5-10-2018

Examining the Predictive Validity of the Strengths and Difficulties Questionnaire Using Office Discipline Referrals

Ashley Jacinth Salmon
Georgia State University

ABSTRACT

EXAMINING THE PREDICTIVE VALIDITY OF THE STRENGTHS AND DIFFICULTIES QUESTIONNAIRE USING OFFICE DISCIPLINE REFERRALS
By

ASHLEY J. SALMON

April 19, 2018

INTRODUCTION: There has been a demand for processes and appropriate tools to identify and route students with challenging behavior to the proper school-based supports. However, there is also a need for intuitive outcome measures that are feasible to be used in school settings.

AIM: In this study, we aim to assess the predictive validity of the Strengths and Difficulties Questionnaire (SDQ) using office discipline referrals (ODRs), a valid metric of student behavior.

METHODS: A series of non-parametric count models were used for analysis including Poisson (P), Negative Binomial (NB), Zero-Inflated Poisson (ZIP), and Zero-Inflated Negative Binomial (ZINB) models. These approaches were used to examine whether the internalizing and externalizing scales from the SDQ, administered at the beginning of the year, predicts total ODRs by the end of the year, even when controlling for the first 3 months of ODRs.

RESULTS: The ZINB model was chosen as the final model, as it had the best model fit (AIC). Our findings indicate that the SDQ's internalizing and externalizing subscales are significant predictors of ODRs. Specifically, the internalizing scale was a significant negative predictor of total ODRs, while the externalizing scale was a significant positive predictor of total ODRs. Additionally, the externalizing scale is a significant negative predictor of excess zeros in ODRs.

DISCUSSION: Findings from our study suggest that the SDQ is a psychometrically valid tool with predictive utility in relationship to an outcome of interest in schools across an array of statistical approaches. Future studies should validate the SDQ with other samples.

EXAMINING THE PREDICTIVE VALIDITY OF THE STRENGTHS AND DIFFICULTIES QUESTIONNAIRE USING OFFICE DISCIPLINE REFERRALS

by

ASHLEY J. SALMON

B.S., UNIVERSITY OF GEORGIA

A Thesis Submitted to the Graduate Faculty
of Georgia State University in Partial Fulfillment
of the
Requirements for the Degree

MASTER OF PUBLIC HEALTH

ATLANTA, GEORGIA
30303

APPROVAL PAGE

EXAMINING THE PREDICTIVE VALIDITY OF THE STRENGTHS AND DIFFICULTIES QUESTIONNAIRE USING OFFICE DISCIPLINE REFERRALS

by

ASHLEY J. SALMON

Approved:

Brian Barger, PhD
Committee Chair

Emily Graybill, PhD
Committee Member

April 19, 2018

Acknowledgments

Author's Statement Page

In presenting this thesis as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this thesis may be granted by the author or, in his/her absence, by the professor under whose direction it was written, or in his/her absence, by the Associate Dean, School of Public Health. Such quoting, copying, or publishing must be solely for scholarly purposes and will not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential financial gain will not be allowed without written permission of the author.

Ashley Salmon

TABLE OF CONTENTS

ACKNOWLEDGMENTS 4

INTRODUCTION..6-9

REVIEW OF THE LITERATURE..9-15
- 2.1 The Strengths and Difficulties Questionnaire……………………………………..…...9-10
- 2.2 Office Discipline Referrals…………………………………………………………..…..10-12
- 2.3 Approaches to Analyzing ODRs…………………………………………………….…..12-14
- 2.4 Other Factors Influencing ODR Receipt……………………………………………...14-15
- 2.5 Research Questions………………..15

METHODS AND PROCEDURES………………...16-19
- 3.1 Sample…………………………………………………………………………….......16
- 3.2 Measures……………………………………………………………………………16-17
- 3.3 Statistical Analysis…………………………………………..………………..17-19

RESULTS..19-21
- 4.1 Bivariate Analyses..……………..19
- 4.2 Multi-Model Predictive Analyses…………………………………………………19-20
- 4.3 Zero-Inflated Negative Binomial Predictive Analysis...20-21

DISCUSSION AND CONCLUSION...21-25
- 5.2 Discussion of Research Questions...21
- 5.3 Implications of the ODRs...21
- 5.4 Implications of Zero-Inflated Negative Binomial...22-24
- 5.5 Limitations………………..………..24-25
- 5.7 Recommendations and Conclusion…………………………………………………...25-26

LIST OF TABLES…………………………………………………………………......…27-29

REFERENCES..30-35

Introduction

Mental health difficulties in young people are of increasing concern because of their relationship to problem behavior, delinquency, substance abuse, educational attainment, and other public health concerns (Irvin et al., 2004; Esin et al., 2015; Eklund et al., 2017; Polgar et al., 2016). With as many as 1 in 5 young people developing mental health disorders in their youth (Burns, et al., 2016), it is imperative that early identification systems are established to help route children with mental health conditions to qualified professionals for intervention, thereby reducing the larger impact these conditions could have for the young person and their communities (Jones et al., 2002; Burns, et al., 2016). Calls for such systems and processes are not only coming from clinical professionals, but also from economists and policymakers who understand the negative impact of mental health difficulties on children's development and educational outcomes, which may have continued influence well into adulthood (Burns, et al., 2016).

Schools are increasingly considered an ideal place to implement preventive programs as most children attend school regularly and are likely to receive their initial identification and mental health support in school settings (Esin et al., 2015; McKown et al., 2016; & Doll et al., 2017). According to The White House (2013), schools can be helpful for ensuring students and young adults receive the necessary treatment for mental health issues that they need. Specifically, this report emphasizes schools as sources of early identification, referral for treatment, training for schoolteachers in early detection, and response to mental illness and clinical training.

To ensure students with behavior challenges get the necessary supports they need, they must first be identified. Many schools employ educators trained in mental and behavioral health who could contribute to a preventive mental health intervention (Bradshaw et al., 2008). Currently, federal regulations require that students are monitored and screened to identify mental health needs for special education services, such as psychosocial supports, particularly if there are negative educational impacts (IDEA, 1997). The Individuals with Disabilities Education Improvement Act (2004) allows educational agencies to use a response to intervention (RTI) model to aid in the identification of students with challenging behaviors. This current system includes problem-solving (Ikeda et al., 2002), response to instruction (Vaughn, Linan-Thompson, & Hickman, 2003), and standard protocol approaches (Vellutiono et al., 1996).

In recent years, three-tier RTI models have grown in use to deliver academic (Kame'enui & Carnine, 1998) and behavioral (Walker et al., 1996) support. Three-tier RTI models are unique in their emphasis on remediation of problem behavior as well as a key focus on preventive measures yielding a continuum of behavior support ranging from universal strategies focused on prevention for all students, to highly coordinated, individualized student interventions for those with more severe behavior challenges (McIntosh, Campbell, Carter & Dickey, 2009; McIntosh, Chard, Boland, & Horner, 2006). When supports are effective, about 80 to 90 percent of students will thrive from universal strategies (Sugai, Horner, & Gresham, 2002).

Tier 1 (universal strategies) strategies allow teachers to reward expected academic and social behaviors all while providing a structured environment promoting success (Lewis & Sugai, 1999). Examples of tier 1 strategies are defining and teaching school expectations, recognizing proper student behavior (Sugai et al., 2002), and providing active oversight of areas outside of the classroom (Colvin, Sugai, Good, & Lee, 1997). Of those students receiving individualized supports, 5 to 15 percent receive secondary-level (Tier 2) interventions which focuses on those who are at risk for problem behavior, while tertiary-level (Tier 3) interventions target 1 to 7 percent of students who present problem behavior that is more chronic (Crone & Horner, 2003; Sugai, Horner, & Gresham, 2002). Tier 2 supports include social skills interventions (Gresham, 2002) and behavior programs utilizing daily report cards (Hawkin & Horner, 2003), while tier 3 strategies are implemented on the basis of a functional behavior assessment providing highly, individualized supports (Crone & Horner, 2003).

Continuous monitoring of student progress helps inform decisions regarding student placement within the three-tier model (McIntosh, Campbell, Carter & Dickey, 2009). All students continue to receive universal supports even when receiving additional Tier I and Tier II supports so that a student may be removed from additional supports in the absence of its need (McIntosh, Chard, Boland, & Horner, 2008). The purpose of school-based universal screening is to identify children with emerging mental health and behavioral concerns in order to provide intervention and services before a referral process is necessary (Feeney-Kettler, et al., 2010, Gall et al., 2000; Eklund et al., 2017; Kaminski & Good, 1998). This approach to intervention allows students to receive instruction in a more fluid environment aiding in the transition up and down levels of support (Walker et al., 1996). Yet, more is needed to ensure successful early identification and routing to school-based mental health services (Lean & Colucci, 2013). These innovative service models will need to provide coordination of care from universal screening to school-based intervention engaging all key stakeholders to accurately assess the existing mental health needs of children in school settings (Doll et al., 2017).

There are a wide variety of universal screeners available, each with their own strengths and weaknesses (Cullinan & Epstein, 2013; Feeney-Kettler et al., 2010; Kamphaus et al., 2010; Dowdy et al., 2016; Jenkins et al., 2014). The majority of screeners have established good metrics for predictive validity indices and internal reliability (Walker et al., 1995; Squires, Bricker, & Twombly, 2003; Lovibond & Lovibond, 1995; Rusby et al., 2007; Hartman et al., 2017), however, some studies report weak findings for those metrics (Pelham, Gnagy, Greenslade, & Milich, 1992; McDougal et al., 2011). Other screeners like the Behavioral Assessment System for Children, 2nd Edition Behavioral and Emotional Screener System (BASC-2 BESS; Reynolds & Kamphaus, 2007) have numerous studies that establish acceptable to strong predictive validity (Feeney-Kettler et al., 2010; Kamphaus et al., 2010; Dowdy et al., 2016; Jenkins et al., 2014). However, some universal screeners target very young age groups (Walker et al., 1995; Squires, Bricker, & Twombly, 2003; Feeney-Kettler, Kratochwill, & Kettler, 2009) or have yet to be validated at the middle/high school level (Walker & Severson, 1992), while some are more expensive than many school systems can afford (Reynolds & Kamphaus, 2007; Walker & Severson, 1992).

Strengths and Difficulties Questionnaire

The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997), in particular, is a cost-free emotional and behavioral screening questionnaire validated for grades K-12 (Goodman, 2001). It is widely used for measuring child adjustment in relation to mental health problems utilized in clinical assessments, epidemiological studies, and survey research (Keller et al., 2017). As such, it could potentially be utilized as a universal screener in school settings. It is a well-established 25-item, Likert response survey that can be completed by parents, teachers, or self-reported by children between the ages of 11 and 17 years old that is based on fundamental domains of child symptoms described by the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) (Keller et al., 2017). It is derived from a child’s strengths and deficits in five speculative core dimensions namely, Emotional Problems, Conduct Problems, Hyperactivity/Inattention, Peer Problems, and Prosocial Behavior, and upon completion of the questionnaire, a total difficulties score (TDS) is obtained by summing up the scores for the four problem subscales, intentionally excluding the Prosocial score (Goodman, 1997).

In total, 15 items reflect problems (Ortuño-Sierra et al., 2015) that can be categorized into internalizing and externalizing problem behavior. Based on the literature on the five subscales, Emotional Problems and Peer Problems, which produces the internalizing score, have acceptable to good internal consistency, while Conduct Problems and Hyperactivity/Inattention, which produces the externalizing score, have acceptable to good internal consistency (Keller et al., 2017). The SDQ as a whole had acceptable to good internal reliability (Goodman et al., 1998; Goodman, 2001; Goossens et al., 2016). The SDQ had high predictive validity indices when predicting teacher ratings, daily behavioral performance, and quarterly grades (Owens et al., 2015) and low predictive validity indices when predicting a DSM-IV diagnosis and ADHD (Jenkins et al., 2014; Rimvall et al., 2014).

Office Discipline Referrals

Three-tier models require continuous progress monitoring (McIntosh, Campbell, Carter & Dickey, 2009). However, there is a demand to provide intuitive outcome measures related to a range of behavioral concerns to identify specific aims for intervention (McIntosh et al., 2009; Predy et al., 2014). Office Discipline Referrals (ODRs), in particular, are often used broadly in school settings as disciplinary action for students with challenging or problem behavior (Martell et al., 2010; Pas et al., 2011; McIntosh et al., 2010). According to Sugai et al. (2000), an ODR is “an event in which (a) a student engaged in a behavior that violated a rule/social norm in the school, (b) a problem behavior was observed 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”. ODRs are often used in progress monitoring to evaluate the impact of interventions and school policies (Sugai et al., 2000), for monitoring student behavior (Irvin et al., 2004; Pas et al., 2011), and as measures of decision making for student support services (Irvin et al., 2006; McIntosh et al., 2010).

A key advantage of using ODRs is the ability to sample behavior that would otherwise be problematic to directly observe (McIntosh et al., 2009), such as low-frequency, high-intensity problem behavior (Sprague, Horner, & Walker, 1999), due to the vast amount of time needed to produce accurate behavioral rates (McIntosh et al., 2009). Therefore, the records of ODR receipts provide for analysis of behavior in a more pragmatic way as opposed to observing students and waiting for the behavior to surface (McIntosh et al., 2009). ODRs also help inform school staff in the decision-making process through the documentation of student name, referring teacher, time of day, and nature/location of the problem behavior (Irvin et al., 2006). Through the assessment of referral patterns, the information gathered in ODRs help school staff with constructing and cultivating universal intervention programs as well as allowing school personnel to evaluate school safety status and behavioral climate (Sugai et al., 2000).

Although ODRs are useful for the purposes of mental health concern identification, they are limited. Since about 20% of students develop mental health issues (Burns, et al., 2016), using ODRs leaves the majority of students unidentified. “Research strongly suggests that 80 to 90 percent of children respond well to simple, school-wide discipline policies that emphasize good behavior” (Cortese, 2007, p. 7). With the majority of students unaffected by problem behavior, one would anticipate a low variability of ODRs causing it to be more difficult to make inferences. Also, with the nature of ODRs capturing disruptive behavior, internalizing issues are likely to be overlooked leaving some at-risk children undetected (McIntosh, Frank, & Spaulding, 2010; Tobin & Sugai, 1999; Tobin, Sugai, & Colvin, 1996).

Approaches to Analyzing ODRs

Despite their limitations, ODRs are a measure of interest to schools and researchers interested in negative behavioral outcomes. In the literature, ODRs are analyzed in a variety of ways (McIntosh et al., 2010; McIntosh 2009; Miller et al., 2015; Pas et al., 2011; Esin et al., 2015; Predy et al., 2014). Previous studies have used ODRs as a predictor variable (McIntosh et al., 2009; Pas et al., 2011; Miller et al., 2015; Esin et al., 2015) while others use ODRs as an outcome (McIntosh et al., 2010; Predy et al., 2014; Martinez et al., 2015). As a predictor, Tobin and colleagues (1996) found that receiving 2 or more ODRs in the first 3 months (August, Septmeber, and October) was a significant predictor of chronic ODRs in middle school students. Furthermore, studies have found that ODRs within the first 3 months of the school year significantly predicted total ODRs by the end of the school year (McIntosh et al., 2010) and mean ODR growth throughout the year (Predy et al., 2014). As an outcome variable, data to date has focused on particular ODRs (e.g., aggression, illicit behavior, etc.; McIntosh et al., 2010; Predy et al., 2014) or ODR cut points (e.g., 0-1, 2-5, and 6 or more; McIntosh et al., 2009; Miller et al., 2015; Pas et al., 2011). Collectively, available data indicates that ODRs are a flexible metric of child behavior.

Across multiple analyses, ODRs are associated with a number of relevant outcomes. For example, ODR receipt is related to a range of socio-demographic factors including race/ethnicity, gender, and age (Pas et al., 2011; Martinez et. al., 2016; Girvan et al., 2017). Other classroom and school factors may play a significant role in the receipt of an ODR like the consistency of implementing classroom rules, overall management in the classroom by the teacher, the number of students in a classroom, and a highly disruptive class (Pas et al., 2011). Martinez and colleagues (2016) found that student-teacher ratio and racial/ethnic concentration could also contribute to the receipt of an ODR. Whether a teacher has advanced training may also influence students' receipt of an ODR as well as school level factors like high faculty turnover, large school size, and how well the school is resourced (Pas et al., 2011; Martinez et. al., 2016).

There is currently no standard way of analyzing an ODR leading to numerous analytical approaches. Some studies treat ODRs as a continuous variable by looking at mean differences between various outcomes (e.g. suspensions, internalizing, externalizing, and adaptive scales) using MANOVA (McIntosh, Campbell, Carter & Dickey, 2009; McIntosh, Campbell, Carter & Zumbo, 2009). Some studies analyzed ODRs as a binary outcome utilizing logistic regression (Predy et al., 2014; McIntosh et al., 2010; Esin et al., 2015). Other studies analyzed ODRs as count data (Rusby et al., 2007; Flannery et al., 2014) utilizing Poisson and negative binomial regression (Martinez et al., 2015). Some studies accounted for clustered data using logistic hierarchical linear modeling (Pas et al., 2011), multilevel Poisson and zero-inflated Poisson models (Rusby et al., 2007) and negative binomial multilevel analysis (Martinez et al., 2015). The literature to date indicates a variety of analytical approaches for ODRs as a functional metric of student behavior.

Although one could approach ODR research in various ways, selected analyses may limit interpretations. ODR data exhibits a nonparametric distribution, yet many studies utilize parametric approaches to analyze ODRs (McIntosh, Campbell, Carter & Dickey, 2009; McIntosh, Campbell, Carter & Zumbo, 2009), while others analyze ODRs using a binary approach (Predy et al., 2014; McIntosh et al., 2010; Esin et al., 2015). Adopting parametric approaches, such as ANOVA and MANOVA, may result in inaccurate estimates, while binary approaches limit the scope of predictability by reducing the amount of information used in analysis (e.g. the receipt of an ODR vs. no receipt of an ODR) as opposed to making statistical inferences based on the actual ODR counts (Vives, Losilla, & Rodrigo, 2006; Long, 1997). Because ODRs are a count variable, it is appropriate to use a Poisson regression (Flannery et al., 2014) or a negative binomial regression in the case of overdispersion (UCLA: Statistical Consulting Group, 2017; Martinez et al., 2015), however, in the case of an inflated amount of zeros (due to a vast amount of students with 0 ODRs each year) predictive analyses may require other analytical approaches (Loeys et al., 2012; Zeileis, Kleiber, & Jackman, 2008; Hall, 2000). As stated by Rusby and colleagues (2007), “The zero-inflated Poisson model predicts the value of the dependent variable (the count of discipline referrals) as well as the probability of being unable to assume any value except zero using a logistic regression estimated with maximum likelihood with robust standard errors to account for the multilevel structure of the data.” Furthermore, a zero-inflated negative binomial model will account for both the excess zeros and any overdispersion that may exist (UCLA: Statistical Consulting Group, 2017). It is critical that the level of measurement for the dependent variable matches the model used (Long, 1997), so researchers should carefully consider analytical approaches to count data.

Other Factors Affecting Office Discipline Referrals

Challenging behaviors may indicate social-emotional problems in children including both internalizing and externalizing factors (Esin et al., 2015). The number of ODRs received is associated with negative student outcomes including school dropout, lower achievement, academic failure, and antisocial behaviors (McIntosh et al., 2008; Tobin & Sugai, 1999). Pas and colleagues (2011) suggest that students who receive ODRs may have social skill deficits, clinically significant problems with aggression, delinquent behavior, and attention problems. ODRs may also be used as screening measures to identify students who may require subsequent behavioral or mental health support (Tobin & Sugai, 1999).

ODRs are considered a good metric for monitoring challenging behavior for some researchers (Lane et al., 2008; Pas et al., 2011; McIntosh et al., 2009), while others have found it to be lacking (Miller et al., 2015). In a study conducted by McIntosh and colleagues (2009), ODRs predicted suspensions, internalizing and externalizing scores, and adaptive measures in elementary school students. Lane and colleagues (2008) established good predictive validity of the Student Risk Screening Scale (SRSS; Drummond, 1994) to predict level of risk using ODRs as a behavioral measure. Pas and colleagues (2011) found that students with 2 or more ODRs were rated by teachers as having significantly more disruptive behavior/concentration problems and fewer prosocial behaviors. Esin and colleagues (2015) found that psychiatric disorders were significantly higher amongst those with ODRs. Miller and colleagues (2015), however, found that ODR data did not perform significantly better than chance in identifying students at-risk on the BESS.

The literature to date establishes the SDQ as a reliable measure of internalizing and externalizing child behavior in school settings. Additionally, the SDQ is a reliable predictor of behavioral and academic outcomes including teacher ratings, daily behavioral performance, and quarterly grades. This positions the SDQ as a feasible tool for universal screening as it relates to student outcomes, such as ODRs. The literature establishes ODRs as a valid metric of child behavior, and ODRs are a behavioral outcome of interest to schools.

Current Study

The purpose of this study is to evaluate the extent to which the SDQ predicts the number of ODRs by the end of the school year. Specifically, the study examined the following questions:

1. Do screening results from the SDQ administered at the beginning of the academic year predict total ODRs by the end of the year even when controlling for preliminary (first 3 months) ODRs?
2. Do grade, gender, and race affect prediction of total ODRs by the end of the school year?

Understanding these relationships will display the psychometric utility of the SDQ while providing an intuitive outcome (ODRs) to measure student behavior. This will assist schools in identifying students with at risk behavior in order to route them to the proper school-based supports.

Methods

Participants and Setting

The data for this study was collected in the 2015-2016 school year. The sample consisted of 1134 male (n=617) and female (n=517) Georgia students from 2 middle schools. School 1 (n=328) and School 2 (n=806) consisted of grades 6 (n=362), 7 (n=366), and 8 (n=406). The majority of students were categorized as Black (n=712) followed by White (n=289), Hispanic (n=83), and Multi-other (n=50). These frequencies are summarized in Table 2.

Measures

Office Discipline Referrals. An ODR is a permanent product issued to a student who has engaged in problem behavior violating a school rule or social norm resulting in an administrative consequence. The ODR documents student name, referring teacher, time of day, and nature/location of the problem behavior. The number of ODRs received by each student [according to SWIS] was used in analyses as a covariate. The number of ODRs was calculated for the entire school year and cumulatively by month. For the purpose of this study, we focused on the number of ODRs received by each student ignoring the type of offense; therefore, all

ODRs were considered including both major and minor offenses (McIntosh et al., 2009; Martinez et al., 2016).

Strengths and Difficulties Questionnaire. Internalizing and externalizing scores were calculated using the SDQ subscales. The internalizing problems score is derived from adding the Emotion and Peer Problems questions together, while the externalizing score is derived from adding the Conduct Problems and Hyperactivity/Inattention questions together. The SDQ subscales are scored from 0-10, so the internalizing and externalizing scores range from 0-20.

Office Discipline Referrals (October). The number of ODRs from August, September, and October (Preliminary ODRs) was included in the models, because it is a stringent covariate of total ODRs (McIntosh et al., 2010). Specifically, the receipt of 2+ ODRs by October (2 or more ODRs = 1) was used. By including this valid metric in our predictive analyses, it provides a comparable standard for the internalizing and externalizing scores from the SDQ.

Socio-demographic Information. Student grade, race, and gender were included in all models, because the literature shows us that there are significant differences in ODR receipt based on these variables (Pas et al., 2011; Martinez et. al., 2016). Our study specifically looked at students in grades 6, 7, and 8 and whether differences in the outcome variable emerged. Gender (female=1) and race (Black, White, Hispanic, and multi-other) were also analyzed for group differences. Students who were black, male, and in the 6th grade served as the reference groups for our analyses (Martinez et. al., 2016).

Analyses

Preliminary. Data analyses were performed in *SAS* version 9.4. Bivariate analyses were examined as a preliminary step to identify any existing collinearity as well as variables that significantly correlate with the outcome to include in our models. The threshold for model inclusion was set to $p < 0.2$. We looked at the distribution of ODRs, which appeared to display a Poisson distribution. A Poisson distribution assumes that the mean and variance are equal; therefore, the mean and variance of ODRs were examined for overdispersion (Loeys et al., 2012). The mean, median, standard deviation, minimum and maximum values of each subscale were assessed. The data was also examined for outliers and influential cases.

Bivariate analyses were conducted with externalizing and internalizing scales predicting ODRs utilizing Spearman's Rho due to the violation in normality (Kitchen, 2009). It is important to note that the internalizing scale was not a significant predictor of ODRs, but it was still included in the model. Bivariate analyses were also conducted to examine group differences using Kruskal-Wallis and Wilcoxon Rank Sum test statistics for non-parametric data (Kitchen, 2009). It is important to note that grade was not a significant predictor of ODRs, but it was still included in the model to further examine group differences. For the prediction analyses, a series of models were fit to determine the best model for the data.

Predictive models. A Poisson model was first considered. Predictor variables considered for our model included the SDQ subscales (internalizing and externalizing), receipt of 2+ ODRs in the first 3 months of the school year (August, September, and October), and demographic information (grade, gender, and race). The model building process began with observing how the externalizing and internalizing scales predict ODRs. Model building continued with the externalizing and internalizing scales and receipt of 2+ ODRs by October predicting total ODRs. Finally, the full model was considered including the externalizing and internalizing subscales, receipt of 2+ ODRs by October, and socio-demographic information (grade, gender, and race). The Poisson (P) model was adjusted for overdispersion by adding the scale parameter (scale =

Pearson) (Hilbe, 2011). We also considered a Negative Binomial (NB) model to account for the overdispersed data (Hilbe, 2011).

Due to a high number of zeros (observed number of zero counts exceeds the predicted number of zero counts) in our outcome variable, a Zero-inflated Poisson (ZIP; Lambert, 1992) model and a Zero-Inflated Negative Binomial (ZINB) model were also considered (Loeys et al., 2012). These models are considered mixed models, because the distribution of the outcome is modeled by two separate components, one of which represents the probability of excess zeros and another in which accounts for the non-excess zeros and non-zero counts (Loeys et al., 2012). Model fit was determined by comparing model fit diagnostics, specifically the Akaike Information Criterion (AIC) (Loeys et al., 2012). The results of our model fit diagnostics are summarized in Table 1.

Results

Bivariate analyses were conducted with externalizing and internalizing scales predicting ODRs utilizing Spearman's Rho. Analyses indicated a significant positive correlation between ODRs and the externalizing scale (*rho* = 0.27, $p < 0.0001$). Bivariate analyses were also conducted to examine group differences using the Kruskal-Wallis test statistic for the race and grade variables, while the Wilcoxon Rank Sum test statistic was used for the gender variable. A Wilcoxon Rank Sum test indicated a significant difference for gender ($U = 274872.5$, $p < 0.0001$). Kruskal-Wallis indicated a significant difference for race ($H = 35.5$, $p < 0.0001$), and follow-up Wilcoxon Rank Sum tests indicated Blacks differed from Whites ($p < 0.0001$) and Hispanics ($p < 0.0001$). Wilcoxon Rank Sum tests also indicated Multi-other differed between Hispanics ($p < 0.01$) and Whites ($p < 0.01$). Group differences are summarized in Table 2. Furthermore, although not attaining significance, the internalizing scale and grade variable were still included in the final model based on group differences in ODR counts established in the literature (Pas et al., 2011; Martinez et. al., 2016; Girvan et al., 2017). Therefore, all variables were included in our final models.

To examine whether SDQ predicts the total number of ODRs by the end of the year, a series of models were fit (see Table 1). In the Poisson model, the externalizing scale was a positively significant predictor of ODRs ($p < 0.0001$) and the internalizing scale ($p < 0.01$) was a significantly negative predictor of ODRs even when controlling for receipt of 2+ ODRs by October. Grade was not a significant predictor of ODRs in this model, but being female ($p < 0.01$), White ($p < 0.01$), and Hispanic ($p < 0.01$) were significant negative predictors of ODRs. However, the Poisson model was overdispersed, so a Negative Binomial (NB) model was considered next (Hilbe, 2011). The externalizing ($p < 0.0001$) and internalizing ($p < 0.01$) scales positively and negatively predicted ODRs, respectively, in this model as well, even when controlling for receipt of 2+ ODRs by October. Additionally, being female ($p < 0.01$), White ($p < 0.01$), Hispanic ($p < 0.01$), and in 7th grade ($p < 0.01$) were significant negative predictors of total ODRs. To account for the excess zeros, a Zero-Inflated Poisson (ZIP) model was fit (Lambert, 1992). Again, the externalizing scale ($p < 0.0001$) was a significant positive predictor of ODRs and the internalizing ($p < 0.01$) scale was a significant negative predictor of ODRs even when controlling for receipt of 2+ ODRs by October. However, gender was not a significant predictor of ODRs, but being Multi-other ($p < 0.01$) was a significant positive predictor of ODRs and being in the 7th grade ($p < 0.01$) was a significant negative predictors of ODRs. As for the zero component of the model, the prediction of excess zeros in ODR receipts (i.e. no ODR coded as the event; no ODRs=1), externalizing scores was a significant negative predictor of excess zeros in ODRs ($p < 0.0001$), while being female ($p < 0.0001$) and White ($p <$

0.0001) were significant positive predictors of excess zeros in ODRs. For every one unit increase in externalizing scores, the odds of observing excess zeros in ODRs decreases by 17%. The odds of observing excess zeros in ODRs for females ($p < 0.0001$) was 2.05 times the odds for males. The odds of observing excess zeros in ODRs for Whites ($p < 0.0001$) was 2.94 times the odds for blacks.

The Zero-inflated Negative Binomial Model was ultimately selected as it displayed the best fit to the data (i.e. AIC = 2117.2; Loeys et al., 2012). Results of our final model are summarized in Table 3. In this model, the externalizing ($p < 0.01$) scale was a significant positive predictor of ODRs and the internalizing ($p < 0.01$) scale was a significant negative predictor of ODRs even when controlling for receipt of 2+ ODRs by October. Compared to 6th grade, 7th grade ($p < 0.01$) and 8th grade ($p < 0.05$) receive less ODRs. Compared to Blacks, Hispanics ($p < 0.01$) and Whites ($p < 0.01$) receive less ODRs. For the prediction of excess zeros (i.e. no ODR receipt coded as the event; no ODR=1), the externalizing scale ($p < 0.01$) was a significant negative predictor of no ODR receipts. For every one unit increase in externalizing scores, the odds of observing excess zeros in ODRs decreases by 55%. The odds of observing excess zeros in ODRs for females ($p < 0.0001$) was 17.6 times the odds for males. The odds of observing excess zeros in ODRs for Whites ($p < 0.0001$) was 10.2 times the odds for Blacks.

Discussion

This study provides data indicating the SDQ, administered at the beginning of the school year, significantly predicts the number of ODRs by the end of the school year. The externalizing scale, in particular, also predicted excess zeros in reported ODRs. Thus, data from the study shows that the SDQ can predict relevant child behavior outcomes in school settings. Furthermore, this study is unique in that it displays the predictive utility of the SDQ as a forecaster of total ODRs as well as excess zeros in ODRs by the end of the school year. Additionally, this study resonates with other studies by finding certain socio-demographic markers predict total ODRs. Ultimately, this indicates that the SDQ may be an effective tool in identifying students who may be at risk for problem behavior, which can assist school personnel with routing students to necessary supports provided by the school.

This study is in agreement with other ODR studies (Martinez et al., 2016;). Particularly, this study shows that externalizing behaviors predict ODRs (Pas et al., 2011; Rusby et al., 2007). Additionally, being White and female predict fewer ODRs (Martinez et al., 2016;), while being male and black predict greater ODRs (Martinez et al., 2016; Pas et al., 2011). However, our study contradicts some ODR studies by predicting higher ODR counts for 6th grade students compared to 7th and 8th grade students (Martinez et al., 2016;). For example, Martinez et al. (2016) found that younger students had fewer ODRs than older students; however, they compared middle schools and elementary schools without accounting for each grade level. Pas et al. (2011) also found that younger students received less ODRs than older students, however, they only examined elementary school students and compared fourth/fifth grades to kindergarten through third grade, again not accounting for each grade level. Thus, differences in studies may be due to the way the grade variable was coded for their analyses. Specifically, other studies described lumped multiple grades together into a binary, while our study split grade into more refined categories. Additionally, Pas et al., 2011 predicted the receipt of any ODR, a binary construct. Thus differences between studies may also be due to the treatment of the outcome variable in analysis.

This study is a unique addition to the literature by providing analyses of the probability of excess zeros. In particular, higher externalizing scores decrease predictive odds of observing a student with excess zeros in ODRs, while being female and White increases the odds of excess zeros in ODRs. This approach allows for the prediction of ODR counts while shedding light on the disproportionate count of ODR receipts for males and non-White students. Additionally, it exhibits a unique utility of the externalizing scale, in particular, as a predictor of excess zeros in ODRs. Furthermore, this may highlight the unique psychometric contribution to the receipt of no ODRs as opposed to only focusing on behaviors that may predict ODR receipt.

This study is unique in its analytic approach. We utilized a Zero-Inflated Negative Binomial model that accounted for the prediction of ODR counts as well as the probability of excess zeros in ODRs. Only a small number of ODR studies utilize Poisson (Rusby et al., 2007; Flannery et al., 2014), Zero-Inflated Poisson (Rusby et al., 2007) and, Negative Binomial (Martinez et al., 2016) models for predictive analyses, while a large proportion of ODR studies utilize logistic regression (McIntosh et al., 2010; McIntosh et al., 2012; McIntosh et al., 2006; McIntosh et al., 2009). No studies to date have utilized a Zero-Inflated Negative Binomial model, and none use the multi-model approach used here. This comparative count statistical method improves on binary approaches like logistic regression. Additionally, this method reveals the predictive utility of the SDQ across multiple statistical approaches. Looking at data one statistical method at a time may result in a loss of information.

This study is also unique in the universe of SDQ psychometric studies (Owens et al., 2015; Jenkins et al., 2014; Rimvall et al., 2014; Goosens et al., 2016). Most SDQ studies investigate the reliability of the SDQ in terms of its predictive relationship to other screeners (Owens et al., 2015; Jenkins et al., 2014) and some diagnostic tools (Rimvall et al., 2014). However, this study aims to showcase the applied predictive validity of the SDQ by addressing its relationship to ODRs, an applied outcome of interests to schools. Specifically, this study exhibits the SDQ as a valid predictor of student behavior, particularly showing how internalizing and externalizing behavior scores predict total ODRs and excessive zero counts in non-ODR receipt. By addressing ODRs, this study broadens the scope of usability for the SDQ as a socially valid tool with predictive utility in relationship to an outcome of interest to schools.

Our findings mirror some findings in other ODR studies (Rusby et al., 2007; Lane et al., 2008; Hartman et al., 2017). For example, results from our study are similar to a study conducted by Rusby et al. (2007) that used the Child Adolescent Disruptive Behavior Inventory (CADBI; Burns, Taylor, & Rusby, 2001), a measure of externalizing behaviors, in their prediction of total School Discipline Referrals (SDRs). They found that gender and risk status (65th percentile on CADBI defined the at-risk sample) significantly contributed to the prediction of SDR counts. Specifically, females had more SDR counts than males, and the at-risk sample had more SDRs in comparison to the universal sample. Lane et al. (2008) examined the predictive validity of the Student Risk Screening Scale (SRSS; Drummond, 1994), a measure of internalizing and externalizing behaviors, in predicting behavioral (i.e. ODRs) and academic (i.e. GPAs) outcomes. They found significant differences in ODRs and GPAs comparing low-risk, moderate-risk, and high-risk groups. They also examined the convergent validity of the SRSS with the SDQ and found them to correlate in a positive and statistically significant way. Hartman et al., (2017) examined the relationship between internalizing/externalizing scores and school outcomes (i.e. ODRs and GPAs). Specifically, they assessed the Student Internalizing Behavior Screener (SIBS; Cook et al., 2011) and the Student Externalizing Behavior Screener (SEBS; Cook et al., 2012) and found that all screening scores significantly predicted school outcomes (i.e. ODRs and GPAs). However, the SIBS alone was not significantly correlated to ODRs.

Furthermore, GPA was more strongly correlated with screening scores than ODRs. Therefore, our study contributes to the literature of other ODR studies.

This study highlights the predictive utility of the SDQ's internalizing and externalizing scores in relation to ODRs, an intuitive outcome of interest to schools. In particular, higher externalizing scores predict greater ODRs in older, Black, and male students compared to younger, White, and female students. Additionally, higher externalizing scores decrease the predictive odds of observing a student with excess zeros in ODRs in Black and male students compared to female, White, and Multi-other students.

Limitations

There are a number of limitations to consider in this study. First, the SDQ was a self-reported questionnaire, which may, in some cases, be unreliable. Besides, the effect sizes of the externalizing scales were not much bigger than the internalizing scales bringing into question its predictive utility. Furthermore, there are many adverse behaviors grouped into ODRs. Separating them for analysis may show moderating factors that were not assessed in this study (e.g. comparing aggressive behaviors with illicit drug use). Also, ODRs were used as a predictor and an outcome and may warrant caution when interpreting results. Moreover, external validity may be affected by various methods of office discipline referral processes across schools.

Future Research and Implications

Creating and implementing school-wide supports for students with behavior challenges, bypassing the referral process, will foster an environment that yields positive behavioral outcomes in our young people. It is imperative that we provide the tools for early identification as well as intuitive and measurable outcomes for assessing student behavior. Universal screening is a pragmatic tool that can be implemented in school settings that will assist in routing students to the proper supports. The findings in this study provide evidence that the SDQ's internalizing and externalizing scales are predictive of problem behavior in schools; however, these findings should be validated with other samples. Furthermore, it sheds light on the need to assess the predictive utility of screeners with other behavioral metrics aside from ODRs in an effort to capture more internalizing concerns that may go undetected.

Table 1. Predictive Analyses with Multi-Model Approach

		P	NB	ZIP	ZINB
COUNT COMPONENT					
(intercept)		-0.58	-0.66**	0.75***	0.004
		(0.1714)	(0.1933)	(0.1234)	(0.2427)
externalizing		0.14***	0.18***	0.05***	0.08**
		(0.0163)	(0.0218)	(0.0126)	(0.0245)
internalizing		-0.06**	-0.07**	-0.05**	-0.06**
		(0.019)	(0.0226)	(0.0143)	(0.0238)
preliminary ODR		1.51***	1.90***	0.80***	1.68***
		(0.2029)	(0.4703)	(0.1259)	(0.4076)
grade					
	6	REF	REF	REF	REF
	7	-0.37	-0.51**	-0.34**	-0.59**
		(0.15)	(0.1724)	(0.1114)	(0.1819)
	8	-0.24	-0.32	-0.19	-0.38*
		(0.1436)	(0.167)	(0.1063)	(0.1854)
gender					
	male	REF	REF	REF	REF
	female	-0.37**	-0.51**	0.03	-0.03
		(0.132)	(0.1473)	(0.104)	(0.1756)
race					
	Black	REF	REF	REF	REF
	White	-0.76**	-0.96***	-0.13	-0.55**
		(0.1684)	(0.1756)	(0.1276)	(0.2069)
	Hispanic	-1.15**	-1.26**	-0.48	-1.31**
		(0.409)	(0.3455)	(0.369)	(0.3589)
	multi-other	0.16	-0.03	0.52**	0.40
		(0.2326)	(0.3101)	(0.153)	(0.3775)
ZERO COMPONENT					
(intercept)				1.07***	0.08
				(0.2289)	(0.7907)
externalizing				-0.18***	-0.80**
				(0.0255)	(0.2098)
internalizing				0.03	0.08
				(0.0276)	(0.092)
grade					
	6			REF	REF
	7			0.10	-0.82
				(0.2104)	(0.7464)
	8			0.10	-0.18
				(0.2015)	(0.613)
gender					
	male			REF	REF

	P	NB	ZIP	ZINB
female			0.72***	2.87**
			(0.1805)	(0.7974)
race				
Black			REF	REF
White			1.08***	2.32**
			(0.2176)	(0.7733)
Hispanic			1.06*	-0.73
			(0.4778)	(1.5487)
Multi-other			0.71*	2.89**
			(0.3626)	(1.1452)
log *L*	-302.65	-514.41	-535.96	-1020.17
AIC	2646.47	2117.12	2176.30	2080.34

Notes: P = Poisson, NB = Negative Binomial, ZIP = Zero-Inflated Poisson, ZINB = Zero-Inflated Negative Binomial, *** = p<0.0001, ** = p<0.01, * = p<0.05

Table 2. Socio-demographic Group Differences

	Mean	SD	Min	Max
grade				
6	0.78 a	1.81	0	11
7	0.57 a	1.29	0	8
8	0.60 a	1.47	0	10
gender				
male	0.80 b*	1.71	0	11
female	0.46 b*	1.27	0	9
race				
Black	0.77 a*	1.65	0	11
White	0.38 a*	1.12	0	8
Hispanic	0.19 a*	0.65	0	4
Multi-other	1.1 a*	2.35	0	10

Notes: Wilcoxon Rank Sum[a] , Kruskal-Wallis[b], * = p<0.0001

Table 3. Predictive Analysis with the Zero-Inflated Negative Binomial Model

		Estimate	Standard Error	2.5% CI	97.5% CI	*p*
COUNT COMPONENT						
(intercept)		0.004	0.24	-0.47	0.48	
externalizing		0.08	0.02	0.04	0.13	**
internalizing		-0.06	0.02	-0.11	-0.01	**
preliminary ODR		1.68	0.41	0.88	2.48	***
grade						
	6	REF				
	7	-0.59	0.18	-0.94	-0.23	**
	8	-0.38	0.19	-0.74	-0.02	**
gender						
	male	REF				
	female	-0.03	0.18	-0.38	0.31	
race						
	Black	REF				
	White	-0.55	0.21	-0.96	-0.14	**
	Hispanic	-1.31	0.36	-2.01	-0.60	**
	Multi-other	0.40	0.38	-0.34	1.14	
ZERO COMPONENT						
(intercept)		0.08	0.79	-1.47	1.63	
externalizing		-0.79	0.2098	-1.2013	-0.3787	**
internalizing		0.08	0.09	-0.10	0.26	
grade						
	6	REF				
	7	-0.82	0.75	-2.28	0.65	
	8	-0.18	0.61	-1.38	1.03	
gender						
	male	REF				
	female	2.87	0.80	1.31	4.43	**
race						
	Black	REF				
	White	2.32	0.77	0.81	3.84	**
	Hispanic	-0.73	1.55	-3.76	2.31	
	Multi-other	2.89	1.15	0.64	5.13	**

*** = p<0.0001, ** = p<0.01, * = p<0.05

References

Bradshaw, C. P., Buckley, J. A, & Ialongo, N. S. (2008). School-based service utilization among urban children with early onset education and mental health problems: The squeaky wheel phenomenon. *School Psychology Quarterly*, *23*, 169–186. https://doi.org/10.1037/1045- 3830.23.2.169

Burns, J. R., & Rapee, R. M. (2016). Screening for mental health risk in high schools: The development of the Youth RADAR. Psychological Assessment, 28(10), 1220-1231. doi:10.1037/pas0000237

Colvin, G., Sugai, G., Good, R. H., & Lee, Y. (1997). Effect of active supervision and precorrection on transition behaviors of elementary students. *School Psychology Quarterly, 12,* 344–363.

Cortese, A. (2007, Spring). Get real: Here's the boost that poor children, their teacher, and their schools really need. American Educator, 4-9.

Crone, D. A., & Horner, R. H. (2003). Building positive behavior support systems in schools. New York: Guildford Press.

Cullinan, D., & Epstein, M. H. (2013). Development, Reliability, and Construct Validity of the Emotional and Behavioral Screener. *Preventing School Failure*, *57*(4), 223-230.

Daniel B. Hall, a. (2000). Zero-Inflated Poisson and Binomial Regression with Random Effects: A Case Study. *Biometrics*, (4), 1030.

Doll, B., Nastasi, B. K., Cornell, L., & Song, S. Y. (2017). School-Based Mental Health Services: Definitions and Models of Effective Practice. *Journal Of Applied School Psychology*, *33*(3), 179-194. doi:10.1080/15377903.2017.1317143

Dowdy, E., Harrell-Williams, L., Dever, B. V., Furlong, M. J., Moore, S., Raines, T., & Kamphaus, R. W. (2016). Predictive Validity of a Student Self-Report Screener of Behavioral and Emotional Risk in an Urban High School. *School Psychology Review*, *45*(4), 458-476.

Drummond, T. (1994). The Student Risk Screening Scale (SRSS). Grants Pass, OR: Josephine County Mental Health Program.

Eklund, K., Kilgus, S., von der Embse, N., Beardmore, M., & Tanner, N. (2017). Use of universal screening scores to predict distal academic and behavioral outcomes: A multilevel approach. Psychological Assessment, 29(5), 486-499. doi:10.1037/pas0000355

Esin, İ., Dursun, O., Acemoğlu, H., & Baykara, B. (2015). Psychopathology and Related Psychosocial Factors in Children with Office Discipline Referrals at School: Evidence from a Developing Country. Community Mental Health Journal, 51(7), 867-878. doi:10.1007/s10597-015-9861-9

Flannery, K. B., Fenning, P., Kato, M. M., & McIntosh, K. (2014). Effects of school-wide positive behavioral interventions and supports and fidelity of implementation on problem behavior in high schools. School Psychology Quarterly, 29(2), 111.

Gall, G., Pagano, M., Desmond, M., Perrin, J., & Murphy, J. (2000). Utility of psychosocial screening at a school-based health center. *Journal Of School Health*, *70*(7), 292-298. doi:10.1111/j.1746-1561.2000.tb07254.x

Girvan, E. J., Gion, C., McIntosh, K., & Smolkowski, K. (2017). The relative contribution of subjective office referrals to racial disproportionality in school discipline. *School Psychology Quarterly*, *32*(3), 392-404. doi:10.1037/spq0000178

Glover, T. A., & Albers, C. A. (2007). Considerations for evalu- ating universal screening instruments. *Journal of School Psy- chology*, *45*, 117–135.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. Journal of Child Psychology and Psychiatry, 38, 581–586. doi.org/10.1111/j.1469-7610.1997.tb01545.x

Goossens, F. X., Lammers, J., Onrust, S. A., Conrod, P. J., de Castro, B. O., & Monshouwer, K. (2016). Effectiveness of a brief school-based intervention on depression, anxiety, hyperactivity, and delinquency: a cluster randomized controlled trial. European Child & Adolescent Psychiatry, 25(6), 639-648. doi:10.1007/s00787-015-0781-6

Gresham, F. M. (2002). Teaching social skills to high-risk children and youth: Preventative and remedial strategies. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Pre- ventive and remedial approaches.* Bethesda, MD: National Association of School Psychologists.

Hartman, K., Gresham, F. M., & Byrd, S. (2017). Student Internalizing and Externalizing Behavior Screeners: Evidence for Reliability, Validity, and Usability in Elementary Schools. Behavioral Disorders, 42(3), 108. doi:10.1177/0198742916688656

Hawken, L. S., & Horner, R. H. (2003). Evaluation of a targeted group intervention within a school-wide system of behavior support. *Journal of Be- havioral Education, 12,* 225–240.

Hilbe, J. (2011). *Negative binomial regression.* Cambridge: Cambridge University Press.

Individuals with Disability Education Act Amendments of 1997 [IDEA]. (1997). Retrieved from https://www.congress.gov/105/plaws/publ17/PLAW- 105publ17.pdf

Individuals with Disabilities Education Act, 20 U.S.C. § 1400 (2004)

Ikeda, M. J., Grimes, J. P., Tilly, W. D., Allison, S., Kurns, S., & Stumme, J. (2002). Implementing an intervention-based approach to service delivery: A case example. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Inter- ventions for academic and behavioral problems II: Preventive and remedial approaches* (pp. 53–69). Bethesda, MD: National Association of School Psychologists.

Introduction to SAS. UCLA: Statistical Consulting Group. From https://stats.idre.ucla.edu/sas/modules/sas-learning-moduleintroduction-to-the-features-of-sas/ (accessed April 22, 2018).

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 school-wide behavioral interven tions. *Journal of Positive Behavior Interventions, 6,* 131-147.

Irvin, L., Horner, R., Ingram, K., Todd, A., Sugai, G., Sampson, N., & Boland, J. (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*(1), 10-23.

Jones, D. E.. Dodge, K. A., Foster, E. M., Nix, R., & Conduct Prevention Research Group. (2002). Early identification of children at risk for costly mental health service use. *Prevention Science, 3,* 247-256.

Jenkins, L. N., Demaray, M. K., Wren, N. S., Secord, S. M., Lyell, K. M., Magers, A. M., & Tennant, J. (2014). A Critical Review of Five Commonly Used Social-Emotional and Behavioral Screeners for Elementary or Secondary Schools. *Contemporary School Psychology*, *18*(4), 241-254.

Kamphaus, R. W., DiStefano, C., Dowdy, E., Eklund, K., & Dunn, A. R. (2010). Determining the Presence of a Problem: Comparing Two Approaches for Detecting Youth Behavioral Risk. *School Psychology Review*, *39*(3), 395-407

Kamphaus, R. W., & Reynolds, C. R. (2007). BASC-2 Behavioral and Emotional Screening System manual. Bloomington: Pearson.

Keller, F., & Langmeyer, A. N. (2017). An Item Response Theory Analysis of the Strength and Difficulties Questionnaire (SDQ). European Journal Of Psychological Assessment, doi:10.1027/1015-5759/a000390

Kame'enui, E. J., & Carnine, D. W. (1998). *Effective teaching strategies that accommodate diverse learners.* Des Moines, IA: Prentice Hall.

Kaminski, R. A., & Good, R. H. (1998). Assessing early literacy skills in a problem-solving model: Dynamic Indicators of Basic Early Literacy Skills. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 90–129). New York: Guilford Press.

Kitchen, C. M. R. (2009). Nonparametric versus parametric tests of location in biomedical research. *American Journal of Ophthalmology*, *147*(4), 571–572. http://doi.org/10.1016/j.ajo.2008.06.031

Lambert, D. (1992). Zero-Inflated Poisson regression, with an application to defects in manufacturing. *Technometrics, 34*(1), 1-17.

Lane, K. L., Kalberg, J. R., Parks, R. J., & Carter, E. W. (2008). Student Risk Screening Scale: Initial Evidence for Score Reliability and Validity at the High School Level. Journal Of Emotional And Behavioral Disorders, 16(3), 178-190.

Lewis, T. J., & Sugai, G. (1999). Effective behavior support: A systems ap- proach to proactive schoolwide management. *Focus on Exceptional Children, 31,* 1–24.

Levitt, J. M., Saka, N., Hunter Romanelli, L., & Hoagwood, K. (2007). Early identification of mental health problems in schools: The status of instrumentation. Journal of School Psychology, 45, 163–191.

Loeys, T., Moerkerke, B., De Smet, O., & Buysse, A. (2012). The analysis of zero-inflated count data: Beyond zero-inflated Poisson regression. *British Journal Of Mathematical & Statistical Psychology*, *65*(1), 163-180.

Martinez, A., McMahon, S. D., & Treger, S. (2016). Individual- and School-Level Predictors of Student Office Disciplinary Referrals. Journal Of Emotional & Behavioral Disorders, 24(1), 30-41. doi:10.1177/1063426615588289

McIntosh, K., Campbell, A. L., Carter, D. R., & Zumbo, B. D. (2009). Concurrent Validity of Office Discipline Referrals and Cut Points Used in Schoolwide Positive Behavior Support. Behavioral Disorders, 34(2), 100.

McIntosh, K., Campbell, A. L., Carter, D. R., & Dickey, C. R. (2009). Differential Effects of a Tier Two Behavior Intervention Based on Function of Problem Behavior. *Journal Of Positive Behavior Interventions*, *11*(2), 82-93.

McIntosh, K., Chard, D., Boland, J., & Horner, R. (2006). Demonstration of Combined Efforts in School-Wide Academic and Behavioral Systems and Incidence of Reading and Behavior Challenges in Early Elementary Grades. *Journal Of Positive Behavior Interventions*, *8*(3), 146-154.

McIntosh. K,, Flannery, K. B., Sugai, G., Braun, D., & Cochrane, K. L. (2008). Relationships between aca demics and problem behavior in the transition from middle school to high school. *Journal of Positive Be havior Interventions, 10,* 243-255.

McIntosh, K., Frank, J. L., & Spaulding, S. A. (2010). Establishing Research-Based Trajectories of Office Discipline Referrals for Individual Students. School Psychology Review, 39(3), 380.

McKown, C. (2017). Social-Emotional Assessment, Performance, and Standards. Future Of Children, 27(1), 157-178.

McKown, C., Russo-Ponsaran, N. M., Allen, A., Johnson, J. K., & Warren-Khot, H. K. (2016). Social-Emotional Factors and Academic Outcomes Among Elementary-Aged Children. Infant & Child Development, 25(2), 119-136.

Miller, F. G., Cohen, D., Chafouleas, S. M., Riley-Tillman, T. C., Welsh, M. E., & Fabiano, G. A. (2015). A comparison of measures to screen for social, emotional, and behavioral risk. School Psychology Quarterly, 30(2), 184-196. doi:10.1037/spq0000085

Now is the time : the President's plan to protect our children and our communities by reducing gun violence. (2013). Washington, D.C. : The White House, 2013.

Ortuño-Sierra, J., Chocarro, E., Fonseca-Pedrero, E., Riba, S. i., & Muñiz, J. (2015). Original article: The assessment of emotional and Behavioural problems: Internal structure of The Strengths and Difficulties Questionnaire. International Journal Of Clinical And Health Psychology, 15265-273. doi:10.1016/j.ijchp.2015.05.005

Owens, J. S., Storer, J., Holdaway, A. S., Serrano, V. J., Watabe, Y., Himawan, L. K., & Andrews, N. (2015). Screening for Social, Emotional, and Behavioral Problems at Kindergarten Entry: Utility and Incremental Validity of Parent Report. *School Psychology Review*, *44*(1), 21-40.

Pas, E. T., Bradshaw, C. P., & Mitchell, M. M. (2011). Examining the Validity of Office Discipline Referrals as an Indicator of Student Behavior Problems. Psychology In The Schools, 48(6), 541-555.

Polgar, M., Cabassa, L., Morrissey, J., Polgar, M. F., Cabassa, L. J., & Morrissey, J. P. (2016). How Community Organizations Promote Continuity of Care for Young People with Mental Health Problems. *Journal Of Behavioral Health Services & Research*, *43*(2), 200-213. doi:10.1007/s11414-014-9409-8

Predy, L. l., McIntosh, K., Frank, J. L., & Flitchcock, J. (2014). Utility of Number and Type of Office Discipline Referrals in Predicting Chronic Problem Behavior in Middle Schools. School Psychology Review, 43(4), 472-489.

Rimvall, M., Elberling, H., Rask, C., Helenius, D., Skovgaard, A., & Jeppesen, P. (2014). Predicting ADHD in school age when using the Strengths and Difficulties Questionnaire in preschool age: a longitudinal general population study, CCC2000. European Child & Adolescent Psychiatry, 23(11), 1051-1060. doi:10.1007/s00787-014-0546-7

Rusby, J. C., Taylor, T. K., & Foster, E. M. (2007). A Descriptive Study of School Discipline Referrals in First Grade. Psychology In The Schools, 44(4), 333-350. doi:10.1002/pits.20226

Sprague, J. R., Sugai, G., Horner, R. H., & Walker, H. M. (1999). Using office discipline referral data to evaluate school-wide discipline and violence prevention interventions. *Oregon School Study Council Bulletin, 42*(2). Eugene, OR: University of Oregon, College of Education.

Sugai, G., Horner, R. H., & Gresham, F. M. (2002). Behaviorally effective school environments. In M. R. Shinn, H. M. Walker, & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 315–350). Bethesda, MD: National Association of School Psychologists.

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 & Behavioral Disorders, 8(2), 94.

Tobin, T. J., & Sugai, G. M. (1999). Using sixth-grade school records to predict school violence, chronic dis cipline problems, and high school outcomes. *Journal of Emotional and Behavioral Disorders, 7,* 40-53. doi:10.1177/106342669900700105

Tobin, T., Sugai, G., & Colvin, G. (1996). Patterns in middle school discipline records. *Journal of Emotional and Behavioral Disorders, 4,* 82–94.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruc- tion as a means of identifying students with reading/learning disabilities. *Exceptional Children, 69,* 391–409.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, S., Chen, R., et al. (1996). Cognitive profiles of difficult-to-remediate and readily reme- diated poor readers: Early intervention as a vehicle for distinguishing be- tween cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88,* 601–638.

Vives, J., Losilla, J., & Rodrigo, M. (2006). COUNT DATA IN PSYCHOLOGICAL APPLIED RESEARCH. *Psychological Reports*, *98*(3), 821-835. doi:10.2466/PR0.98.3.821-835

Walker, H. M., Horner, R. H., Sugai, G., Bullis, M., Sprague, J. R., Bricker, D., et al. (1996). Integrated approaches to preventing antisocial behavior pat- terns among school-age children and youth. *Journal of Emotional and Be- havioral Disorders, 4,* 194–209.

Zeileis, A., Kleiber, C., & Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software, 27*(8), 1-25.