PREDICTING STUDENT OUTCOMES USING OFFICE DISCIPLINE REFERRAL DATA FROM A NATIONAL SAMPLE OF MIDDLE SCHOOL STUDENTS

by

Larissa Kristine Predy

B.A. (Honours), Vancouver Island University, 2008

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in

The Faculty of Graduate Studies

(School Psychology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2012

© Larissa Kristine Predy, 2012
Abstract

This study examined the adequacy of using Office Discipline Referral (ODR) data to predict student behaviour outcomes. Specifically, the study identified rates of ODRs and student trajectories in a middle school population and assessed whether end-of-year behaviour outcomes can reliably be predicted as early as the first few months of Grades 6, 7, and 8. Participants were 401,852 students from 593 public middle schools (serving Grades 6 to 8) in the United States whose ODR data had been entered in the School-wide Information System (SWIS, 2012) during the 2009-2010 school year. ODRs were categorized by final ODR cut points used in positive behaviour support systems (0-1, 2-5, and 6+ total ODRs per year). Descriptive analyses and multilevel multinomial logistic regression analyses were performed; Receiver Operating Characteristic (ROC) analyses were used to assess diagnostic accuracy. Results showed relatively stable mean increases in ODRs throughout the school year for students in each final ODR category, though median growth trajectories revealed a comparatively large increase in ODRs received in November and February. Results also showed that ODRs received in September, October, and November were statistically significant predictors of final behaviour outcomes, and the inclusion of types of referrals (especially for defiance) significantly improved prediction of the final ODR category. These results are discussed with regards to previous and future research, limitations, and the utility of ODRs for intervention decision-making in middle schools.
Preface

This thesis consists of original research conceived by the graduate student, with advisement from her research supervisor. The research used archival records from the *School-wide Information System* (SWIS, 2011), and access was granted to the graduate student by ECS at the University of Oregon. The graduate student was responsible for analysis and writing, and thus, this thesis represents her work as lead researcher and author. Ethics Approval was required by the UBC Behavioural Research Ethics Board (BREB) to conduct this research. The UBC BREB certificate number is H11-01800.
# Table of Contents

ABSTRACT ............................................................................................................................... ii
PREFACE ................................................................................................................................. iii
TABLE OF CONTENTS .............................................................................................................. iv
LIST OF TABLES ..................................................................................................................... v
LIST OF FIGURES .................................................................................................................. vi
ACKNOWLEDGEMENTS .......................................................................................................... vii

CHAPTER 1: INTRODUCTION ............................................................................................. 1
   Patterns of Problem Behaviour in Schools ......................................................................... 3
   From Reactive to Proactive ................................................................................................. 4
   Measuring Behaviours with Office Discipline Referrals ................................................... 5
   The Current Study .............................................................................................................. 8

CHAPTER 2: METHOD .......................................................................................................... 10
   Participants and Settings .................................................................................................... 10
   Measure .............................................................................................................................. 12
   Design and Analysis .......................................................................................................... 12

CHAPTER 3: RESULTS ......................................................................................................... 16
   Descriptive Analyses ........................................................................................................ 16
   Predictive Analyses .......................................................................................................... 21

CHAPTER 4: DISCUSSION .................................................................................................... 28
   ODR Frequencies and Trajectories .................................................................................. 28
   Prediction with Cumulative Number of ODRs Received Per Month ............................. 30
   Prediction with Type of ODR (e.g., defiance) ................................................................... 31
   Limitations and Future Research .................................................................................... 33
   Implications for Practice .................................................................................................. 35

REFERENCES ......................................................................................................................... 39
List of Tables

Table 2.1  Proportion of ODRs by Type of Problem Behaviour (N = 401,852)..........................11
Table 3.1  Mean Growth Trajectories by ODR Cut Point Group ........................................16
Table 3.2  Multinomial Logistic Regression and Accuracy of Prediction of Final ODR Category
.............................................................................................................................................22
Table 3.3  Accuracy of Prediction of Final ODR Category ......................................................23
Table 3.4  Prediction of Total ODRs by Type of ODR (by End of September) ......................25
Table 3.5  Accuracy for Prediction of Total ODRs by Referred Behaviour (by End of September)
.............................................................................................................................................27
List of Figures

Figure 3.1 Mean growth in ODRs for students in total ODR categories by month ............... 17
Figure 3.2 Median growth in ODRs for students in total ODR categories by month .......... 18
Figure 3.3 Percent of students in 2-5 final ODR category in June with 2-5 ODRs per month (n = 46,886) ........................................................................................................................................ 19
Figure 3.4 Percent of students in 6+ final ODR category in June with 6+ ODRs per month (n = 19,812) ........................................................................................................................................ 20
Figure 3.5 Percent of students in 6+ final ODR category in June with 2+ ODRs per month (n = 19,812) ........................................................................................................................................ 21
Acknowledgements

To my Research Supervisor, Dr. Kent McIntosh, I owe you many thanks. Your consistent support and commitment have been invaluable, and I sincerely appreciate all your efforts, thoughtful direction, constant encouragement, and guidance. A special thanks to my committee members, Dr. Sterett Mercer and Dr. Joseph Lucyshyn, for your interest in this project as well as your advice, support, and guidance. A special thanks to Jenny Frank, at Pennsylvania State University, who cleaned and prepared the archival dataset in preparation for this study: your time and efforts are greatly appreciated.

Finally, I am beyond grateful to my loving family and friends who have been more than supportive throughout this process. Without your unending confidence and encouragement, I would not have endeavoured to reach so high. Thank you for celebrating successes with me and for listening and supporting me through the setbacks.
Chapter 1: Introduction

“It is easier to build strong children than to repair broken men”

-- Frederick Douglass

For children and youth, research has explicitly demonstrated that schools are the most common setting for violent and antisocial behaviours, including bullying, physical aggression, harassment, and defiance/disrespect (Beaudoin & Taylor, 2004). Although not all students engage in violent behaviours, exposure alone has the capacity to affect student academic and social-emotional outcomes (Mayer & Furlong, 2010). Witnessing and growing up surrounded by violence has been shown to have a detrimental impact on child and youth development (Guerra, Huesmann, & Spindler, 2003; Swick & Williams, 2006).

Although violent crimes (e.g., school shootings) represent a miniscule proportion of problem behaviours at all grade levels (Spaulding et al., 2010), they have often been equated with school safety, regardless of the prevalence of less severe problem behaviours (Mayer & Furlong, 2010). For example, Spaulding et al. reported that only 4% of office referrals were issued for harassment/bullying and 12% for fighting in a given year. Although the rate of violent crimes in schools has consistently decreased from 1994 to 2007, comparatively, less directly violent behaviours, such as harassment/bullying and defiance, have not decreased over this time period (NCES, 2010).

Given that children spend a large proportion of their time in school from early childhood to late adolescence, the influence of school culture, and the natural availability of social structures designed for support, schools are demonstrably the most effective place for preventive action. Fortunately, teaching and promoting appropriate behaviour in schools is increasingly viewed as just as important as teaching academics. In a large-scale CBC poll of Canadian
families, the vast majority of Canadians supported the idea of schools educating students in moral and social development (93% and 99%, respectively; CBC News, 2008). The priority for both behaviour intervention and early prevention strategies is to support healthy social development and school success as early in a student’s school experience as possible.

In response to the need for more extensive preventive action, school-wide approaches for promoting appropriate behaviour have become more common in schools across North America (e.g., Positive Behaviour Interventions and Supports, PBIS; Sugai & Horner, 2009). Within a three-tiered model of behaviour intervention, such as that used by PBIS, tier 1 reflects universal or school-wide behaviour principles or programs, tier 2 represents small groups of students who receive specific supports with common intervention goals, and tier 3 represents yet fewer students receiving individualized programs for specified areas of concern.

Still, even when implementing preventive support, identification of students at risk of significant behaviour problems in schools, particularly by middle and high school, continues to be a challenge for school teams. A critical activity for school teams implementing these systems is determining which students require which level of intervention by use of effective screening measures. Although tier 1 systems, such as those in PBIS, provide an effective means for reducing the numbers of students requiring additional support, a proportion of students continue to be at risk of developing serious behaviour problems. As such, school teams can benefit from implementing screening systems shown to effectively identify students by which behaviour interventions can be efficiently employed.

Severson et al. (2007) identified a number of instruments that schools might consider for identifying behaviourally at risk students. They discussed the importance of weighing a number of factors for selecting appropriate behavioural screening measures. Effective screeners were
identified as cost efficient, able to accurately identify a high proportion of students requiring support (sensitivity), able to accurately identify students not requiring support (specificity), and capable of identifying students quickly, early, and with fidelity, and can be used by various raters. In their review of multiple screening tools, the most highly rated measures provided information to guide intervention, used teacher ratings on common behavioural factors, and teacher evaluations of social interactions.

When unidentified and unaddressed, conduct problems can lead not only to negative individual and family outcomes, but also significant financial costs to society (Jones, Dodge, Foster, Nix, & Conduct Prevention Research Group, 2002). According to Cohen and Piquero (2009), the monetary value of individually supporting one child at-risk for challenging behaviour from birth is between $2.6 and $4.4 million US dollars but increases to over $5.8 million US if intervention begins after age 14. Cohen and Piquero noted that early identification of children who are at-risk can lead to reduced cost for social services and more promising long term outcomes.

**Patterns of Problem Behaviour in Schools**

Greater action in middle school appears to be vital. Rates of specific behaviour problems and behaviour trajectories identify middle school as a period of particular need for support (NCES, 2010). By identifying students needing support earlier in middle school and providing a consistent level of support, school teams can capitalize on a critical window for intervention.

To provide more effective and efficient behaviour support for students in middle school, school teams can examine what specific problem behaviours are most prevalent, thereby identifying specific targets for prevention. Behaviour problems have been found to present differently across elementary, middle, and high schools. In a large-scale study by Spaulding et
al. (2010), defiance was found to be one of the most commonly exhibited behaviour problems across all grade levels (elementary school: 29%, middle school: 31%, and high school: 24%). Peer-directed problem behaviours, such as fighting (32%), were among the most commonly exhibited problem behaviours in elementary school, whereas in middle school, problem behaviours were more often adult-directed, such as defiance (31%) and disruption (18%). By high school, tardiness (24%) and truancy (21%) were among the most common behaviour referrals, highlighting the need to provide school-based support before they may disengage from school.

**From Reactive to Proactive**

Schools typically identify students at risk for behaviour challenges only once they have become too disruptive for classroom teachers to manage alone. By middle school, administrative decisions are primarily reactive and exclusive (e.g., detention, in-school suspension, or out-of-school suspension; Spaulding et al., 2010). Given the consistently high rates of behaviour problems in schools and the number of students at risk of poor behaviour outcomes, it is worthwhile to examine the effectiveness of punitive systems. Effective discipline systems can be evaluated according to their focus on instruction (e.g., teaching interpersonal skills and emotional regulation) as opposed to the immediate effects of excluding and punishing the student. Relying on exclusion and suspension in reaction to problem behaviour (sometimes termed zero-tolerance approaches) has been shown to decrease exposure to instruction and school engagement and even increase levels of problem behaviour (Skiba & Peterson, 2000). In such reactive systems, school teams tend to use discipline data only to determine suspensions and expulsions; however, by analyzing the type and frequency of problem behaviours from
discipline referrals, school teams can act to support student behaviours before they become entrenched (Sugai, Sprague, Horner, & Walker, 2000; Walker & Sprague, 1999).

**Measuring Behaviours with Office Discipline Referrals**

Office Discipline Referrals (ODRs) are standardized records used by many schools in the US and Canada for tracking incidents of problem behaviour. ODRs are structured referral forms for recording specific information about behaviour incidents, and typically include information regarding the date, time, location, specific type of behaviour (including minor or major offences), others involved, and administrative actions. ODRs provide a means of recording referral information in an easy, systematic, and consistent manner that offers ease of collection, ease of decision-making, and an observable product for teachers and school staff (McIntosh, Frank, & Spaulding, 2010; Sugai, Sprague, Horner, & Walker, 2000).

ODRs have been shown to be associated with broader social constructs (e.g., student and teacher engagement and perceptions of school climate; McIntosh et al., 2010), and they are related to negative student outcomes, in that number and type of ODRs received are predictive of academic and behavioural trajectories, including school failure and antisocial behaviours (McIntosh, Flannery, Sugai, Braun, & Cochrane, 2008; Tobin & Sugai, 1999). However, Spaulding et al. (2010) advise that the external validity for comparing rates of ODRs across schools can lack strength because of inconsistency within and between schools in completing referral forms (Spaulding, 2010). Furthermore, using ODRs as the sole source of data for screening for behaviour problems may not identify students with internalizing disorders (McIntosh, Campbell, Carter, & Zumbo, 2009).

However, provided that they are used consistently, ODRs have been shown to be an appropriate means of assessing and monitoring individual behaviours for decision-making.
purposes within a problem-solving model (McIntosh et al., 2010). Consistent use includes standard protocols and systems for recording, reporting, and storing ODR data. Such standards are critical not only for quantifying the data across schools and districts, but also to ensure appropriate data-based decision making. Without these standards, data can be too variable and unreliable to be useful for decision-making (Nelson, Benner, Reid, Epstein, & Currin, 2002).

Final ODR categories (whether students receive 0-1 ODRs, 2-5 ODRs, and 6+ total ODRs by the end of each school year) have been demonstrated as an effective means of categorizing student behaviour needs (McIntosh et al., 2009; Walker, Cheney, Stage, & Blum, 2005). McIntosh et al., (2009) examined the relation between final ODR category and behaviour rating scale data for a group of 40 elementary school students identified for behaviour support. Levels of problem behaviour and adaptive skills, as measured through rating scales, were significantly different based on the student’s final ODR category (i.e., students in higher categories had higher ratings of externalizing behaviours and lower ratings of adaptive skills).

Tobin, Sugai, and Colvin (1996) sought to determine whether the frequency of ODRs during the first term of middle school were predictive of referral rates in later terms. In a study examining discipline records from 36 students in Grades 6 to 8, they found that the number of ODRs received in the first term of Grade 6 significantly predicted high numbers of referrals in subsequent terms. In addition, students who were suspended in the first term of Grade 6 had consistently higher numbers of ODRs in Grade 8 compared to other students. Because their study was conducted with schools that were implementing consistent, team-based approaches to discipline, the results may be an underestimate of predictability, given that other schools may rely even more on exclusionary discipline. However, the results provide interesting information regarding the utility of ODRs for identifying behaviour problems early in middle school.
Tobin and Sugai (1999) found that ODRs were predictive of later school failure and antisocial behaviours. Tobin and Sugai conducted regression analyses with archival ODRs of 526 students from Grades 6 to 8. They found that some referrals in Grade 6 were predictive of violent referrals in Grade 8 and that students with chronic discipline problems throughout middle school could have been identified in Grade 6 based on the number or type of ODR. Tobin and Sugai acknowledged the importance of using data to identify students as early as possible to improve outcomes for these students. More recent research has continued to indicate the predictive power of ODRs. The Rochester youth development study and the Denver youth study found that male and female gang members in both elementary and middle school made up 63% of all delinquent behaviour referrals (Thornberry, Huizinga, & Loeber, 2004). The Denver youth study also found that gang members made up 80% of serious and violent crimes, not including gang fights. These findings further support the potential for ODRs to identify students for intervention.

As a means of investigating patterns in a large-scale sample of elementary schools, McIntosh et al. (2010) conducted an archival analysis of 990,908 student records from Kindergarten to Grade 6. In this study, ODRs were categorized according to established cut points (0-1, 2-5, and 6+ ODRs per year) for predicting outcomes. They found that the number of ODRs received by an individual student in the fall was moderately predictive of the total number of ODRs received that year. Specifically, they found that receiving 1 or more ODRs in September was predictive of the number of ODRs received in later months. The authors also examined whether the types of behaviours referred enhanced prediction of student behaviours. Results demonstrated less additional predictive power based on the type of behaviour in an
elementary population as opposed to middle schools; however, receiving an ODR for defiance or physical aggression were the most accurate predictors of having 6 or more total ODRs.

Because patterns of problem behaviours vary by grade levels, it is unknown whether similar trends are apparent in a middle school population as have been observed with an elementary population. Moreover, although Tobin and Sugai (1999) demonstrated the predictive power of ODRs in middle school, their data are nearly 20 years old and relied on end-of-the-term and end-of-the-year behaviour data. Therefore, a method for earlier identification is proposed at the middle school level based on the promising findings of McIntosh et al. (2010). In addition, the current study proposes to replicate Tobin and Sugai’s (1999) research study but on a larger scale, with a population of schools that have adopted a common system for collecting and reporting school-wide social behaviour data. Tobin and Sugai’s research, though informative, is in need of replication with a larger, more current sample of schools to better understand the utility of ODRs in middle schools today.

**The Current Study**

The present study identified trajectories of ODRs in a middle school population in an effort to predict end-of-year behaviour outcomes as early as the first few months of the school year. The study explored patterns of ODRs in middle school and examined the accuracy of using ODRs for early identification based on specific months (e.g., the first, second, and third month) and types of referral problem behaviour to determine the earliest point at which accurate identification for additional support can be made. Specifically, this study examined the following questions:

1. What are the ODR trajectories, by month, for students in each final ODR category (0-1, 2-5, and 6+ total ODRs)?
2. How accurately does the cumulative number of ODRs received per month, for the first three months of the school year, predict the final ODR category?

3. Does screening by type of ODR (e.g., fighting, arson) received by the end of September enhance prediction of final ODR category in a middle school population?
Chapter 2: Method

Participants and Settings

Participants in this study were 401,852 students from 593 public middle schools (serving Grades 6 to 8) in the United States. The research used archival records from the School-wide Information System (SWIS, 2011; www.swis.org) in the 2009-2010 academic year. SWIS is a web-based information system that provides a means of recording ODR data with fidelity across schools, districts, and states. The system is currently used by 1532 middle schools for gathering, entering, reporting, and analysing individual ODR data (SWIS, 2012). Private schools, alternative/juvenile justice schools, and year-round schools were excluded from analysis.

Twenty-three percent of the schools in this study were considered urban (n = 136), 34% were suburban (n = 201), 17% were small-towns (n = 100), and 26% were rural (n = 154). A total of 59% of schools in the sample were identified as Title 1 eligible (i.e., schools with at least 40% of enrolment from low-income families, thus eligible to receive government funds to meet the needs of high numbers of at-risk and low-income students). Most schools using SWIS were also implementing SWPBS to some extent. Fidelity of implementation data were available for some schools to provide an indication of the extent to which SWPBS implemented with fidelity. A total of 37% of the schools in the sample provided fidelity of implementation data; 14% provided data using the Benchmarks of Quality (BoQ) tool and 23% reported fidelity data using the School-Wide Evaluation Tool (SET). A total of 75% of reporting schools (27% of the entire sample) met criteria for fidelity of implementation (BoQ ≥ 70% fidelity; SET ≥ 80% fidelity).

The average student enrollment for schools in the sample was 677.30 (SD = 292.58, MIN/MAX = 20 to 2149). An average of 48% of students in a given school were receiving free or reduced lunch (SD = 35%, MIN/MAX = 13%-83%), and the mean student to teacher ratio was
16:1 (SD = 3.60). Schools in the sample comprised of a mean of 58.3% Caucasian students (SD = 30.4%, MIN/MAX = 0.0% to 100%) and 40.7% non-Caucasian students (SD = 30.7%, MIN/MAX = 0.0% to 100%). The total number of ODRs for the entire year was 403,172, received by 118,582 students. The number of students without ODRs was calculated using each school’s enrollment data from the National Center for Education Statistics. As expected, by the end of June, the largest proportion of students made up the 0-1 final ODR category: 83% received 0-1 total ODRs ($n = 335,154$), 12% received 2-5 total ODRs ($n = 46,886$), and 5% received 6 or more total ODRs ($n = 19,812$). As shown in Table 3.1, the largest proportion of ODRs were issued for defiance/disrespect (33.2%), disruption (16.2%), and physical aggression/fighting (14.0).

Table 2.1
Proportion of ODRs by Type of Problem Behaviour (N = 401,852)

<table>
<thead>
<tr>
<th>Type of ODR</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defiance/Disrespect</td>
<td>133,958 (33.2%)</td>
</tr>
<tr>
<td>Disruption</td>
<td>65,357 (16.2%)</td>
</tr>
<tr>
<td>Physical Aggression/Fighting</td>
<td>56,620 (14.0%)</td>
</tr>
<tr>
<td>Inappropriate Language</td>
<td>29,219 (7.3%)</td>
</tr>
<tr>
<td>Other Major</td>
<td>29,158 (7.2%)</td>
</tr>
<tr>
<td>Tardy</td>
<td>24,296 (6.0%)</td>
</tr>
<tr>
<td>Harassment/Bullying</td>
<td>21,120 (5.2%)</td>
</tr>
<tr>
<td>Dress Code Violation</td>
<td>6372 (1.6%)</td>
</tr>
<tr>
<td>Theft/Forgery</td>
<td>5924 (1.5%)</td>
</tr>
<tr>
<td>Technology</td>
<td>5850 (1.5%)</td>
</tr>
<tr>
<td>Property Damage/Vandalism</td>
<td>4976 (1.2%)</td>
</tr>
<tr>
<td>Lying</td>
<td>4842 (1.2%)</td>
</tr>
<tr>
<td>Out of Bounds</td>
<td>3299 (0.8%)</td>
</tr>
<tr>
<td>Inappropriate Affection</td>
<td>2265 (0.6%)</td>
</tr>
<tr>
<td>Use/Possession of Drugs</td>
<td>1675 (0.4%)</td>
</tr>
<tr>
<td>Unknown Major</td>
<td>1670 (0.4%)</td>
</tr>
<tr>
<td>Truancy</td>
<td>1650 (0.4%)</td>
</tr>
<tr>
<td>Possession of Weapons</td>
<td>1383 (0.3%)</td>
</tr>
<tr>
<td>Use/Possession of Tobacco</td>
<td>1318 (0.3%)</td>
</tr>
<tr>
<td>Gang Display</td>
<td>905 (0.2%)</td>
</tr>
<tr>
<td>Use/Possess Combustible Items</td>
<td>605 (0.2%)</td>
</tr>
<tr>
<td>Use/Possess Alcohol</td>
<td>490 (0.1%)</td>
</tr>
<tr>
<td>Bomb Threat/False Alarm</td>
<td>134 (&lt;0.1%)</td>
</tr>
<tr>
<td>Arson</td>
<td>86 (&lt;0.1%)</td>
</tr>
<tr>
<td>Total Major ODRs</td>
<td>403,172 (100%)</td>
</tr>
</tbody>
</table>
Measure

ODRs are systematically recorded forms for documenting incidents of problem behaviour, and information regarding the date, time, location, and type of behaviour. The number and type of ODRs received by each student for major offences, based on SWIS classifications, was calculated for the entire school year and cumulatively by month. Minor offences, such as talking out in class, were not included given their varied use in practice. Students’ final count of ODRs in a given year was then categorized into three final ODR categories used in PBIS systems (0-1, 2-5, and 6+ total ODRs by the end of the year). For example, a student may have 3 ODRs in September, 2 in October, 1 in January, and 1 in May, resulting in 7 total ODRs at the end of the year, placing him or her in the 6+ final ODR category.

Design and Analysis

Data analyses were performed using Mplus 6.12 (Muthén & Muthén, 1998-2011) and SPSS 20 (IBM Corp., 2011). As a preliminary step, bivariate correlations were examined to identify any existing collinearity in the predictor variables. Collinearity was not indicated.

**ODR trajectories.** Initial analyses consisted of descriptive statistics for students according to the final ODR cut-point categories (0-1, 2-5, and 6+). Mean and median ODR trajectories per month were calculated, and the proportion of students who already qualified for their final ODR category per month was also examined (e.g., the percent of students with 6+ total ODRs who had 6+ ODRs by the end of September).

**Predictive analyses.** Multilevel multinomial logistic regression was used for all predictive analyses. Multinomial logistic regression reflects categorical data in a generalizable way, in that it provides an odds ratio describing the likelihood of being in a particular category based on specific predictor information. The coefficients in multinomial logistic regression are
interpreted in much the same way as in binary logistic regression, although in this case there are two threshold estimates (for the number of outcome categories minus one) as opposed to a single estimate, as with binary outcomes. The odds ratio is an indicator of the effect size of the predictor, indicating in the current study the likelihood of a student being in a final ODR category (0-1, 2-5, or 6+ ODRs). For example, in the present study, the log of the odds expresses the increased probability of a student being in the 0-1, 2-5, or 6+ final ODR category for each additional ODR received.

Although the outcome variable in the current study contained three categories that were ordinal in nature, the relation between the individual predictors and the logit were found not to be the same for all logits, violating the required parallel lines assumption for ordinal logistic regression. As a result, a multilevel multinomial logistic regression was conducted for all predictive analyses. This approach considered the violation of the parallel lines assumption but corrected for the nonlinear nature of the outcome variable. Intraclass correlation coefficients (ICCs) were calculated on the model containing only the random intercept to determine the shared variance in ODR data across individual schools. The ICCs for the two thresholds were .35 (2-5 ODRs) and .15 (6+ ODRs), indicating substantial variability at the school level (Peugh, 2010). Therefore, multilevel modeling was used to account for school-specific variability by including a school-level random intercept in each predictive model. The random intercept captured school level differences, given the varied use of ODRs, discipline practices, and ODR counts across schools. An a priori significance value of $\alpha = .01$ was selected as a criterion for statistically significant findings due to the size of the sample.
**Diagnostic statistics.** Diagnostic statistics were calculated to identify the accuracy of prediction of ODR categories. These statistics included area under the curve (AUC), sensitivity, and specificity. All diagnostic statistics were calculated using *SPSS*.

**Area under the curve.** The area under the curve (AUC) statistic provides the calculated assessment of diagnostic accuracy. Based on criteria from Rice and Harris (1995), AUC values of 0.50 to 0.59 represent low diagnostic accuracy, 0.60 to 0.65 represent moderate diagnostic accuracy, and 0.66 to 1.00 represent high diagnostic accuracy. Hanley and McNeil’s (1983) approach for comparing AUCs for predictive models was used for comparing the predictive power of the different variables.

**Sensitivity.** Sensitivity reflects the percent of occurrences correctly predicted, or true positives predicted by a given model [P(correct | ODR category met)]. Sensitivity determined the percent of students accurately classified as being in a final ODR category.

**Specificity.** Specificity reflects the percent of occurrences that were correctly predicted not to occur, or true negatives [P (correct | ODR category not met)]. Specificity determined the percent of students accurately classified as not being in a particular outcome category.

**Prediction with cumulative number of ODRs received per month.** Multilevel multinomial logistic regression was performed to assess the accuracy of cumulative ODRs received in September, October, and November in predicting the likelihood of a student receiving 0-1, 2-5, or 6+ ODRs by the end of the school year. Three separate analyses were run, each with one predictor variable: (1) cumulative ODRs by the end of September, (2) cumulative ODRs by the end of October, or (3) cumulative ODRs by the end of November. The categorical dependent variable for each of the three analyses was final ODR cutpoint, which categorized students into one of three groups according to the total number of ODRs received by the end of
June (0-1, 2-5, or 6+ ODRs). The 0-1 final ODR category was taken as the reference category, thus the predicted outcomes for all analyzed models resulted in two calculated threshold estimates: (1) the threshold to the 2-5 final ODR category, and (2) the threshold to the 6+ final ODR category.

**Prediction with ODR type (e.g., fighting).** Multilevel multinomial logistic regression was performed again to assess the effect the type of ODR received in September had for predicting the likelihood of a student being in the 2-5 or 6+ final ODR category at the end of the school year. This extended analysis examined the predictive effect of six predictor variables (total number of ODRs received in September, and number of ODRs received in September for: defiance, disruption, inappropriate language, physical aggression/fighting, and harassment/bullying). These five types of ODRs were most strongly correlated with the outcome variable (final ODR category).
Chapter 3: Results

Descriptive Analyses

ODR trajectories. Mean growth trajectories per month are detailed in Table 3.1, and mean and median ODR trajectories are shown by cut point in Figures 3.1 and 3.2 respectively. As seen in Table 3.1 and Figure 3.1, relatively stable slopes were observed throughout the year for students in each final ODR category. The mean monthly increase in ODRs for students in the 0-1 ODR category through to the end of June was 0.01 ODRs (SD = .01). An average increase of 0.29 ODRs (SD = .10) per month was observed for students in the 2-5 ODR category, and students in the 6+ ODR category by the end of June showed an average increase of 1.10 ODRs (SD = .35) per month.

Table 3.1
Mean Growth Trajectories by ODR Cut Point Group

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0-1 ODRs</td>
<td>335,154</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01(0.01)</td>
</tr>
<tr>
<td>2-5 ODRs</td>
<td>46,886</td>
<td>0.22</td>
<td>0.30</td>
<td>0.29</td>
<td>0.25</td>
<td>0.27</td>
<td>0.33</td>
<td>0.43</td>
<td>0.36</td>
<td>0.38</td>
<td>0.07</td>
<td>0.29(0.10)</td>
</tr>
<tr>
<td>6+ ODRs</td>
<td>19,812</td>
<td>0.80</td>
<td>1.19</td>
<td>1.15</td>
<td>0.96</td>
<td>1.11</td>
<td>1.22</td>
<td>1.53</td>
<td>1.28</td>
<td>1.22</td>
<td>0.22</td>
<td>1.10(0.35)</td>
</tr>
</tbody>
</table>
Figure 3.1 Mean growth in ODRs for students in total ODR categories by month.

Figure 3.2 displays the median growth trajectory and displays a progression that appears less smooth than was seen with the mean growth per month, particularly for students in the 2-5 and 6+ ODR categories. It appears that comparatively large median growth occurred in November and February compared to other months.
Figure 3.2 Median growth in ODRs for students in total ODR categories by month.

Referral trajectories for students in each of the final ODR categories are shown in Figures 3.3 through 3.5 relative to the month by which students met the final category criteria. Figure 3.3 displays only students with a total of 2-5 ODRs and the percent of those students who reached the 2-5 ODR category by month. As shown in Figure 3.3, of students in the 2-5 ODR category by the end of June, 13% were already in that category by September, 23% by October, 32% by November, and 42% by December.
Figure 3.3 Percent of students in 2-5 final ODR category in June with 2-5 ODRs per month (n = 46,886).

Similarly, Figure 3.4 displays only students with a total of 6+ ODRs and the percent of those students who reached the 6+ ODR category by month. As shown in Figure 3.4, of students in the 6+ ODR category by the end of June, 9% were in the 6+ category by September, 18% by October, 27% by November, and 38% by December.
Figure 3.4 Percent of students in 6+ final ODR category in June with 6+ ODRs per month (n = 19,812).

Figure 3.5 displays only students with a total of 6+ ODRs by the end of the year and the percent of those students who reached 2 or more ODRs by month. As shown in Figure 3.5, close to 50% of students in the 6+ final ODR category by the end of June, had 2 or more ODRs by September, and the percent increased drastically each consecutive month, reaching 100% by March.
Figure 3.5  Percent of students in 6+ final ODR category in June with 2+ ODRs per month (n = 19,812).

Predictive Analyses

**Prediction with cumulative number of ODRs received per month.** Multilevel multinomial logistic regression was performed to assess the accuracy and utility of predicting the likelihood of a student receiving 0-1, 2-5, or 6+ ODRs by the end of the school year based on the cumulative number of ODRs received in a) September, b) October, and c) November. As shown in Table 3.2, cumulative numbers of ODRs at the end of each of the three months were statistically significant predictors for whether a student received either 2-5 or 6+ total ODRs by the end of the school year. As would be expected, the strongest model was the November model, with odds ratios of 35.21 for the 2-5 ODR category, and 15.09 for the 6+ ODR category. These
statistics show that for each ODR received by the end of November, students were 35 times more likely to be in the 2-5 ODR category, and 15 times more likely to be in the 6+ ODR category at the end of the year. As shown in Table 3.2, similar odds ratios were observed for all three monthly models for both outcome categories.

Table 3.2
Multinomial Logistic Regression and Accuracy of Prediction of Final ODR Category

<table>
<thead>
<tr>
<th>Outcome and Model Predictor</th>
<th>β (SE)</th>
<th>OR</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-5 ODR Final Category</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September Cumulative ODRs</td>
<td>3.50 (.03)*****</td>
<td>32.96</td>
<td>.63</td>
</tr>
<tr>
<td>October Cumulative ODRs</td>
<td>3.49 (.03)*****</td>
<td>32.74</td>
<td>.73</td>
</tr>
<tr>
<td>November Cumulative ODRs</td>
<td>3.56 (.03)*****</td>
<td>35.20</td>
<td>.80</td>
</tr>
<tr>
<td>6+ ODR Final Category</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September Cumulative ODRs</td>
<td>2.63 (.03)*****</td>
<td>13.82</td>
<td>.71</td>
</tr>
<tr>
<td>October Cumulative ODRs</td>
<td>2.65 (.02)*****</td>
<td>14.20</td>
<td>.83†</td>
</tr>
<tr>
<td>November Cumulative ODRs</td>
<td>2.71 (.02)*****</td>
<td>15.09</td>
<td>.89†</td>
</tr>
</tbody>
</table>

Note. ODR = office discipline referral; OR = odds ratio; AUC = Area Under the Curve

*** p < .001  ** p < .01
† Indicates a statistically significantly improved AUC (p < .01) by including the behaviour.

**Model accuracy.** As seen in the last column of Table 3.2, the AUC results indicate that screening with cumulative ODRs by the end of September offered moderate diagnostic accuracy in predicting membership in the 2-5 final ODR category based on criteria from Rice and Harris (1995). October and November cumulative ODRs, on the other hand, had high diagnostic accuracy for the 2-5 category, but were not statistically significantly stronger predictors than previous months. In predicting membership in the 6+ final ODR category, screening with cumulative ODRs from any of the three months demonstrated high diagnostic accuracy, and each additional month was statistically significantly more accurate than the previous month in predicting student outcomes.

**Sensitivity and specificity.** Sensitivity and specificity outcomes are presented in Table 3.3. Sensitivity results for this analysis examined the ratio of true-positive to false-negative
predictions for students with 2+ total ODRs by the end of June, and 6+ total ODRs at the end of June, based on whether a student had 1 or more or 2 or more ODRs in each of the three months.

Table 3.3
Accuracy of Prediction of Final ODR Category

<table>
<thead>
<tr>
<th>Outcome and Model Predictor</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2+ ODRs by End of June</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 or More ODRs by September</td>
<td>.28</td>
<td>.99</td>
</tr>
<tr>
<td>1 or More ODRs by October</td>
<td>.48</td>
<td>.97</td>
</tr>
<tr>
<td>1 or More ODRs by November</td>
<td>.62</td>
<td>.96</td>
</tr>
<tr>
<td>6+ ODRs by End of June</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 or More ODRs by September</td>
<td>.46</td>
<td>.96</td>
</tr>
<tr>
<td>1 or More ODRs by October</td>
<td>.71</td>
<td>.93</td>
</tr>
<tr>
<td>1 or More ODRs by November</td>
<td>.83</td>
<td>.90</td>
</tr>
<tr>
<td>2 or More ODRs by September</td>
<td>.23</td>
<td>.99</td>
</tr>
<tr>
<td>2 or More ODRs by October</td>
<td>.49</td>
<td>.98</td>
</tr>
<tr>
<td>2 or More ODRs by November</td>
<td>.67</td>
<td>.97</td>
</tr>
</tbody>
</table>

Sensitivity statistics identified a high proportion of false negatives in both September and October, for predicting whether a student would have 2+ ODRs by the end of June, with a slight improvement in the November predictor model. The specificity statistics, however, demonstrated a high degree of accuracy in predicting which students did not have 2+ ODRs by the end of June, suggesting that although such a predictor model may not definitively identify in the first three months the majority of students requiring support, the model is unlikely to inappropriately categorize (i.e., overidentify) students who do not require additional behaviour support.

A relative improvement was observed for accurately predicting whether students would have 6+ ODRs by the end of June, based on receiving 1 or more ODRs in the first three months, demonstrating a particularly high degree of accuracy in both the October and November predictor models. The specificity statistics again demonstrated low rates of false positives.
Additional sensitivity and specificity statistics were calculated to determine how accurately 2 or more ODRs received by September, October, or November would predict whether a student would have 6+ ODRs by the end of June. Sensitivity statistics revealed similar outcomes to predictor models with the lower threshold (i.e., 1 or more ODRs in a month versus 2 or more ODRs in a month as the predictor). A low proportion of true positives (i.e., high proportion of false negatives) were observed when predicting 6+ final ODRs from 2 or more per month, yet specificity statistics again revealed high predictive accuracy for students who did not have 6+ ODRs by the end of June.

**Prediction with ODR type (e.g., defiance).** Multilevel multinomial logistic regression was performed to assess the effect of including type of ODR received in September on the predictive value of the September model for establishing the likelihood of a student being in the 2-5 or 6+ final ODR category at the end of the school year. Model statistics are shown in the first half of Table 3.4 and are displayed according to each outcome variable. Five of the six independent variables (cumulative number of ODRs received through September, and number of ODRs received in September for: defiance, inappropriate language, physical aggression/fighting, and harassment) made a statistically significant contribution to predicting whether a student would receive 2-5 total ODRs by the end of September; disruptive behaviour was not a statistically significant predictor ($p = .02$) given the strict a priori alpha values used. ODRs received for defiance ($\beta = .73$, OR = 2.20), harassment/bullying ($\beta = .72$, OR = 1.67), physical aggression/fighting ($\beta = .51$, OR = 1.19), and cumulative ODR count through September ($\beta = 3.13$, OR = 22.84) were positively related to predicting the 2-5 final ODR category. As an example, for each ODR for inappropriate language (holding constant the total number of ODRs obtained in September, and the number of ODRs obtained for defiance, harassment, physical
aggression, and disruptive behaviour in September), the odds of a student being in the 2-5 final ODR category by the end of June more than doubled (OR = 2.07).

Table 3.4
Prediction of Total ODRs by Type of ODR (by End of September)

<table>
<thead>
<tr>
<th>Outcome and Predictors</th>
<th>β (SE)</th>
<th>OR</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-5 ODR Category by End of June</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September Cumulative ODRs</td>
<td>3.13 (.04)***</td>
<td>22.84</td>
<td>.63</td>
</tr>
<tr>
<td>Inappropriate Language</td>
<td>0.79 (.06)***</td>
<td>2.07</td>
<td>.51</td>
</tr>
<tr>
<td>Defiance/Disrespect</td>
<td>0.73 (.10)***</td>
<td>2.20</td>
<td>.81†</td>
</tr>
<tr>
<td>Harassment/Bullying</td>
<td>0.72 (.09)***</td>
<td>1.67</td>
<td>.59</td>
</tr>
<tr>
<td>Physical Aggression/Fighting</td>
<td>0.51 (.09)***</td>
<td>1.19</td>
<td>.69†</td>
</tr>
<tr>
<td>Disruptive Behaviour</td>
<td>0.17 (.07)</td>
<td>2.05</td>
<td>.70†</td>
</tr>
<tr>
<td>6+ ODR Category by End of June</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September Cumulative ODRs</td>
<td>2.44 (.04)***</td>
<td>11.48</td>
<td>.71</td>
</tr>
<tr>
<td>Inappropriate Language</td>
<td>0.44 (.04)***</td>
<td>1.56</td>
<td>.53</td>
</tr>
<tr>
<td>Defiance/Disrespect</td>
<td>0.41 (.06)***</td>
<td>1.51</td>
<td>.91†</td>
</tr>
<tr>
<td>Harassment/Bullying</td>
<td>0.31 (.09)**</td>
<td>1.36</td>
<td>.65</td>
</tr>
<tr>
<td>Aggression/Fighting</td>
<td>0.17 (.06)**</td>
<td>1.18</td>
<td>.74†</td>
</tr>
<tr>
<td>Disruptive Behaviour</td>
<td>0.32 (.08)***</td>
<td>1.37</td>
<td>.81†</td>
</tr>
</tbody>
</table>

*Note. ODR = office discipline referral; OR = odds ratio; AUC = Area Under the Curve

*** p < .001  ** p < .01
† Indicates a statistically significantly improved AUC (p < .01) by including the behaviour.

For the 6+ final ODR model, all six variables were statistically significant predictors. ODRs received for defiance (β = .41, OR = 1.51), harassment/bullying (β = .31, OR = 1.36), aggression/fighting (β = .17, OR = 1.18), disruptive behaviour (β = .32, OR = 1.37), and cumulative ODR count by September (β = 2.44, OR = 11.48) were positively related to predicting the 6+ final ODR category.

Model accuracy. Although the variables were statistically significant predictors, the findings may simply be a result of the large sample size. Accuracy statistics may help explain the impact of these results. The far right column in Table 3.4 presents the AUC results for the predictive analyses using type of ODR and September ODRs to predict final ODR category. The AUC statistics were calculated independently for each predictor variable. Results showed that
when predicting membership in the 2-5 final ODR category, receiving ODRs for defiance, aggression/fighting, or disruptive behaviour was highly diagnostically accurate, according to Rice and Harris’ (1995) criteria. Defiance, aggression/fighting, and disruptive behaviour were statistically significantly more accurate compared to September cumulative count alone, while all other types of ODRs were statistically significantly weaker. When predicting the 6+ final ODR category, considering three of the five ODR types, defiance, aggression/fighting, and disruptive behaviour, was statistically significantly more accurate than using the September cumulative count alone.

**Sensitivity and specificity.** Table 3.5 displays the sensitivity and specificity results for correctly identifying students who had 2+ total ODRs by the end of the school year based on 1 or more ODRs for a specific behaviour. ODRs for inappropriate language offered the lowest rate of accurate positive predictions for both the 2-5 and 6+ final ODR models. In contrast, ODRs for defiance were the most promising in both models. Specificity remained high across all variables for both the 2+ and 6+ outcomes, indicating that cases predicted not to have 2+ or 6+ total ODRs by the end of the year were accurately identified as such.
Table 3.5
Accuracy for Prediction of Total ODRs by Referred Behaviour (by End of September)

<table>
<thead>
<tr>
<th>Outcome and predictors</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>2+ ODRs by End of June</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 or More of Any ODR in September</td>
<td>.20</td>
<td>.96</td>
</tr>
<tr>
<td>1 or More Inappropriate Language</td>
<td>.03</td>
<td>1.0</td>
</tr>
<tr>
<td>1 or More Defiance/Disrespect</td>
<td>.67</td>
<td>.97</td>
</tr>
<tr>
<td>1 or More Harassment/Bullying</td>
<td>.19</td>
<td>.99</td>
</tr>
<tr>
<td>1 or More Aggression/Fighting</td>
<td>.42</td>
<td>.97</td>
</tr>
<tr>
<td>1 or More Disruptive Behaviour</td>
<td>.42</td>
<td>98</td>
</tr>
<tr>
<td>6+ ODRs by End of June</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 or More of Any ODR in September</td>
<td>.46</td>
<td>.96</td>
</tr>
<tr>
<td>1 or More Inappropriate Language</td>
<td>.05</td>
<td>1.0</td>
</tr>
<tr>
<td>1 or More Defiance/Disrespect</td>
<td>.91</td>
<td>.90</td>
</tr>
<tr>
<td>1 or More Harassment/Bullying</td>
<td>.31</td>
<td>.97</td>
</tr>
<tr>
<td>1 or More Aggression/Fighting</td>
<td>.55</td>
<td>.93</td>
</tr>
<tr>
<td>1 or More Disruptive Behaviour</td>
<td>.68</td>
<td>.95</td>
</tr>
</tbody>
</table>
Chapter 4: Discussion

The goal of the present study was to examine the adequacy of using Office Discipline Referral (ODR) data as an early screening measure for flagging behaviourally at risk students. Specifically, the study examined rates of ODRs and student trajectories in a middle school population, and using multilevel multinomial logistic regression, assessed whether end-of-year behaviour outcomes could be predicted accurately based on number of ODRs and type of ODR received as early as the first few months of middle school. Descriptive results identified defiance, disruption, and aggression/fighting as the most prevalent ODRs received in middle school. The rate of ODRs received followed a steady trajectory throughout the school year. The predictive results revealed that although screening with ODR counts alone by October and November was moderately predictive of final ODRs, the most accurate screening results included the type of referral received, specifically referrals for defiant behaviour, in September for predicting 6 or more ODRs.

These findings add to the current literature in that they shed light on the most valid use of ODRs as screening tools for identifying behaviour problems early in middle school. The results corroborate findings from Tobin, Sugai, and Colvin’s (1996) small-scale study in which behaviour data in subsequent school terms was effectively predicted from ODRs received in the first term of each middle school grade. The current study further supports Tobin & Sugai’s (1999) assessment that students with chronic discipline problems in middle school can be identified early in the school year through ODR data.

ODR Frequencies and Trajectories

Based on the present sample, the most commonly received type of ODRs in middle school were defiance, disruption, and physical aggression/fighting, with inappropriate language,
tardies, and harassment/bullying behaviours falling close behind. Descriptively, results showed relatively stable mean increases in ODRs throughout the school year for students in each final ODR category, and relatively low mean numbers of ODRs at the beginning of the year. These results were consistent with previous research demonstrating similar patterns of mean growth across months for each category (McIntosh et al., 2010). However, median growth trajectories revealed a comparatively large increase in ODRs received in November and February. Presumably, this increase may have a number of causes that were not within the scope of the present study to examine, however, it offers insight into the months in which supplementary universal (tier 1) interventions may be beneficial (i.e., November and February). The present results support the use of ODRs for identifying peak times in a school year for bolstering current intervention practices.

The percent of students in their final ODR category per month increased by approximately 10% in each of the first four months of the school year, and reached approximately 50% by January for both the 2-5 and 6+ final ODR categories. These numbers suggest that 50% of students requiring tier 2 and 3 levels of support (i.e., having 2-5 or 6+ ODRs by the end of June) could have been identified by January; however, it should be noted that when diagnostic accuracy was assessed using sensitivity and specificity statistics for the first three months, false negative predictions were prominent in the first three months relative to the rate of true positives. Although the proportion of true positive predictions increased and false negatives decreased with each additional month, predictions using ODRs alone should be interpreted with caution given the chance of students who required support not being identified based on this simple model. It remains probable, however, that for students in the 2-5 and 6+ final ODR categories, interventions beginning before or by January may contribute in an important way to
efforts to minimize these negative behaviour outcomes (e.g., school failure and antisocial
behaviour problems; Tobin & Sugai, 1999). Identifying specific patterns of behaviour as early as
the first term of middle school would allow for more informed and responsive intervention for
students before behaviour patterns become fixed and/or continue to escalate.

**Prediction with Cumulative Number of ODRs Received Per Month**

ODRs received in each of the first three months of the school year were found to be
statistically significant predictors of whether a student received 2-5 or 6+ total ODRs by the end
of June. Students who received an ODR in September, October, or November were significantly
more likely to have 2-5 ODRs by the end of the school year, and although the overall accuracy of
such predictions were found to be moderate to high, a higher rate of false negative predictions
were observed compared to true positives until November. This finding suggests that although
the results were statistically significant, using monthly ODRs to predict the 2-5 category is not
diagnostically supported as a stand-alone process, at least not until later months (e.g., November
or later), due to low sensitivity. On the other hand, the 6+ final ODR category was accurately
predicted using ODRs received in October and November, with each month demonstrating high
diagnostic accuracy and higher proportions of true positive predictions. October and November
were each statistically significantly more accurate than September as predictors, and November
more accurate than October. According to these findings, students who required tier 2
interventions (i.e., had 2-5 final ODRs) could have been accurately identified for intervention as
early as November, with a moderate risk of not identifying students requiring support, while
students who required tier 3 interventions (i.e., 6+ final ODRs) could have been accurately
identified for intervention as early as October.
Whether a student received 2 or more ODRs compared to 1 or more ODR did not improve predictive accuracy of the 6+ final category, as it did not increase the proportion of true positives. These findings point toward the value of making decisions based on 1 or more ODRs received in the first three months. Such a disparity may have been due to the fact that comparatively fewer students in the sample received 2+ ODRs in September (2%; n = 6791) than 1+ ODRs (6%; n = 22,862).

The rate of false positive predictions of either final category, was very low. In other words, numbers were minimal for students who would have been identified as requiring support who in fact were not in need. Although this is a promising outcome regarding costs associated with providing services to those not in need, the data also demonstrated that a number of students in the early months might have gone unidentified for services where there was a need. Therefore, students who required service may have been overlooked at a higher rate, whereas students who did not require service ran little risk of being identified and improperly supported. Greater accuracy would be ideal in terms of sensitivity to ensure increased identification for students in need of interventions. Although the most serious implication for students who are identified for service that are not in need might be that they receive instruction or strategies for a behaviour problem that may not have persisted, the chance of this error was negligible when screening with ODRs. A higher degree of risk remains for students who would not have been identified for intervention where needed, further implicating the need for incorporating additional decision-making criteria to capture these students, including adding type of ODR.

**Prediction with Type of ODR (e.g., defiance)**

The current study found that including the type of ODR received in September offers significantly more reliable, sensitive, and efficient data for identifying at risk students. ODRs
received in September for defiance, inappropriate language, aggression/fighting, and harassment/bullying were statistically significant predictors of the 2-5 and 6+ final ODR categories. These findings are similar but distinct compared to results from the smaller scale, longitudinal research of Tobin & Sugai (1999), who found that violence, aggression, or harassing behaviours were effective screeners for identifying students in Grade 6 who were at risk for later violent behaviours, chronic discipline problems, or school failure.

Defiance was the most frequently received type of ODR, and it was the strongest predictor, with the highest rate of true positive predictions; even higher than the total number of ODRs received in September. The significant improvement in predictive value observed by including ODRs for defiance in the model suggests strongly that when a student receives an ODR for defiance in September, early intervention is warranted to avert predicted negative outcomes. Including defiance in the screening criteria enhanced the sensitivity of prediction of the 6+ final ODR category, so that the limitations observed by using only the number of ODRs (i.e., a high rate of false negatives) was strongly mitigated. Based on these findings, if a student receives a single ODR for defiance in September, it is highly likely that this student will persist with chronic discipline problems and would benefit from early intervention and support.

It is helpful to hypothesize why an ODR for defiance was so strongly predictive of behaviour outcomes. It may be that teachers are more likely to send a student out of class for defiant behaviour because it is directed at the teacher or staff, and may be viewed as more challenging and threatening compared to more general disruptive behaviour (i.e., loud, irritating). Defiance may also indicate a pattern of negative interactions between teachers and students. These interactions may develop into a coercive relationship between teacher and student by which the student and teacher learn that removal from class serves each of their immediate needs.
for escaping an aversive event (McIntosh et al., 2008; Patterson, 1982). Such a pattern may lead to further exclusion for the student, reducing access to school-wide preventive interventions (including academic instruction), and reinforcement of exclusionary discipline practices by the classroom teacher. It is critical that upon an initial ODR for defiance, the team should assess this pattern of teacher-student interactions to assess whether a coercive process exists. Examination and support of challenges in the teacher-student relationship at this early month may minimize the damage that can occur over the span of a school year, thus reducing the perpetuation of this dangerous cycle. It is also possible that defiant behaviours may be associated more strongly with chronic behaviour problems and negative school outcomes due to the long-lasting effects of ecological factors, such as family systems (e.g., control, abuse, intimidation), modelling, and parenting styles (Swick & Williams, 2006), on long term student outcomes. Therefore, receipt of ODRs for defiance should be considered when planning interventions.

**Limitations and Future Research**

There are a number of limitations that should be considered when interpreting the results of this study. ODRs were used in this study as both predictor variables and outcome variables, thus results should generally be interpreted with some caution. Although this study was designed with standard practice in mind, and accounted for the outcome variable by categorizing it based on support of previous research, future research might consider alternative outcome measures to reduce any potential inflation due to this model. For example, research might consider using results from response to intervention (RTI) data in short-term studies, and longitudinal research might consider the use of specific outcomes, such as identified behaviour disorders (including rating scales and/or medical diagnoses), suspensions/expulsions, dropout, arrests, or criminal charges.
The use of archival ODR data from a wide range of schools across the country, although regulated to some degree by the SWIS data entry system and controlled in part by the size of the sample itself, is limited in the consistency of how behaviours are defined and tolerated within schools, as well as subjectivity in identifying behaviour problems amongst teachers within each individual school. Furthermore, only 27% of the current sample met fidelity of implementation data for PBIS programs, indicating that a small percentage of schools identified that they were reliably implementing PBIS at the time that ODRs were being entered. It is unknown whether schools implementing PBIS varied from those not implementing such programs. Future research might endeavour to obtain referral data from similarly run schools and districts with consistent operational definitions and expectations for behaviour. The present study used previously existing data, thus the consistency of implementation and use of the ODR data is not known. Although SWIS provides operational definitions for each referral type in the database, emphasis should be placed on training school staff to ensure consistency of the use and understanding of these definitions. Even with this in mind, there are a number of personal and ecological factors that may contribute to how a teacher identifies a problem behaviour, and the number of times a behaviour is displayed before an ODR is issued, thus teachers may respond differently even within a more structured environment resulting in a more complex nesting effect that was not examined in this study. School teams should maintain an open dialogue and commit to regular consultation regarding common behaviours, their defining features, functions, and school-based consequences.

The present study was limited in its perspective in that it was unable to examine outcomes longitudinally due to lack of data regarding student dropout, program transfer, or suspension/expulsion. Moreover, the study was unable to track students who may have been
suspended, expelled, or who left the school part way through the school year. As such, future research might consider collecting data specific to such student outcomes and behaviour consequences as a means of tracking students with significant behaviour challenges with extreme consequences and from year to year.

**Implications for Practice**

As the dedication to early intervention progresses in schools, identification of at risk students requires a reliable, sensitive, and efficient method and procedure to support schools in this transition. ODRs have been demonstrated to be effective means of identifying students in need of additional support (McIntosh et al., 2010). The ODR mean growth trajectories shown in the current study support previous findings that ODRs might be utilized as long-term progress monitoring tools, as suggested by McIntosh et al., and the current findings support their use at the middle school level as well. However, although promising and in support of previous findings, the results of the present study also suggest a need for caution in making decisions based on ODR counts alone. ODR counts are not suggested as a sole screening measure, but they may be useful as part of a screening system or in monitoring long-term intervention success. The inclusion of type of ODR, particularly defiance, is an important way in which schools can identify students requiring intervention by September of their school year. It is important to note that effective ODR systems require definitively operationalized definitions for behaviours, and ODRs received outside such a system should be interpreted with caution. The accuracy of ODR model predictions might, however, be improved, compared to that seen in the present study, when objective definitions and clear procedures are implemented. Schools choosing to utilize ODR data as a means of identifying students for additional services should work to establish
consistently implemented, clearly operationalized, behaviour definitions and expectations prior to using their data for decision making purposes.

Given that ODRs consist of existing data required by most schools for reporting purposes, they are reasonable tools for tracking behaviour data and for identifying student needs. The use of ODRs in schools should be considered as a means of identifying students who may benefit from extra supports and interventions, but may not catch all students in need, even when considering the type of ODR. Many additional factors should be taken into consideration in addition to ODR data to improve the accuracy of true positive predictions. Not only should school teams continue to consider other ecological factors (e.g., family, peers, academic and behavioural history) and medical/mental health status to ensure a well-rounded, well-informed picture of the student, but as ODRs are utilized it will be exceedingly important to ensure consistent implementation and regular assurance of fidelity of implementation of ODR use as well as the intervention decisions and practices.

The results of this study provide strong evidence that school teams can identify students at risk for severe behaviour challenges as early as September by flagging those students referred for defiant behaviour. As prominent as detentions, in-school suspensions, and out-of-school suspensions are at the middle school level (Spaulding, 2010), they lack long-term effectiveness. Identifying students early may break cycles of defiance that lead to chronic ODRs. However, ODRs as a means of identification are recognized to some degree as a reactive tool as well, and although their use in the first three months supports earlier identification, they should not substitute and cannot replace the benefits of universal behaviour interventions as true methods of prevention.
As a result of the present findings, it is suggested that school behaviour support teams meet at the end of September each year to review the ODR data from the month. A member of the team can bring all ODR data collected to date, and review the students who received specific referrals. Students who received 2 or more ODRs or even a single ODR for defiance should be flagged, along with students identified through other means, such as teacher nomination or formal screening systems. At this time, for each flagged student, a file review and assessment of other data, such as academic skills, along with assessment of the teacher-student relationship, can be examined. The results of this study support intervention for each student flagged by this process, and the type and level of intervention would be selected by the team based on the additional information obtained (e.g., coercive student-adult relationships, problematic peer relationships, mental health concerns, significant life events, trauma, academic difficulties).

Schools might also consider defiance and disruptive behaviours for school-wide interventions, given their prominence in proportions of ODRs received. In fact, proportions ranging from 14% to 33% for the top three most frequent ODRs suggest that middle school teams may wish to focus universal interventions and expectations on the top three types of behaviours (i.e., defiance, disruption, and physical aggression/fighting). Universal level program planning can be positively informed in a community by either building from the results of the current sample, or by examining ODR data from previous school years in one’s own school or district.

Lastly, understanding the nature of specific types of behaviour problems as they are related to school dynamics and ecological constructs may be important in elementary and high schools that incorporate the same grade ranges seen in the present study. Given the influence of teacher expectations and school culture on prevalence’s of problem behaviours, as demonstrated by Gietz and McIntosh (2013), elementary and high schools designed to serve grades overlapping
the middle school range (e.g., Grades K-7; Grades 8-12) should consider the implications of these results for their Grade 6, 7, and 8 students. For example, elementary schools that include students from Grades 6 and 7 might consider that behaviours commonly observed at the elementary level may not be as relevant for these grade levels. School wide universal interventions that focus mainly on behaviours relevant to Grades K-6, may be missing out on an opportunity to begin teaching appropriate behaviours, skills, and tools to contend with behaviours such as defiance and disruption for students transitioning into a new age. Likewise, students entering high school in Grade 8 may require ongoing support and intervention for behaviours such as defiance and disruption in a high school environment that may have different, possibly more severe consequences for such behaviours. Considering that common behaviours may be different for students in these grades compared to earlier or later grades, school teams would benefit from examining their ODR data for these grade levels in particular, and work within their community to determine suitable school based interventions appropriate for their students. In this manner, as well as those stated above, the present study demonstrates how ODR data, a commonly collected form of behavioural information in schools, can effectively be used to assist school teams in identifying students for early interventions, working to offer at risk students their best chance for success.
References


