing Intercept	1.931862	0.094320	20.482	190	0.000	6.902349 [5.732, 8.312]
Internalizing Intercept	-0.153666	0.057282	-2.682	190	0.008	0.857558 [0.766,0.960]
<i>Random Effect</i>						
	Standard Deviation	Variance Component	<i>df</i>	$\chi^2$	<i>p</i> value	
	0.46316	.21452	2	223.23411	0.000	

Table 4

*Hierarchical Nonlinear Modeling Estimates Predicting ODR rate with Significant Main Effects Nested within Graduate Status*

<i>Fixed Effects</i>						
	Coefficient	SE	T-ratio	df	p-value	Event Rate Ratio [95% CI]
Intercept	0.331521	0.263695	1.257	2	0.272	1.393085 [0.778, 2.494]
Ethnicity Intercept	0.421402	0.063770	6.608	196	0.000	1.524097 [1.344, 1.728]
Grad Slope	0.731115	0.137441	5.319	196	0.000	2.077396 [1.585, 2.723]
Externalizing Intercept	1.987406	0.138455	14.355	196	0.00	7.296582 [5.556, 9.583]
Grad Slope	-0.067355	0.180256	-0.374	196	0.709	0.934863 [0.656, 1.333]
Internalizing Intercept	-0.102014	0.062835	-1.624	196	0.106	0.903017 [0.798, 1.022]
Grad Slope	0.145278	0.153806	0.945	196	0.364	1.156361 [0.854, 1.565]
<i>Random Effect</i>						
	Standard Deviation	Variance Component	df	$\chi^2$	p value	
	0.42115	.17737	2	189.59554	0.000	



*Figure 1. The distribution of ODRs within the sample*

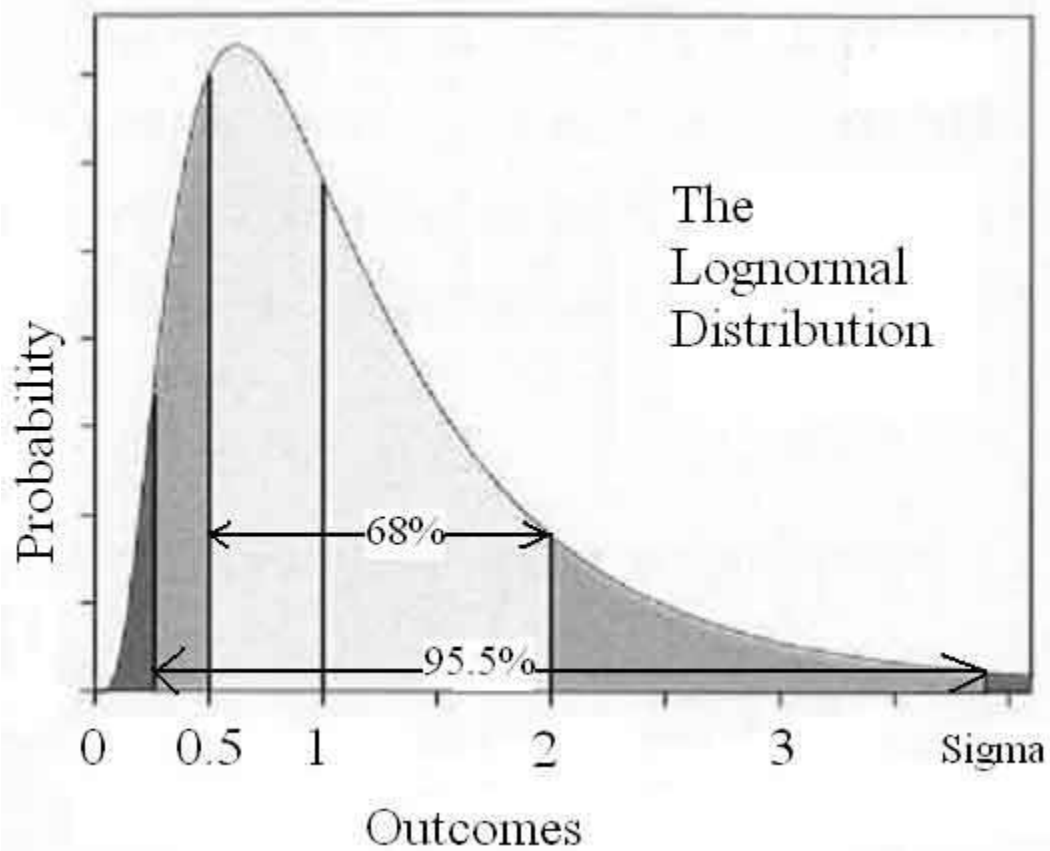


Figure 2. The lognormal distribution.

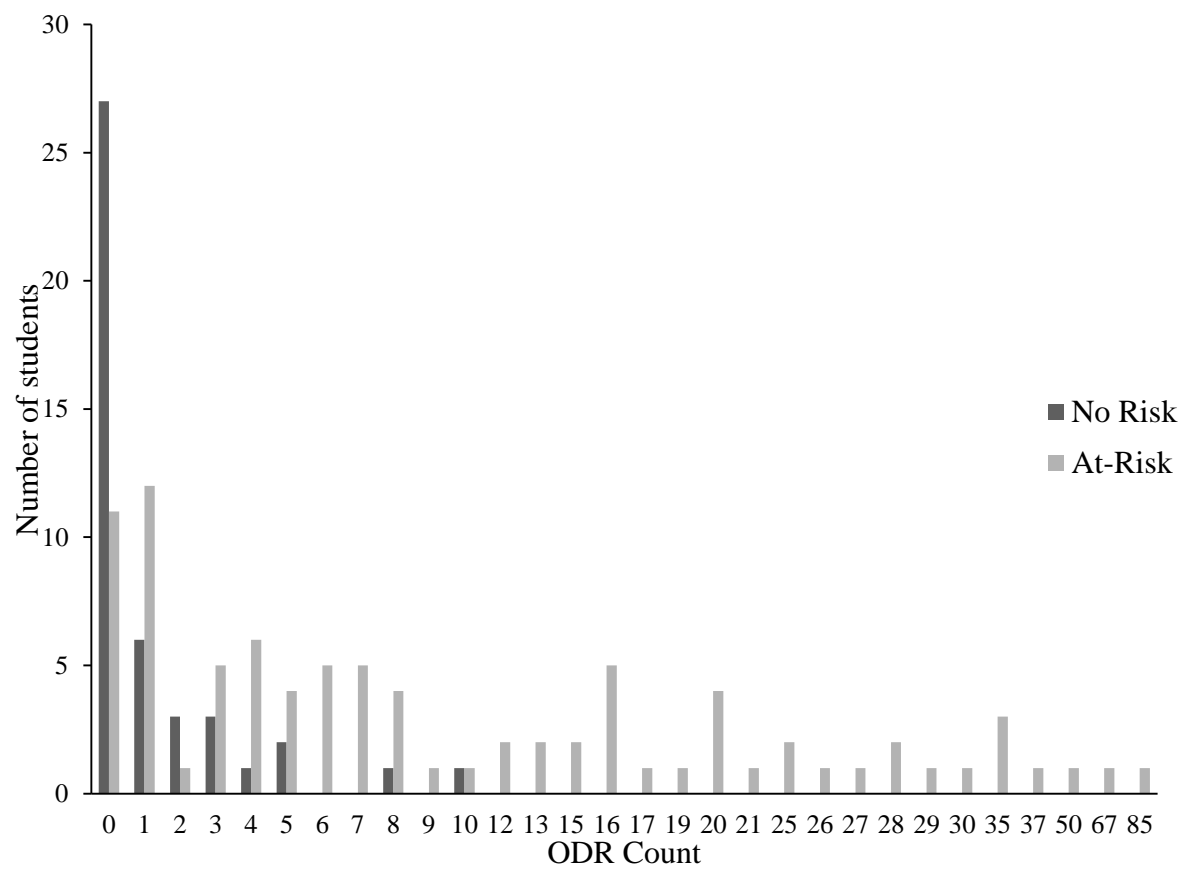


Figure 3. ODR counts for non-graduates on externalizing problem behavior

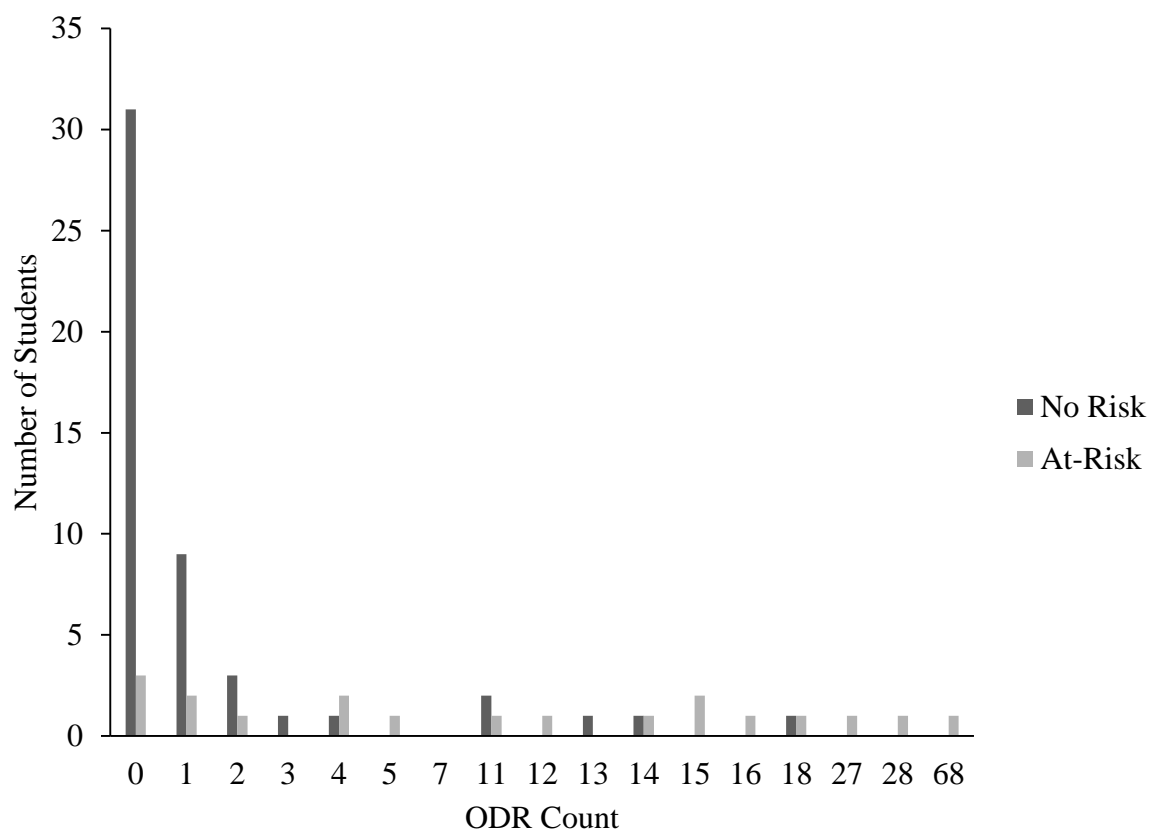


Figure 4. ODR counts for graduates on externalizing problem behavior

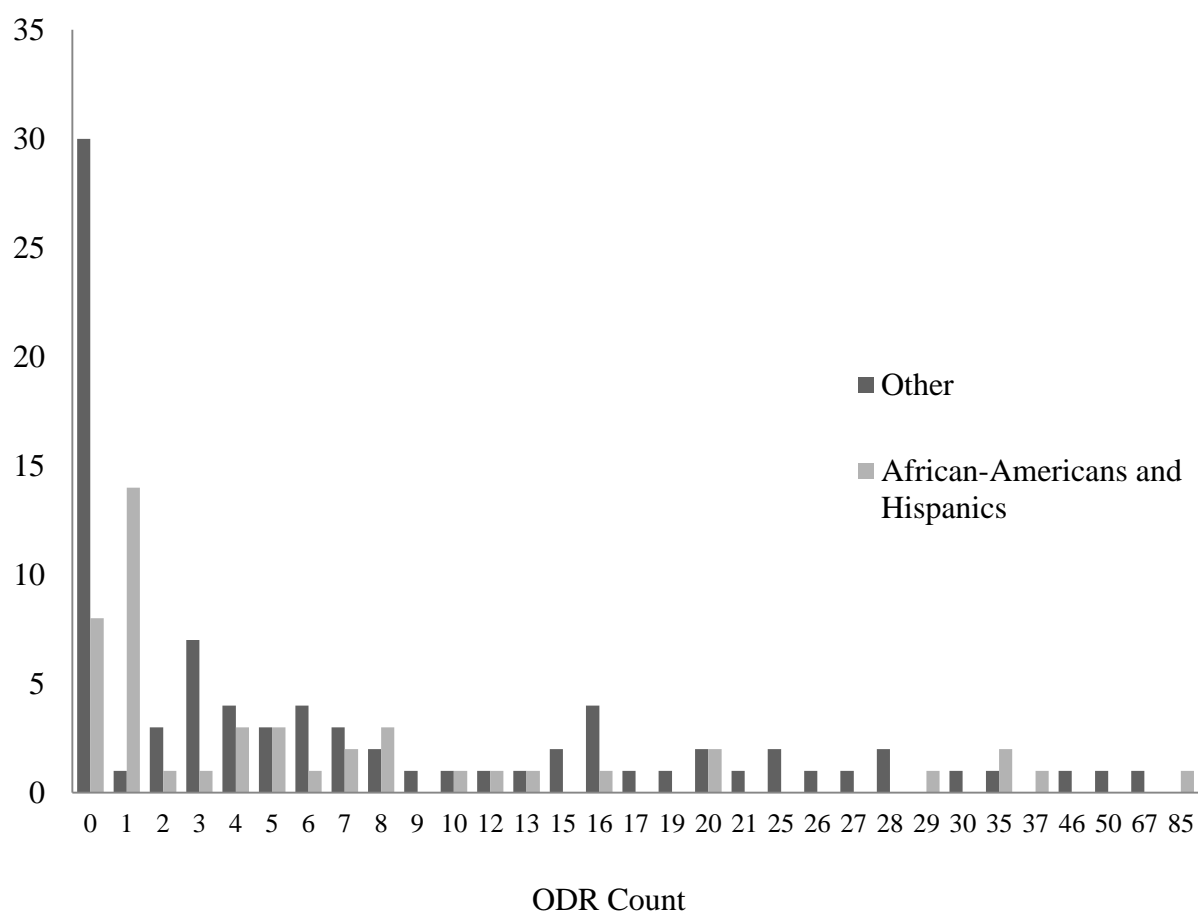


Figure 5. ODR counts by ethnicity for non-graduates



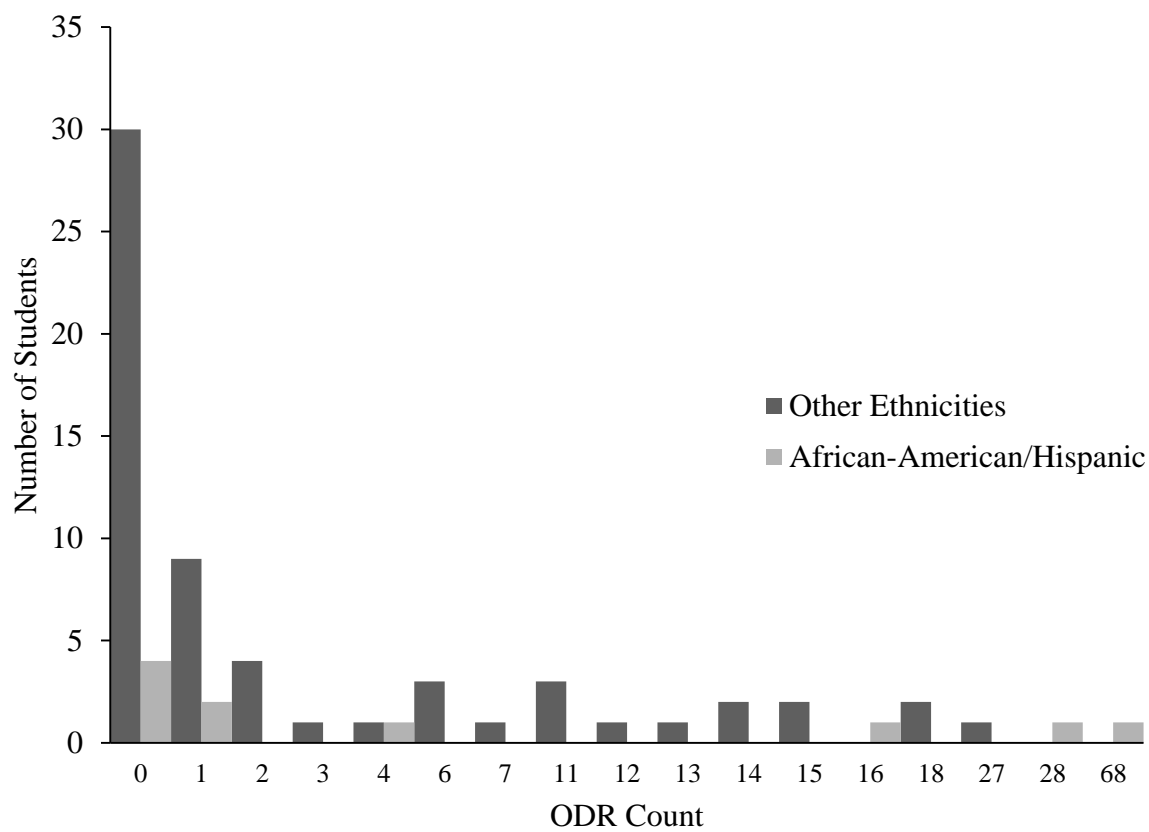


Figure 6. ODR counts by ethnicity for graduates

**APPENDICES**

## Appendix A

SWIS OFFICE DISCIPLINE REFERRAL FORM						
Student(s) _____		Referring Staff _____		Grade Level _____	Date _____	Time _____
<b>Location</b>						
<input type="checkbox"/> Classroom	<input type="checkbox"/> Cafeteria	<input type="checkbox"/> Bus loading zone		<input type="checkbox"/> Other _____		
<input type="checkbox"/> Playground	<input type="checkbox"/> Bathroom/restroom	<input type="checkbox"/> Parking lot				
<input type="checkbox"/> Commons/common area	<input type="checkbox"/> Gym	<input type="checkbox"/> On bus				
<input type="checkbox"/> Halfway/ breezeway	<input type="checkbox"/> Library	<input type="checkbox"/> Special event/assembly/ field trip				
<b>Problem Behaviors (check the most intrusive)</b>						
<input type="checkbox"/> Minor/ warning	<input type="checkbox"/> Tardy	<input type="checkbox"/> Tobacco	<input type="checkbox"/> Bomb threat			
<input type="checkbox"/> Abusive lang./ inapprop. lang	<input type="checkbox"/> Skip class/ truancy	<input type="checkbox"/> Alcohol/drugs	<input type="checkbox"/> Arson			
<input type="checkbox"/> Fighting/ physical aggression	<input type="checkbox"/> Forgery/ theft	<input type="checkbox"/> Combustibles	<input type="checkbox"/> Weapons			
<input type="checkbox"/> Defiance/disrespect/insubordination/non-compliant	<input type="checkbox"/> Dress code violation	<input type="checkbox"/> Vandalism	<input type="checkbox"/> Other _____			
<input type="checkbox"/> Harassment/ tease/ taunt	<input type="checkbox"/> Lying/cheating	<input type="checkbox"/> Property damage				
<input type="checkbox"/> Disruption						
<b>Possible Motivation</b>						
<input type="checkbox"/> Obtain peer attention	<input type="checkbox"/> Avoid tasks/activities	<input type="checkbox"/> Don't know				
<input type="checkbox"/> Obtain adult attention	<input type="checkbox"/> Avoid peer(s)	<input type="checkbox"/> Other _____				
<input type="checkbox"/> Obtain items/ activities	<input type="checkbox"/> Avoid adult(s)					
<b>Others involved</b>						
<input type="checkbox"/> None	<input type="checkbox"/> Peers	<input type="checkbox"/> Staff	<input type="checkbox"/> Teacher	<input type="checkbox"/> Substitute	<input type="checkbox"/> Unknown	<input type="checkbox"/> Other _____
<b>Administrative Decision</b>						
<input type="checkbox"/> Time in office	<input type="checkbox"/> Detention	<input type="checkbox"/> Saturday School	<input type="checkbox"/> In-school suspension			
<input type="checkbox"/> Loss of privilege	<input type="checkbox"/> Parent contact	<input type="checkbox"/> Individualized instruction	<input type="checkbox"/> Out-of-school suspension			
<input type="checkbox"/> Conference with student	<input type="checkbox"/> Other _____					
<b>Comments:</b>						
<b>Follow up comments:</b>						

Appendix B

**THE PAW PRINT**  
TIFFANY PARK ELEMENTARY  
RENTON, WASHINGTON

Checked in	YES	NO
	TA	AB
Checked out	YES	NO
Parent Signature	YES	NO

STUDENT: \_\_\_\_\_

DATE: \_\_\_\_\_

TEACHER: \_\_\_\_\_

GOAL: \_\_\_\_\_

**READING**

**WRITING**

Expectations	Tough Time	OK	Good	Purrfect	Expectations	Tough Time	OK	Good	Purrfect	SUB-TOTAL
Participate Respectfully	1	2	3	4	Participate Respectfully	1	2	3	4	
Act Responsibly	1	2	3	4	Act Responsibly	1	2	3	4	
Will Follow Directions	1	2	3	4	Will Follow Directions	1	2	3	4	
Will Be Safe	1	2	3	4	Will Be Safe	1	2	3	4	

Total \_\_\_\_\_

Total \_\_\_\_\_

**MATH**

**SCIENCE/SOCIAL STUDIES**

Expectations	Tough Time	OK	Good	Purrfect	Expectations	Tough Time	OK	Good	Purrfect
Participate Respectfully	1	2	3	4	Participate Respectfully	1	2	3	4
Act Responsibly	1	2	3	4	Act Responsibly	1	2	3	4
Will Follow Directions	1	2	3	4	Will Follow Directions	1	2	3	4
Will Be Safe	1	2	3	4	Will Be Safe	1	2	3	4

Total \_\_\_\_\_

Total \_\_\_\_\_

Parent Signature: _____ Comments:	<b>TOTAL:</b> _____ Comments:
--------------------------------------	----------------------------------

**PURRFECT (4):** Met expectations with positive behavior; worked independently without corrections or reminders.

**GOOD (3):** Met expectations with only 1 correction or reminder.

**OK (2):** Needed 2 or 3 corrections or reminders.

**TOUGH TIME (1):** Needed 4 or more corrections or reminders

