HHS Public Access

Predicting School Suspension Risk from Eighth through Tenth Grade Using the Strengths and Difficulties Questionnaire

Thomas J. Gross,

Western Kentucky University, Psychology Department, 3045 Gary A. Ransdell Hall, 1906 College Heights Blvd., #21030, Bowling Green, KY 42101-1030

Jenna Duncan,

Lipscomb University, College of Education, One University Park Drive, Nashville, TN 37204

Samuel Y. Kim,

Texas Woman's University, Department of Psychology and Philosophy, CFO 807B, P.O. Box 425470, Denton, TX 76204-5470

W. Alex Mason,

National Research Institute for Child and Family Studies, Boys Town, NE, 14100 Crawford Street, Boys Town, NE 68010

Kevin P. Haggerty

Social Development Research Group, University of Washington, Seattle, UW Box #358734, 9725 Third Ave NE, Suite #401, Seattle WA 98115

Abstract

The current study examined (1) if the Strengths and Difficulties Questionnaire (SDQ) would yield alternative factor structures related to either symptoms or strengths with early adolescent students when an exploratory factor analysis (EFA) is used; (2) which scales best predicted suspensions of typically developing early adolescents; and (3) what cut-off scores were useful for identifying youth at risk for suspensions. The current study included 321 parent-student dyads, who were followed from the middle of eighth grade until the end of tenth grade. A symptoms-based EFA yielded three factors: Misbehavior, Isolation, and Agitation. A strength-based EFA yielded three factors, as, well: Emotional, Social, and Moral competence. Logistic regression path analyses were

used to predict risk of any suspension at the end of eighth, ninth, and tenth grades. The predictor variables were the original SDQ Conduct Problems and Hyperactivity scales in one model, the Misbehavior and Agitation scales in a second model, and the Emotional and Moral competence scales in the third model. Only the Misbehavior scale consistently predicted suspensions across each grade ($b = .27$, OR = 1.32, $p < .001$; $b = .15$, OR = 1.18, $p = .029$; $b = .17$, OR = 1.18, $p = .029$, respectively). For the Misbehavior scale, cut-off scores were established that reflected the 75th and 90th percentile; however, each cut-off demonstrated strengths and weaknesses for identifying at-risk students. The expectation of screening to identify youth at-risk for suspensions, a complex school discipline decision, is discussed.

Keywords

school suspension; adolescence; behavioral screening; exploratory factor analysis; parent report

School suspensions are the temporary removal of a child from school to punish violations of school rules (Gibson, Haight, & Kayama, 2014). According to the U.S Department of Education (2015), over three million students are suspended from school each year. A broad range of negative academic outcomes are associated with school suspensions (Brooks, Schriraldi, & Ziedenberg, 2000), some include creating a school climate that is rejecting and harsh toward students (Christle, Jolivette, & Nelson, 2007), higher dropout rates (Suh & Suh, 2007), and the "school-to-prison-pipeline" that links suspensions to drop out and eventual criminal conduct (Losen & Skiba, 2010; Fabelo et al., 2011). Suspensions by the tenth grade are estimated to cost U.S. taxpayers $11 billion (e.g., health, welfare, and crime) and accrue a societal cost of $35 billion (e.g., lost wages and productivity, and expenditures on health care; Rumberger & Losen, 2016). Moreover, unsystematic or decontextualized implementation of suspensions might limit educational opportunities necessary for work skills (Suh, Suh, & Houston, 2007) and increase antisocial behaviors in the future (Parker, Paget, Ford, & Gwernan-Jones, 2016). Identifying risk or protective factors for students could lead to preventive measures to reduce suspensions and potentially their adverse consequences.

Contextual Factors Related to Suspensions

It might be that individual risk and protective factors are related to suspensions, but it is important to clarify that suspensions occur within dynamic contexts. These dynamic contexts can be examined using the ecological transactional theory, which posits that domains, such as social settings, caregiver and peer relationships, and intra-individual context, influence youth bonding to social norms and decision making regarding pro-social or misconduct behaviors (Liu, 2004). It is likely that the interactions of these domains impact adolescents' attention to caregivers' standards, trust in their environment, and decisions to victimize others (Flouri, Midouhas, Joshi, & Tzavidis, 2015; Lynch & Cicchetti, 2002; Matjasko, Needham, Grunden, & Farb, 2010). In light of this, contexts such as school climate, issues related to racial disparity, socialization influences from family life, and individual dispositions could be included in understanding suspensions in schools. The range

of domains influencing suspensions make it critical for educators to account for the various areas and discern which can be impacted within schools.

Admittedly, these contextual issues are typically systemic and difficult to directly assess or efficiently address in consideration of one another within schools or school systems (Cohen, McCabe, Michelli, & Pickeral, 2009). For instance, the use of zero tolerance policies in schools tends to increase suspensions, especially for youth identified as Black/African American (Curran, 2016; Heilbrun, Cornell, & Lovegrove, 2015). However, some evidence indicates that increased professional development in classroom behavior management skills for teachers is related to decreased suspensions over time (Flynn, Lissy, Alicea, Tazartes, & McKay, 2016).

In addition to the difficulty of assessing and addressing multiple domains at school, there are potential family contributions to suspensions risk that are similar across racial groups, such as disruption to the home ($OR_{Black} = 1.34$, $p < .001$; $OR_{White} = 1.31$, $p < .001$; Ganao, Silvestre, & Glenn, 2013). However, protection from suspensions might be related to parental participation in a student's education. For example, parent involvement in school activities, checking homework, and parent volunteering have predicted decreased likelihood of receiving a suspension (Neymotin, 2014). As promising as it may be to provide comprehensive family services to all students, it may overburden the limited resources within a school to do so. It might be more practical to engage in practices that have been identified as effective for reducing behaviors associated with negative school outcomes, such as suspensions. One such practice might include monitoring behavior within a school-wide support system (Dodge, Dishion, & Lansford, 2006). This does not focus on suspensions directly, but does attend to behavioral symptoms and strengths (e.g., Reinke, et al., 2017), which in turn could be related to suspensions.

Student Specific Context and Suspensions

Generally, suspensions are associated with an increased risk of negative behaviors in school settings, such as the use of foul language, defiance, disrespectful behavior, threatening others, fighting, destroying property, and bullying (Skiba et al., 2014), as well as for drug and weapon possession (Skiba et al., 2011). In addition to these negative behaviors, students who exhibit violent behaviors are more likely to be suspended, as well as continue to engage in violent behavior at school after receiving suspensions (Breunlin, Bryant-Edwards, Hetherington, & Cimmarusti, 2002).

However, less severe infractions are more frequently the reason for suspensions (Skiba et al., 2014). A pattern of suspensions due to nonviolent behaviors such as attendance problems (Richard, Brooks, & Soler, 2003) and classroom disruptions (Skiba et al., 1997) have been documented. Other data from The National Center for Education Statistics (2012) has found that nonviolent behaviors such as excessive tardiness and absenteeism are more often the reason for a suspension than more severe conduct problems, such as violent behavior. Whereas more severe infractions are frequently associated with poor impulse control (Pratt & Cullen, 2000), over activity might be implicated in less severe infractions (Thapar et al., 2001). Impulsive and over active students are more likely to violate social norms at school

(Fletcher & Wolfe, 2009) which can lead to eventual aggression at school (Sasser, Kalvin, & Bierman, 2016). Further, these difficulties might be alarming to others and reduce a general sense of security and school engagement (De Voe et al., 2004).

Problem behaviors associated with suspensions might be attributed to skill deficits, rather than an inherent flaw within the youth. A positive youth development (PYD) perspective could be informative in understanding skill development needs related to suspensions. Within a PYD framework, the assessment of five competency areas to promote protective factors, rather than solely relieving symptoms is endorsed (Catalano, Berglund, Ryan, Lonczak, & Hawkins, 2004). *Social competence* is the demonstration of a range of interpersonal skills, *emotional competence* is ability to identify and skillfully manage emotional reactions, and *cognitive competence* is problem solving in social situations, and using abstraction and inductive and deductive strategies to solve academic problems (Catalano et al., 2004). Additionally, *behavioral competence* is the effect use of verbal and nonverbal behavior to help oneself or others meet needs; whereas, *moral competence* relates to having skills for empathizing with others and making decisions related to society's standards of right and wrong (Catalano et al., 2004).

Due to the focus on behavioral problems, monitoring conduct problems and hyperactivity symptoms, or overall externalizing behaviors could be useful for schools to assess suspension risk. However, this might perpetuate the view that these problems are internal to the student or independent of the broader context in which they occur. Alternatively, conceptualizing conduct problems as low levels of behavioral or moral competence could be useful. That is, emphasizing the competence development aspect could help reframe the problems as target areas for skill acquisition, fluency, or generalization instruction. For instance, rather than decreasing classroom disruptions, one could focus on developing goal setting for completing schoolwork.

Universal Screening

One potential way to assess risk of suspensions-related student behaviors is through universal screening within schools' behavioral response to intervention (RTI) systems. Behavioral RTI is a multi-tiered method for the identification of emotional or behavioral problems in students. The first tier of RTI begins with all students receiving scientifically-based curriculum in the general education setting, as well as universal screening to identify students at risk of behavioral skills deficits (Hughes & Dexter, 2015). In the second tier of RTI, students typically receive small group or low intensity individual interventions to address behavior skill deficits. At the third tier of RTI, students receive highly individualized and intensive research based instruction for the student's needs (Crawford & Ketterlin-Geller, 2008).

Universal screening is the brief systematic assessment of students to determine skill development and identify students at-risk for difficulties (Hughes & Dexter, 2015), as well as to determine which students would benefit from available services (Berkeley, Bender, Peaster, & Saunders, 2009). Universal screening has a role to decrease the use of suspensions, as it could address intra-individual strengths and growth areas in a data-

informed manner. That is, it can be one piece of a complex picture that helps educators target at-risk youth to receive skill development before more stringent consequences are enacted. For instance, within the RTI system, students are transitioned into tier 2 services if they are identified as at-risk through universal screening and other evidence corroborates specific skills deficits (Hughes & Dexter, 2015). Additionally, there is emerging evidence to support the use of behavioral screeners to predict risk of suspension due to behavioral problems exhibited by a student (Kettler, Glover, Albers, Feeney-Kettler, 2014). Behavioral screeners that may be used to identify students at risk for suspensions include the Behavior Assessment System of Children- Second Edition (BASC-2), The Behavioral and Emotional Screening System (BESS), Systematic Screening for Behavior Disorders (SSBD), Social, Academic, and Emotional Behavior Risk Screener (SAEBRS), and the Strengths and Difficulties Questionnaire (SDQ).

The BASC-2 includes parent, teacher, and youth self-report rating scales across multiple areas of behavioral symptoms and adaptive skills (Reynolds & Kamphaus, 2004). The BASC-2 is psychometrically sound, but it has between 64 and 185 items depending on reporter form and student age (Reynolds & Kamphaus, 2004). This could make it cumbersome for routine administration. The BESS is an extension of the BASC-2 and it has 30 items that may be well suited to determine externalizing, internalizing, or overall problems (Kamphaus & Reynolds, 2007). However, it could be expensive, as it requires purchasing materials and elements such as a subscription to the AIMSweb behavioral screening service (Kamphaus & Reynolds, 2007). The SBBD is a low cost tool, but has three stages that can be time consuming for school personnel (Walker, Severson, & Feil, 2014). Another screening tool is the SAEBRS, which is a teacher rating form that was created to assess students' behavioral, emotional, academic, and total behavioral skills, and discriminate at-risk and not at risk students (Kilgus, Chafouleas, & Riley-Tillman, 2013; Kilgus, Eklund, von der Embse, Taylor & Sims, 2016; Kilgus, Sims, von der Embse, Riley-Tillman, 2015). While the research for the SAEBRS is promising and frames student behaviors positively, it is relatively new, is included with a for-purchase service, and more research is needed to support its utility.

The SDQ is a 25-item screener that is psychometrically well-established. Its psychometrics have been extensively assessed and it is valid for use with all grade levels in a K-12 setting (Lane, Parks, Kalberg, & Carter, 2007). The items are used to calculate five scales: (a) conduct problems, (b) hyperactivity/inattention, (c) emotional symptoms, (d) peer problems, and (e) prosocial behaviors. Some advantages of the SDQ are that it is a brief questionnaire, interferes minimally with instructional time, can be easily accessed on the internet, pre and post tests can be used to determine progress, and it is free under certain circumstance, including paper-and-pencil forms given to caregivers without charge ("*Information for researchers and professionals*", n.d.; Kamphaus, DiStefano, Dowdy, Elklund, & Dunn, 2010).

However, it is indicated that for non-identified youth, or typically-developing youth, broader measures of externalizing and internalizing on the SDQ might function better for identifying elevated risk levels than individual scales (Goodman, Lamping, & Ploubidis, 2010). The parent report SDQ demonstrated the predictive value of the internalizing and externalizing

subscales to identify children at risk for developing disruptive behavior disorders. Subsequently, researchers have confirmed the factorial validity of the internalizing and externalizing second order factors (Niclasen, Skovgaard, Anderson, Somhovd, & Obel, 2012). Overall, the SDQ measure of externalizing problems may be a more efficient and psychometrically sound screener than the individual conduct problems and hyperactivity/ inattention scales for typically developing youth.

Nonetheless, studies that have employed exploratory factor analysis (EFA) have supported three first order factors for the SDQ that include internalizing, externalizing, and either social or prosocial factors; these are different from the published scales (Azzopardi, Camilleri, Sammut, & Cefai. 2016; Dickey & Blumberg, 2004; Gómez-Beneyto et al., 2013). For instance, it has been reported that adolescent behavioral and emotional difficulties are related to constructs of distress associated with interpersonal problems, fear, and externalizing problems (Doyle, Murphy, & Shevlin, 2016). The SDQ studies that employed EFA used data from children 4 to 17 years of age; therefore, factorial models of the SDQ should be explored at discrete developmental periods, such as early adolescence, due to well-documented developmental differences in the expression of emotional and behavioral problems (Petersen et al., 2015; Reef et al., 2009).

Further, the SDQ has 20 difficulties items and five strengths items; this distribution of items is arguably symptoms-based. Symptoms-based factor models that use the items as designed might over emphasize an intra-youth symptoms focus. However, if all items are scored so that they represent strengths, or acquired skills, then the strengths-based factors could represent levels of competency. Because of the recoding, there might also be alternative combinations of items, in that the strengths and difficulties might exist across independent continuums, rather than as polar opposites. This might be more useful for directing attention to social and emotional needs planning. In any event, the SDQ factors extracted from EFA could be used to predict which early adolescent youth are at risk for suspensions. It would be useful to determine if a symptoms-based model predicts suspensions differently than a strengths-based model. Moreover, it might be practical for schools to know if the EFA extracted factors primarily focused on symptoms or solely focused on strengths are better predictors of suspensions over the original conduct problems and hyperactivity scales during early adolescence.

Current Study

The purpose of the current study is to examine (1) if the SDQ would yield an alternative to the original factor structure with early adolescent students when EFA is used for symptoms-based versus solely strength-based items; (2) which SDQ scales or factors best predict suspensions of typically developing early adolescents; and (3) what cut-off scores could be useful for identifying which youth are at risk for suspensions. To accomplish this, two EFAs were completed with the parent report SDQ from the middle of eighth grade. The factors extracted from the EFA were used to calculate parent report SDQ scales at the middle of eighth grade, and end of eighth and ninth grade. The original parent report SDQ conduct problems and hyperactivity/inattention scales were used to predict the likelihood of receiving any suspensions at the end of eighth, ninth, and tenth grade in one logistic

regression analysis. Similarly, the EFA factors were used to predict suspensions at the end of eighth, ninth, and tenth grade. Separate logistic regression were used for the symptoms-based and strengths-based factors. Further, cut-scores based on percentile rank were used to examine discrimination between those who received any suspensions or no suspensions at the end of eighth, ninth, and tenth grade, respectively.

Method

Participants

Participants included 321 low-income parent-child dyads, who were enrolled in a randomized controlled trial (RCT) of two parenting programs, Common Sense Parenting (CSP; Burke, Schucman, & Barnes, 2006) and CSP*Plus*, which included all elements of CSP and two parent-child sessions regarding transitioning to high school. Participants were assigned to one of three conditions in the RCT: CSP ($n = 118$), CSP*Plus* ($n = 95$), or a control condition ($n = 108$). The parenting program was designed as a prevention program; therefore, all youth participants were recruited from the general education setting. Each family included a target parent and target eighth grader, who attended one of five low-performing middle schools in the urban Pacific Northwest. Approximately 70% of all the students at each of the five schools received free or reduced-price school lunch. Data were collected from the participants at four time points: middle of eight grade (baseline), end of eighth grade (post-test), end of ninth grade (one-year follow-up from post-test), and end of tenth grade (two-year follow-up from post-test). The retention rate at the end of tenth grade was 94% for parents ($n = 303$) and children ($n = 301$), and was similar across each time point for each condition. Moreover, completion rates were similar across race/ethnicity of youth or parents, age of parent, marital status of parent, and whether families had an annual household income above or below $24,000. Parent and youth demographic data are included in Table 1 and Table 2, respectively. Details regarding recruitment and study procedures can be found in Mason et al. (2016).

Measures

School Suspensions.—Suspensions data was collected for the academic year at the end of eighth, ninth, and tenth grade through youth self-report. In eighth grade, students were asked, "During the past 12 months, how many times have you been suspended from school for disciplinary reasons?" At subsequent follow-up survey, youth were asked, "Since the last interview, how many times have you been suspended from school for disciplinary reasons?" The suspensions data was coded *no suspensions* = 0 and *any suspensions* = 1. For this variable and all other variables, descriptive statistics are provided in the results section.

Strength and Difficulties Questionnaire (SDQ).—Parent report SDQ data was collected at middle of eight grade, end of eighth grade, end of ninth grade, and end of tenth grade. However, the SDQ data from the middle of eighth grade, end of eighth grade, and end of ninth grade were used for the predictor variables. All 25 SDQ items were given to participating parents at each data collection. While the Conduct Problems scale and Hyperactivity Symptoms scales were of primary interest for this study, items from the

Emotional Symptoms, Peer Problems, and Prosocial scales were also administered in the context of the RCT.

The Conduct Problems scale (CP) and Hyperactivity Symptoms scales (HS; Goodman, 1997) each included 5 items, and each were rated by parents on a 3-point Likert-like scale (0 = *not true*; 1 = *somewhat true*; 2 = *certainly true*) for a given item over the last six months, and summed to give a respective scale scores between 0 and 10. A higher score indicates increased problems. Sample CP items are "Often fights with other youth or bullies them" and, "Often lies or cheats;" and sample HS items are "Restless, overactive, cannot stay still for long" and, "Easily distracted, concentration wanders." The Emotional Symptoms, Peer Problems, and Prosocial scales had the same item rating scale and scoring procedures. Emotional Symptoms items included "Often unhappy, downhearted" and "Many fears, easily scared." Peer Problems items have items, such as "Rather solitary, tends to play alone" and "Picked on or bullied by other children." Also, the Prosocial items include "Considerate of other people's feelings" and "Often volunteers to help others."

In order to compare models, items were scored in two separate manners. First, the items were scored as intended by the SDQ developers, where some items represented symptoms or negative behavioral attributes, and other items represented strengths or positive interpersonal attributes. The same items were also scored so all represented levels of strengths, or positive behavioral attributes. For example, the item, "Nervous in new situations, easily loses confidence," was recoded positively so that it represented "confident in new situations."

Covariates.—Student gender and race were used as covariates. Youth were asked to identify if they were a girl or boy, and responses were coded *girl* = 0 and *boy* = 1. This was selected as a covariate because boys are more likely to be suspended than girls (Mendez & Knoff, 2003; Office of Civil Rights [OCR], 2016). Youth responded to the race/ethnicity item, "What best describes your racial background?" They indicated one of the following responses: (a) White, (b) Caucasian or European not Hispanic, (c) Black or African American, (d) Asian or Asian American, (e) Hispanic, Spanish or Latino, (f) American Indian or Alaskan Native, or (g) Native Hawaiian or other Pacific Islander. They also had the opportunity to provide no response; however all students provided a response. Students were divided into two race groups for this study coded *non-African American* = 0 (i.e., any response option other than Black or African American) and *Black or African American* = 1. This was selected as a covariate because African American youth are more likely to be suspended than non-African American youth (Heilbrun, Cornell, & Lovegrove, 2015; OCR, 2016). Students were also divided into two ethnicity categories *non-Hispanic* = 0 (i.e., any response option other than Hispanic, Spanish or Latino) and *Hispanic, Spanish or Latino* = 1. This was selected as a covariate because there is evidence that Hispanic youth are suspended at comparably lower rates for persons or color (OCR, 2016).

Analyses

A chi-squared analysis indicated non-significant differences between experimental conditions (i.e., CSP, CSP*Plus*, control) on suspensions in eighth grade, χ^2 (n = 321, df = 2) = 3.17, p = .205, ninth grade, χ^2 (n = 246, df = 2) = 3.12, p = .210, and tenth grade, χ^2 (n =

284, $df = 2$) = 0.54, $p = .763$. Therefore, analyses were based on the total sample pooled across experimental conditions.

Exploratory factor analysis (EFA).—Two EFAs with Varimax Rotation were completed in Mplus 8.0 (Muthén & Muthén, 1998–2017) using maximum likelihood estimation with robust standard errors (MLR). The first EFA had all items scored as the SDQ authors had specified. The second EFA had all items scored to reflect positive attributes. At the item level, missing data was low (0 to <0 .01%). The EFA analyses was completed with middle of eighth grade SDQ items. One item failed to have acceptable levels of skewness (> |3|) or kurtosis (> |8|; Kline, 2010). The MLR estimator has the advantage of adjusting model fit and standard errors to account for non-normal distributions. Moreover, the use of MLR over other missing data techniques, such as a weighted least squares means estimator (e.g., WLSM) for EFA, is supported since the estimations are consistent with WLSM, but provide robust standard errors in light of non-normal distributions (Muthén & Muthén, 1998–2017; Yuan & Bentler, 2000). Varimax rotation is an orthogonal rotation procedure that constrains factors to be uncorrelated and has the benefit of more clearly associating items with extracted factors (Henson & Roberts, 2006). Moreover, previous studies have established that the factors extracted via EFA on the SDQ are conceptually separate (e.g., internalizing or externalizing).

The EFAs tested models with 1 through 5 factors. Chi-square (χ^2) values were used to determine if there was significant model misfit; however, sample sizes larger than 200 cases could result in poor model fit due to potential inflation of the χ^2 statistic from increased sample size (Bentler & Bonett, 1980). Therefore, other model fit indicators need to be included, and model fit is best assessed when evaluating the model across multiple fit statistics (Vandenberg & Lance, 2000). Model fit was also assessed with the root mean squared error of approximation (RMSEA) and root mean square residual (RMSR), which range from 0 to 1.00. It is recognized that an RMSEA and RMSR less than .08 indicate acceptable model fit (Hooper, Coughlin, & Mullen, 2008). In addition, EFA results should be viewed in light of factor eigenvalues, scree plot inspection, inspecting if there are any negative residual items, and the number of items for each factor, and the each factor's interpretability. Further, Eigen values greater than 1.00 for factors are considered acceptable and the scree plot should be used to visually inspect where a leveling-off of the points occurs. Typically, models with one or more items that have negative residuals are considered to be invalid. Further, the number of items for a factor should be greater than two, the items should have a discernable and defensible congruence, and practical use when combined (Henson & Roberts, 2006).

Descriptive and preliminary analyses post-EFA.—The factors retained from the EFAs were used to compute summed scores from the respective SDQ items at the middle of eighth, end of eighth, and end of ninth grades. The number and proportion of student suspensions, students' gender and race by eighth, ninth, and tenth grade were calculated. SDQ CP and HS, and the scores for the scales based on the EFAs means and standard deviations were calculated across the middle of eighth, end of eighth, and end of ninth grades. Pearson's r correlations were computed between these variables. Cronbach's alpha

(α) was assessed for the SDQ CP, HS, and scales derived from the EFAs at the middle of eighth, end of eighth, and end of ninth grades.

Logistic Regression Path Analyses.—Logistic regression was selected because it allows for predicting the likelihood of group membership (i.e., suspensions or no suspensions) using both categorical (e.g., gender) and continuous (i.e., SDQ scales' scores) measures. The path analysis approach was selected because the sequential and correlated nature of the variables across time is accounted for. When predictive paths are specified from one time point to the next within a criterion variable, this allows the model to account for previous outcomes of the criterion variable to predict outcomes on the criterion variable. Therefore, it accounts for the variance of prior performance on the criterion variable in addition to estimates made from other predictors.

Prior to modeling the logistic regressions, missing data was assessed. Little's test of missing completely at random (MCAR) was significant, $\chi^2 = 171.07$, $df = 137$, $p < .026$. This indicated that listwise deletion would be inappropriate. Logistic regressions were completed in Mplus 8.0 (Muthén & Muthén, 1998 – 2012) using MLR. It has been demonstrated that while maximum likelihood estimates assume missing at random (MAR) data, they are likely unbiased when data is missing not at random (MNAR; Schafer & Graham, 2002). Mplus 8.0 uses Monte Carlo integration, which assesses the accuracy of the model function in predicting group membership through randomly sampling predictor values from the score distribution until the most probable outcome is estimated (Forster, McDonald, & Smith, 1996).

Three logistic regression models were tested. For each logistic regression, receiving suspensions at the end of eighth grade was set to predict suspension at the end of ninth and tenth grades, and suspension at ninth grade were set to predict tenth grade suspension at the same time. The first logistic regression was modeled so (a) the SDQ CP and HS scales from the middle of eight grade were predictors of suspensions at the end of eighth grade; (b) the SDQ CP and HS scales from the end of eight grade were predictors of suspensions at the end of ninth grade; and (c) the SDQ CP and HS scales from the end of ninth grade were predictors of suspensions at the end of tenth grade. SDQ CP and HS were auto-regressed across time points, and were cross-lagged across each data collection period, meaning these variables were set to predict each other from one time point to the next. SDQ CP and HS scales were specified to have correlated residual variables at the end of ninth grade in the model due to the high intercorrelations between the measures. The second and third logistic regression were similarly modeled with the scales derived from each EFA. That is, (a) the scales from the end of eighth grade were predictors of suspensions at the end of ninth grade; (b) the scales from the end of eighth grade were predictors of suspensions at the end of ninth grade; and (c) the scales from the end of ninth grade were predictors of suspensions at the end of tenth grade. Correspondingly, these scales were auto-regressed and cross-lagged. Because of the high intercorrelations between the measures, the EFA scales were specified to have correlated residual variables at the end of ninth grade in the model. Additionally, gender and race were entered as predictors of suspensions only, across all grades levels in both models.

Results of a logistic regression in Mplus include unstandardized regression coefficients and the respective standard error, and odds ratios. The regression coefficients represent the log odds of changes in group membership for every one unit increase in a specific predictor variable. For example, if a coefficient equals .51 for a predictor and the predictor increases by two units, the number of suspensions increases by 1.02. Additionally, the odds ratio (OR) is the multiplicative change in the odds of group membership for one unit of change in a predictor. An OR of 1.00 indicates that a change in a predictor variable does not impact the odds of suspension over non-suspension. A change in OR represents the proportional change in odds of suspension based on a change in one unit of the predictor; changes that are more than 1.00 mean higher odds, whereas, less than 1.00 mean lower odds. For example, if the OR is 2.00 that means as the predictor increases by one unit, the odds of being suspended double, or increase by 100%.

Chi-square analyses.—A priori cut-off points were established if scales consistently predicted suspensions from the logistic regression path analyses using percentile rank to group students in to low-risk and elevated-risk categories. If the scales in the logistic regression failed to predict suspensions, then the cut-off scores were not used. Two sets of cut-off points were used, 75th percentile and 90th percentile for the original and EFA symptom-based scales, and 25th and 10th percentile for the EFA strength-based scales. While cut-off points related to 1 and 1.5 standard deviations have been cited as useful for assessing symptoms or impairment (Feil et al., 2005), percentile rank was selected because it is more intuitive than standard deviation cut-off points, and percentile rank indicates an expected range that a comparative group of children should be below or above (Wang & Chen, 2012).

First, scores related to quantile grouping and then scores related to decile grouping were examined. The 75th percentile score for the original and symptom-based EFA scales was used for dummy coding, where scores below the 75th percentile were considered low-risk (= 0) and scores at or above the 75th percentile were considered elevated-risk (= 1). The same coding procedure was used for 90th percentile score cut-off, where below the 90th percentile were considered low-risk and above the 90th percentile were considered elevated-risk. Similarly, for the strengths-based EFA scale scores at or below the 25th percentile were considered low-skilled, as well as for scores at or below the 10th percentile.

The chi-square analyses were performed to examine the relation between the scale cut-off score groupings and any suspensions received. The chi-square analysis does not included the previous suspensions data for predicting 9th and 10th grade suspensions because including a repeated measure over time in a chi-square analysis violates the assumptions of the statistic. That is, assessing change over time in participants on the same measure is not allowed (McHugh, 2013). However, the logistic regression provided evidence of the scales predicting suspensions over the variance accounted for by previous suspensions. Chi-square analyses were as follows: (a) middle of eighth grade cut-off score groupings by end of eighth grade suspension; (b) end of eighth grade cut-off score groupings by end of ninth grade suspension; and (c) end of ninth grade cut-off score groupings by end of tenth grade suspension. The models were evaluated using the Pearson's chi-square (χ^2) value and phi coefficient (ϕ) value. The chi-square value is used to determine statistical significance. The phi value provides an effect size for the comparison by assessing the strength of association

between groupings. Phi ranges in value from –1.00 to 1.00, with negative values indicating an inverse relationship and positive values indicating a positive relationship. Phi values can be classified as no relationship (0 to <|.01|), negligible relationship (|.01| to |.19|), weak relationship (|.20| to |.29|), moderate relationship (|.30| to |.39|), strong relationship (|.40| to |.69|), and very strong relationship (|.70|) (Davis, 1971). Comparisons between the effect sizes from percentile grouping were made using the phi coefficients at each grade level. Due to the number of analyses, the significance level was set to $p < .01$. Additionally, sensitivity, specificity, positive predictive value (*ppv*) and negative predictive value (*npv*) were calculated for each score cut-off grouping.

Results

EFA

The SDQ items descriptive statistics, correlation, covariance matrices, and scree plots are available upon request to the first author. Table 3 provides the model fit statistics for both EFA models with one through five factors. In both EFAs, all models had a statistically significant chi-square value ($p < .001$), but models with two through five factors had adequate fit statistics (RMSEA and RMSR < .08). According to an examination of the Eigen values in both EFAs, up to five factors could be acceptable; however, when inspecting the scree plot, leveling off could be inferred after three and then five factors. Parallel across EFA models, the review of factor loadings indicated that (a) a five-factor solution yielded six items with substantial loadings (.30) across factors, which contributed to factors that were difficult to define or assess to a priori templates, (b) a four-factor model with eleven items with substantial loadings across factors, (c) a three-factor model with five items with substantial loading across factors, and (d) a two-factor model that had three items with substantial loading across factors, but factors consisting of items which were difficult to interpret. Additionally, the strengths-based EFA four and five factor solutions had at least on factor with all loading values >.30 having negative values. It was decided to retain the three-factor model for both EFAs because of the adequate model fit, match with the scree plot, factor Eigen values all greater than 1.00, few cross-loaded items, and interpretability of the factors and the fit of a three factor symptoms model with previous research. Across both EFA models, Table 4 contains the results from the three factor models from the symptoms- and strength-based EFAs. Two items, "Shares readily with other youth, for example books, games, food" and "Kind to younger children" did not substantially load on any factor (factor loading < .30). These items were omitted from the interpretation of factors and scales created from the EFAs.

Symptoms-based EFA.—For the symptoms-based EFA model, factor 1 contained eleven items that were positively related to conduct problems and impulse control, and negatively related to being considerate of others. This factor was labeled "Misbehavior" (MB). Factor 2 contained nine items related to worry and social disengagement from peers. It was labeled "Isolation" (ISO). Factor 3 consisted of eight items that connected to movement, distractibility, nervousness and somatic complaints. This factor was labeled "Agitation" (AG).

For the purpose of scale creation, it was decide to omit items that had negative factor loadings, and include items on scales with which they had the highest factor loading if they loaded   .30 on more than one factor. This yielded an MB scale with eight items, and an ISO and an AG scale with six items each. However, to improve internal consistency on the MB scale, the "Often loses temper" item was removed to improve internal consistency (from α = .37 to .75), resulting in a seven item scale. The internal consistency coefficients for the three scales were adequate. See Table 5 for means and standard deviation, and scale Cronbach alphas.

Strength-based EFA.—Factors were named to reflect PYD competence areas the items with which they were most associated. Concerning the strength-based EFA model, factor 1 consisted of eight items that reflected being calm and physically still. This factor was labeled "Emotional competence" (EMC). Factor 2 contained nine items related to confidence in social engagement. It was labeled "Social competence" (SOC). Factor 3 had eleven items related to being considerate of others. This factor was identified as "Moral competence" (MOC).

When considering scale development, we followed the same criteria as for the symptoms-based EFA scales. This resulted in an EMC and SOC scales with six items each, and a MOC scale with eleven items. See Table 6 for means and standard deviation, and scale Cronbach alphas.

Descriptive Statistics Post-EFA for All Study Variables

Table 7 contains the intercorrelations between the developer conduct problem and hyperactivity scales, EFA symptoms-based scales, and EFA strengths-based scales, respectively. Cronbach's alpha was included for each scale at each time point in the table, as well. In general, scales within each set of predictive measures (i.e., original, symptoms EFA, and strengths EFA) were positively and significantly correlated with each other. Table 8 contains the correlations between each set of predictive measures and suspensions, gender, and race; it also contains descriptive statistics for each variable. These correlations were completed to examine which predictors should be included in the logistic regression. The SDQ CP and HS scales were positively associated with suspensions from one time point to the next ($p < .001$). Similarly, the symptoms-based EFA MB ($p < .001$) and AG ($p < .01$) scales had the same associations. Conversely, the strengths-based EFA scales of EMC ($p < .01$ to $p < .05$) and MOC ($p < .001$) were negatively associated with suspensions. The remaining EFA scales were non-significantly associated with suspensions from one time point to the next. Moreover, youth identification as Hispanic was non-significantly associated with suspensions in 8th ($r = -.10$, $p = .075$), 9th ($r = -.09$, $p = .128$), and 10th ($r = -.04$, $p = .539$) grades. The full correlation matrix with associations between all measures in the study is available from the first author by request.

Logistic Regression Path Analyses

See Figure 1 for the path analyses diagram. The results of the logistic regression path analysis using the SDQ CP and HS scales indicated that gender only predicted suspensions in eighth grade (OR = 2.38; $p = .020$), and that identification as African-American predicted

suspensions in eighth (OR = 4.11; $p < .001$) and tenth grade (OR = 3.03; $p = .010$). The conduct problems scale predicted suspension in eighth (OR = 1.48; $p < .001$) and tenth grade (OR = 1.32; $p = .015$), whereas the hyperactivity scale predicted suspension in the ninth grade (OR = 1.27; $p = .007$). Regression path coefficients between the conduct problems and hyperactivity scale are available upon request to the first author. Table 9 contains the estimates, standard errors, and odds ratio for all predictors on suspension.

The second logistic regression path analysis used the misbehavior and agitation scales derived from the EFA. The decision to exclude the isolation scale was based on the limited correspondence between the scale and suspension across grades. Similarly, gender predicted suspension at eighth grade (OR = 2.45; $p = .015$), and race predicted suspensions in eighth (OR = 3.62; $p < .001$) and tenth grade (OR = 2.72; $p = .021$). The misbehavior scale predicted suspension at eighth (OR = 1.32; $p < .001$), ninth (OR = 1.18; $p = .029$), and tenth grade (OR = 1.18; $p = .029$). The agitation scale predictions of suspension across grades were non-significant.

The EFA derived emotional competence and moral competence scales were used in the third logistic regression path analysis. The decision to exclude the social competence scale was based on the limited correspondence between the scale and suspension across grades. In this model, gender predicted suspension at eighth grade (OR = 2.39; $p = .016$), and race predicted suspensions in eighth (OR = 4.09; $p < .001$) and tenth grade (OR = 2.65; $p = .025$). The moral competence scale predicted suspension at eighth grade only (OR = .80; $p < .001$). The emotional competence scale did not predict suspensions at any time point. Table 10 contains the estimates, standard errors, and odds ratio for all predictors on suspension for the second and third logistic regression models.

Chi-square Analyses

Two sets of 2×2 chi-square analyses were completed for the misbehavior scale only, because it was the only scale to consistently predict suspensions across all grades. The 75th percentile scores across eighth (MB = 5), ninth (MB = 4), and tenth (MB = 5) grade were similar, but not the same. However, to keep cut-off points consistent across grade levels it was determined to use MB = 5 as the cut-off score. The 90th percentile scores across eighth, ninth, and tenth grade were the same (MB = 7).

In the first chi-square analysis, the low-risk group had scores < 5 and the elevated-risk group had scores 5 for the middle of eighth grade, end of eighth grade, and end of ninth grade, respectively. The two groups in (a) the middle of eighth grade were used to predict suspension at the end of eighth grade, (b) the end of eighth grade were used to predict suspension at the end of ninth grade, and (c) the end of ninth grade to predict suspension in the end of tenth grade. There were significant associations between being in the elevated-risk group and receiving a suspension in eighth, $\chi^2(1, n = 317) = 19.35$, $p < .001$, $\phi = .25$, ninth, $\chi^2(1, n = 279) = 11.02$, $p = .001$, $\phi = .20$, and tenth, $\chi^2(1, n = 271) = 9.67$, $p = .002$, $\phi = .19$, grades. Table 11 contains the 75th percentile cut-off groups and suspension category results across grades. The elevated-risk cut-off score accurately predicted 58% in eighth, 43% in ninth, and 53% in tenth grade of the youth who were suspended; whereas the low-risk group predicted 74% in eighth, 80% in ninth, and 73% in tenth grade of the youth who were not

suspended. That is the cut-off score of MB = 5 had low sensitivity and adequate specificity. Further, the positive predictive values were low across each time point (28%, 31%, and 24%, respectively), and the negative predictive values were consistent and high at each grade level (91%, 87%, and 90%, respectively).

In the second chi-square analysis, the low-risk group had scores < 7 and the elevated-risk group had scores 7 for the middle of eighth grade, end of eighth grade, and end of ninth grade, respectively. The low-risk and elevated-risk groups were used to predict suspension as in the previous chi-square analysis. There were significantly more youth in the elevated-risk group having been suspended for eighth, χ^2 (1, $n = 317$) = 16.85, $p < .001$, $\phi = .23$, ninth, χ^2(1, $n = 279$) = 12.68, $p < .001$, $\phi = .21$, and tenth, χ^2(1, $n = 271$) = 11.62, $p = .001$, $\phi = .21$, grade. Table 10 contains the 90th percentile cut-off groups and suspension category results across grades. The elevated-risk cut-off score accurately predicted 31% in eighth, 24% in ninth, and 29% in tenth grade of the youth who were suspended; whereas the low-risk group predicted 90% in eighth, 93% in ninth, and 91% in tenth grade of the youth who were not suspended. That is the cut-off score of MB = 7 had low sensitivity and high specificity. Additionally, the positive predictive values were low at each grade level (37%, 41%, and 33%, respectively), and the negative predictive values were high across time points (88%, 85%, and 89%, respectively).

A comparison of the chi-square analyses for all suspensions at all grade levels indicated that both cut-off scores had a weak relation to suspension, which is reflected in the respective cut-off scores' sensitivity coefficients. Moreover, positive predictive values were low for both cut-off scores, but the negative predictive values and specificity coefficients were high. This indicated either cut-off score would be more useful for identifying students who are not-at-risk than those at-risk for suspensions.

There were some differences between the 75th and 90th percentile cut scores by grade. Specifically, the phi coefficient for the 75th percentile cut-off score was greater only for eighth grade. For ninth grade and tenth grade, the phi coefficients for the 90th percentile were slightly greater. This indicated that different cut-off scores for eighth grade versus ninth and tenth grade might be more useful. However, only in tenth grade did the significance level change by cut-off score, where the higher cut-off score had a lower p-value ($p = .002$ vs. .001). However, the similarity in phi coefficients, as well as the changes to sensitivity, specificity, and positive and negative predictive values coefficients followed a very similar pattern.

Discussion

The current study sought to evaluate the SDQ as a universal screening measure that may be useful for identifying students who are at risk for school suspensions based on symptoms-based scales or strengths-based scales. The purpose of this study was to determine if there were alternative factors to the SDQ when used with early adolescents; which factors best predicted suspensions; and what cut-off scores might be helpful to identify who might need extra support to prevent school suspensions. Two EFAs yielded three factors, each. The symptoms-based EFA produced the three scales Misbehavior, Isolation, and Agitation. The

strengths-based EFA produced the three scales Emotional competence, Social competence, and Moral competence.

The Misbehavior scale of the SDQ was found to be the most consistent predictor of school suspension, when compared with the original Conduct Problems and Hyperactivity scales and the three strengths-based scales. Additionally, cut-off scores were examined for the Misbehavior scale because it was the only consistent predictor of suspension. Cut-off criteria at the 75th and 90th percentile rank both resulted in significant associations between elevated scores and suspension; however, the different cut-off scores varied in sensitivity and specificity for predicting suspensions. A higher cut-off score was found to yield a greater proportion of true positives for suspension, but identify less youth who might be at-risk. Conversely, a lower cut-off score identified more students as at risk for suspensions, but at a cost of decreased precision.

Student Difficulties and Strengths Factors

The SDQ is a brief screener that covers some breadth of behavioral symptoms (Goodman et al., 2010); however, multiple studies provide emerging evidence of varying factors and predictions of behavior problems in children (Doyle et al., 2016; Niclasen, et al., 2012). Previous studies used a broad range of school aged children (4 to 17 years of age; Azzopardi, Camilleri, Sammut, & Cefai. 2016; Dickey & Blumberg, 2004; Gómez-Beneyto et al., 2013), but there is evidence for changing problem presentation due to developmental differences in emotional and behavioral constructs in adolescence (Petersen et al., 2015; Reef et al., 2009). Moreover, analyses of the SDQ have yet to be published looking at items as all strengths with the PYD competence framework.

The current study found keeping the items in a primarily symptoms-based format produced three broad factors in this sample, consisting of misbehavior, isolation, and agitation symptoms. The misbehavior factor appears consistent with externalizing factors of other EFA studies of the SDQ (Niclasen, et al., 2012). Nonetheless, the inclusion of inattention and distractibility might better reflect behavioral symptoms that parents find bothersome, or would identify as misbehavior. In addition, instead of an internalizing factor (e.g., Niclasen, et al., 2012) or internalizing and prosocial factors (e.g., Azzopardi, Camilleri, Sammut, & Cefai. 2016; Dickey & Blumberg, 2004; Gómez-Beneyto et al., 2013), the current analysis yielded isolation and agitation factors. In fact, the original pro-social items failed to load onto a distinct factor. While these factors are not an exact match to previous factors, they are consistent with internalizing and social problems.

The emotional and social competency scales had the same items as the agitation and isolation scales. The emotional and social competency scales could be looked at as the strengths equivalents of the agitation and isolation scales, respectively. It is possible that typically developing adolescents experience worry about social engagement and experience difficulty with managing complex and interrelated physical and psychological distress. These factors could be conceptualized as social and coping skills competencies that youth can be taught.

However, the moral competence scale contains additional items when compared to the misbehavior scale. Specifically, the moral competence scale included items related to controlling one's temper, helping someone if they were hurt, and offering to help others. When viewed as a whole the moral competence scale could reflect conscientiousness, as if contains elements related to orderliness, responsibility, industriousness, and self-control (Roberts, Chernyshenko, Stark, & Goldberg, 2005). Moreover, the misbehavior factor did not include three of the items because they had negative factor loadings. It is quite possible that if these items were recoded to reflect social indifference rather than pro-social behavior, the moral competence and misbehavior scales would represent polar ends of the same trait. Nonetheless, the misbehavior scale might be useful for focusing on remediating behavioral deficits, whereas the moral competence scale could help with identifying a larger set of skills to help students become globally successful individuals.

Student Difficulties Predicting Suspensions

Two SDQ scales that are closely related to externalizing problems are the Conduct Problems and the Hyperactivity scales, and theses scales assess behavioral problems (Skiba et al., 2014) and impulsivity (Pratt & Cullen, 2000; Thapar et al., 2001), which have been linked to suspensions. It was found that these scales predicted suspensions inconsistently, but conduct problem and hyperactivity had significant correlations with suspensions across time points (mean $r = .24$, both factors). Nonetheless, it might be that treating these factors as discrete sets of behaviors is inadequate for predicting suspensions through this time period.

The resulting misbehavior scale from the symptoms-based EFA was a combination of conduct problem items ($n = 4$) and hyperactivity items ($n = 3$). It is possible that parents viewed the co-occurrence of these problems as reflecting a global set of problem behaviors. This is consistent with previous studies (e.g., Kettler, Glover, Albers, Feeney-Kettler, 2014), as broad problems are predictive of school suspensions. Moreover, it supports the finding that student engagement in behaviors viewed as problematic by adults lead to suspensions (Breunlin et al., 2002). Contrastingly, agitation was associated with suspensions, but failed to predict suspensions, which may be a result of the significant, yet small associations with suspensions (mean $r = .17$). Further, it may be that some youth who receive suspensions display characteristics, such as somatic complaints, restlessness, and unhappiness; whereas, those who typically experience agitation do not engage in behaviors that lead to suspensions. Agitation could be a frequent concurrent symptom with externalizing (mean $r = .41$) and isolation (mean $r = .36$) problems; whereas externalizing and isolation are less likely to co-occur (mean $r = .23$).

Interestingly, the strengths-based EFA scale moral competence failed to predict suspensions across more than one time point. It was largely correlated with misbehavior at each time point ($r = -.94$, $p < .001$), the two variables had similar associations with suspensions. It could be that the omission of four items on the misbehavior scale have this scale more closely reflect the actions that lead to suspensions. On the other hand, when the items are included, they add elements that are unrelated to the decision to suspend a student. That is, when presented with symptom clusters, school personnel are more likely to suspend students. However, the added items related to helpfulness might have no impact on the

decision to suspend and therefore interfere with the predictive utility of the items relate to following rules and forethought. Overall, it appears that teaching adolescents to follow rules and organize time to complete projects could provide the most protection against suspensions.

Cut-off Scores for Assessing Suspension Risk

The misbehavior scale was the only consistent predictor of suspensions for each grade. Therefore, it was the only scale, which had its cut-off scores at the 75th and 90th percentile investigated. It was found that the misbehavior scale consistently predicted suspensions, but in the chi-square and adjoined analyses, it was found that significant differences in association were not correspondent to high accuracy in identifying those at risk for suspensions. Rather, a greater cut-off score was equivalent to better identification of those at low risk for suspensions in eighth, ninth, and tenth grades (90th percentile sensitivity = 90%, 93%, and 91%, respectively). In fact, positive predictive values did increase with a higher cut score, yet negative predictive values remained somewhat stable. Still, the average increase in positive predictive values (mean ppv = 9%) should not be overlooked, given the severe ramifications for suspending children from school (Breunlin et al., 2002; Rumberger & Losen, 2016). It could likely be that those in the 75th percentile are earmarked for less intensive interventions, pending corroborative data; whereas those identified in the 90th percentile could be considered for more intensive services, pending corroborative information.

Limitations

The findings of the current study have implications for services provided to non-identified students across most schools, but these results are not without notable limitations. In consideration of the outcome variable, suspensions were determined by self-report. In general, youth self-report regarding if discipline issues occurred (i.e., yes/no) are accurate (Thornberry & Krohn, 2000). However, it was not possible to obtain the school records for suspensions, but we believe that student self-report should receive confirmation from an independent source when possible.

While it is desirable to provide a straightforward and effective tool for preventing suspensions, there are likely issues related to generalizability. Both the geographic location and school setting (e.g., level of resources, behavioral RTI consistency, and varying demographics) could interfere with predicting how this would impact a broader range of school in the United States. Though there is evidence to use this assessment as a screener for behavioral problems symptoms, it may not necessarily function as well for predicting suspension risk. Also, elements related to multi-rater screening across various districts would improve our understanding of screening for suspensions. That is, the use of parent, teacher, and student ratings could provide elements of prediction and nuance for constellations of behavior that lead to school suspensions. Overall, accounting for school setting and having multiple perspectives could provide more information about a student's risk with increased data about the relation between behaviors and context. The use of a single datum or decontextualized assessment might be associated with greater harm than good. That is to say, screening is generally likely to indicate a potential problem. However,

more robust assessment processes are more likely to give better information regarding behaviorally defined problems and possible solutions (e.g., Gross, Farmer, & Ochs, 2018). These assessment might include additional gating or environmental assessment, or accounting for policies that could disproportionately affect specific groups of students.

Additionally, participants were enrolled in a randomized trial. Although the tested parenting program showed no effects on suspensions, the sample may have been unique due to willingness to participate in an experimental study and including only non-identified students. Further, other age ranges and transitional periods should be examined to determine if there are idiosyncratic structures for the behavioral symptoms at other developmental periods. Also, findings associated with the cut-points used in this study might be idiosyncratic to the setting and sample and replication studies are needed. Lastly, the overlapping items to factors need to be addressed. There is a wide variety of suggested rules or guidelines for the use of EFA, where attempts to handle overlapping items and codify the procedures are well documented (e.g., Yong & Pearce, 2013). Still, it needs to be kept in mind that EFA is exploratory and that utility of an assessment precedes statistical procedures that are essentially exploratory, non-inferential, and error prone even with large samples (Costello & Osborne, 2005). It is conceded that EFA is more likely data-informed judgement than confirmation of fact.

Implications

There are documented societal costs for school suspensions (Rumberger & Losen, 2016), and this is especially evident when taking into account the potential for increased violent behaviors after a suspension (Breunlin et al., 2002). The need for a useful screener to determine who demonstrates behaviors consistent with those at-risk for school suspension is evident within the school context, especially within the RTI framework. There is a need for screeners that are brief and can provide information related to areas for intervention regarding behavioral issues. The EFA derived misbehavior factor scale of the SDQ is promising in that it predicts suspensions, but its use should be tempered by its limitations. Namely, screener results are most likely to indicate general interfering behaviors, but they are not predictive of complex school discipline responses. Rather, they are better used to identify possible areas for social skills development. Similarly, competency based screening could emphasize identifying and targeting social-emotional and positive behavior needs and instruction within an RTI framework.

Educators could hypothetically quickly assess multiple students serviced solely at tier 1 for suspension risk and emerging behaviors or behavior deficits that correspond to risks for suspension. However, there is a chance that a large proportion of students would not be identified and it is more likely that some students who are not at risk for suspensions will be identified. Pointedly, it needs to be understood that in spite of well-designed and purposeful screening, it is difficult to predict or identify the children that will receive consequences based on complex and potentially arbitrarily made decisions. The possible increase in predicting odds of suspension were small in this study and may indicate that other variables that were unassessed in this study should be reviewed and considered, particularly regarding school climate and policies. Universal screening does have a place in the process, but it

should be part of a process that is systematic, such as RTI, and related to decreasing the identified problem behaviors (Suh et al., 2007). Correspondingly, if schools oriented to PYD, or something similar, it could create a need to screen for competencies or interpersonal skills that would benefit children scholastically and socially. Specifically, as schools adopt positive behavior support and social-emotional learning curricula and systems, it may be necessary to progress on skills related to pro-social and empathetic behaviors (e.g., Cook et al., 2015). Yet, until there are positive behavior skills-based screenings that are better designed, items and scales need to be interpreted with extreme caution.

Funding

References

Azzopardi LM, Camilleri L, Sammut F, & Cefai C (2016). Examining the model structure of the strengths and difficulties questionnaire (SDQ). Xjenza Online, 4, 100–108. doi: 10.7423/XJENZA.2016.2.01

Berkeley S, Bender W, Peaster L, & Saunders L (2009). Implementation of response to intervention: A snapshot of progress. Journal of Learning Disabilities, 42, 85–95. doi:10.1177/0022219408326214 [PubMed: 19103800]

Bornmann L, Leydesdorff L, & Mutz R (2013). The use of percentiles and percentile rank classes in the analysis of bibliometric data: Opportunities and limits. Journal of Informetrics, 7, 158–165. doi:10.1016/j.joi.2012.10.001

Breunlin DC, Cimmarusti RA, Bryant-Edwards TL, & Hetherington JS (2002). Conflict resolution training as an alternative to suspension for violent behavior. The Journal of Educational Research, 95, 349–357. doi:10.1080/00220670209596609

Brooks K, Schiraldi V, & Ziedenberg J (2000). School house hype: Two years later. Washington, DC: Justice Policy Institute/Children's Law Center [Online] Available: http://www.cjcj.org/schoolhousehype/shh2.html.

Catalano RF, Berglund ML, Ryan JA, Lonczak HS, & Hawkins JD (2004). Positive youth development in the United States: Research findings on evaluations of positive youth development programs. The Annals of the American Academy of Political and Social Science, 591, 98–124. doi:10.1177/0002716203260102

Christle C, Jolivette K, & Nelson C (2007). School characteristics related to high school dropout rates. Remedial and Special Education, 28, 325–339. doi:10.1177/07419325070280060201

Cohen J, McCabe L, Michelli NM, & Pickeral T (2009). School climate: Research, policy, practice, and teacher education. Teachers College Record, 111, 180–213.

Cook CR, Frye M, Slemrod T, Lyon AR, Renshaw TL, & Zhang Y (2015). An integrated approach to universal prevention: Independent and combined effects of PBIS and SEL on youths' mental health. School Psychology Quarterly, 30, 166–183. doi: 10.1037/spq0000102 [PubMed: 25602629]

Costello AB, & Osborne JW (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. Practical Assessment, Research & Evaluation, 10, 1–9.

Crawford L, & Ketterlin-Geller LR (2008). Improving math programming for students at risk: Introduction to the special topic issue. Remedial and Special Education, 29, 5–8. doi:10.1177/0741932507309685

Curran FC (2016). Estimating the effect of state zero tolerance laws on exclusionary discipline, racial discipline gaps, and student behavior. Educational Evaluation and Policy Analysis, 38, 647–668. doi:10.3102/0162373716652728

Davis JA (1971). Elementary Survey Analysis. Englewood Cliffs, NJ: Prentice-Hall.

Dickey WC, & Blumberg SJ (2004). Revisiting the factor structure of the strengths and difficulties questionnaire: United States, 2001. Journal of the American Academy of Child & Adolescent Psychiatry, 43, 1159–1167. doi:10.1097/01.chi.0000132808.36708.a9 [PubMed: 15322420]

Dodge KA, Dishion TJ, & Lansford JE (2006). Deviant Peer Influences in Intervention and Public Policy for Youth. Social Policy Report. Volume 20, Number 1. Society for Research in Child Development.

Doyle MM, Murphy J, & Shevlin M (2016). Competing factor models of child and adolescent psychopathology. Journal of Abnormal Child Psychology, 44, 1559–1571. doi:10.1007/s10802-016-0129-9 [PubMed: 26783112]

Durlak JA, Taylor RD, Kawashima K, Pachan MK, DuPre EP, Celio CI, … & Weissberg RP (2007). Effects of positive youth development programs on school, family, and community systems. American Journal of Community Psychology, 39, 269–286. doi:10.1007/s10464-007-9112-5 [PubMed: 17401642]

Feil EG, Small JW, Forness SR, Kaiser AP, Hancock TB, Serna LA, … & Boyce CA (2005). Using different measures, informants, and clinical cut-off points to estimate prevalence of emotional or behavioral disorders in preschoolers: Effects on age, gender, and ethnicity. Behavioral Disorders, 30, 375–391. doi:10.1177/019874290503000405

Fabelo T, Thompson M, Plotkin M, Carmichael D, Marchbanks M, and Booth E (2011). Breaking schools' rules: A statewide study of how school discipline relates to students' success and juvenile justice involvement. New York: Council of State Governments Justice Center Retrieved from justicecenter.csg.org/resources/juveniles.

Fleming C, Mason W, Thompson R, Haggerty K, & Gross T (2015). Child and parent report of parenting as predictors of substance use and suspensions from school. The Journal of Early Adolescence,36, 625–645. doi:10.1177/0272431615574886 [PubMed: 27493444]

Flouri E, Midouhas E, Joshi H, & Tzavidis N (2015). Emotional and behavioural resilience to multiple risk exposure in early life: The role of parenting. European Child & Adolescent Psychiatry, 24, 745–755. doi:10.1007/s00787-014-0619-7 [PubMed: 25300919]

Flynn RM, Lissy R, Alicea S, Tazartes L, & McKay MM (2016). Professional development for teachers plus coaching related to school-wide suspensions for a large urban school system. Children and Youth Services Review, 62, 29–39. doi:10.1016/j.childyouth.2016.01.015

Forster JJ, McDonald JW, & Smith PW (1996). Monte Carlo exact conditional tests for log-linear and logistic models. Journal of the Royal Statistical Society. Series B (Methodological), 445–453.

Ganao JSD, Silvestre FS, & Glenn JW (2013). Assessing the differential impact of contextual factors on school suspension for black and white students. The Journal of Negro Education, 82, 393–407. doi:10.7709/jnegroeducation.82.4.0393

Gibson P, Haight W, & Kayama M (2016). Out-of-school suspensions of black youths: Culture, ability, disability, gender, and perspective. Social Work, 61, 235–243. doi:10.1093/sw/sww021 [PubMed: 27501641]

Gómez-Beneyto M, Nolasco A, Moncho J, Pereyra-Zamora P, Tamayo-Fonseca N, Munarriz M, … & Girón M (2013). Psychometric behaviour of the strengths and difficulties questionnaire (SDQ) in the Spanish national health survey 2006. BMC psychiatry, 13, 95. doi: 10.1186/1471-244X-13-95 [PubMed: 23522343]

Goodman A, Lamping D, & Ploubidis G (2010). When to use broader internalising and externalising subscales instead of the hypothesized five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers, and children. Journal of Abnormal Child Psychology, 38, 1179–1191. doi:10.1007/s10802-010-9434-x [PubMed: 20623175]

Gross T, Farmer R, & Ochs S (2018). Evidence-based assessment: Best practices, customary practices, and recommendations for field-based assessment. Contemporary School Psychology, Advanced On-line. doi: 10.1007/s40688-018-0186-x

Heilbrun A, Cornell D, & Lovegrove P (2015). Principal attitudes regarding zero tolerance and racial disparities in school suspensions. Psychology in the Schools, 52, 489–499. doi:10.1002/pits.21838

Henson RK, & Roberts JK (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. Educational and Psychological Measurement, 66, 393–416. doi:10.1177/0013164405282485

Hughes C, & Dexter D (2015). Field studies of RTI programs, Revised. RTI Action Network Web, National Center for Learning Disabilities, Retrieved January 18, 2017 from http://www.rtinetwork.org/learn/research.

Information for researchers and professionals about the Strengths & Difficulties Questionnaires. (n.d.). Retrieved from http://www.sdqinfo.com/.

Kamphaus R, DiStefano C, Dowdy E, Eklund K, & Dunn A (2010). Determining the presence of a problem: Comparing two approaches for detecting youth behavioral risk. School Psychology Review,39, 395–407.

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

Kettler R, Glover T, Albers C, & Feeney-Kettler K (2014). Universal screening in educational settings: evidence-based decision making for schools. Washington, D.C.: American Psychological Association.

Lane K, Parks R, Kalberg J, & Carter E (2007). Systematic screening at the middle school level: Score reliability and validity of the student risk screening scale. Journal of Emotional and Behavioral Disorders, 15, 209–222. doi:10.1177/10634266070150040301

Lane KL, Wehby J, Robertson E, & Rogers L (2007). How do different types of high school students respond to positive behavior support programs? Characteristics and responsiveness of teacher-identified students. Journal of Emotional and Behavioral Disorders, 15, 3–20. doi:10.1177/10634266070150010201

Liu J (2004). Childhood externalizing behavior: theory and implications. Journal of Child and Adolescent Psychiatric Nursing, 17, 93–103. doi:10.1111/j.1744-6171.2004.tb00003.x [PubMed: 15535385]

Losen DJ, & Skiba RJ (2010). Suspended education: Urban middle schools in crisis. Los Angeles, CA: The Civil Rights Project.

Lynch M, & Cicchetti D (2002). Links between community violence and the family system: Evidence from children's feelings of relatedness and perceptions of parent behavior. Family Process, 41, 519–532. 10.1111/j.1545-5300.2002.41314.x [PubMed: 12395572]

Matjasko JL, Needham BL, Grunden LN, & Farb AF (2010). Violent victimization and perpetration during adolescence: Developmental stage dependent ecological models. Journal of Youth and Adolescence, 39, 1053–1066. doi:10.1007/s10964-010-9508-7 [PubMed: 20111894]

McHugh ML (2013). The chi-square test of independence. Biochemia Medica, 23, 143–149. doi:10.11613/bm.2013.018 [PubMed: 23894860]

Neymotin F (2014). How parental involvement affects childhood behavioral outcomes. Journal of Family and Economic Issues, 35, 433–451. doi:10.1007/s10834-013-9383-y

Niclasen J, Skovgaard A, Andersen A, Sømhovd M, & Obel C (2012). A confirmatory approach to examining the factor structure of the strengths and difficulties questionnaire (SDQ): A large scale cohort study. Journal of Abnormal Child Psychology, 41, 355–365. doi:10.1007/s10802-012-9683-y

Number of students suspended and expelled from public elementary and secondary schools, by sex, race/ethnicity, and state: 2011–12. (2015, 11). Retrieved January 10, 2017, from https://nces.ed.gov/programs/digest/d15/tables/dt15_233.30.asp

Office for Civil Rights. (2016). 2013–2014 civil rights data collection: A first look Key data highlights on equity and opportunity gaps in our nation's public schools. Washington, DC: U.S. Department of Education Retrieved from https://www2.ed.gov/about/offices/list/ocr/docs/2013-14-first-look.pdf.

Parker C, Paget A, Ford T, & Gwernan-Jones R (2016). '.he was excluded for the kind of behaviour that we thought he needed support with…' A qualitative analysis of the experiences and perspectives of parents whose children have been excluded from school. Emotional and Behavioural Difficulties,21, 133–151. doi:10.1080/13632752.2015.1120070

Petersen IT, Bates JE, Dodge KA, Lansford JE, & Pettit GS (2015). Describing and predicting developmental profiles of externalizing problems from childhood to adulthood. Development and Psychopathology, 27, 791–818. doi:10.1017/S0954579414000789 [PubMed: 25166430]

Pratt TC, & Cullen FT (2000). The empirical status of Gottfredson and Hirschi's general theory of crime: A meta-analysis. Criminology, 38, 931–964. 10.1111/j.1745-9125.2000.tb00911.x

Reef J, Diamantopoulou S, van Meurs I, Verhulst F, & van der Ende J (2009). Child to adult continuities of psychopathology: A 24-year follow-up. Acta Psychiatrica Scandinavica, 120, 230–238. doi:10.1111/j.1600-0447.2009.01422.x [PubMed: 19522882]

Reinke WM, Thompson A, Herman KC, Holmes S, Owens S, Cohen D, … & Copeland C (2017). The county schools mental health coalition: A model for community-level impact. School Mental Health, Advanced Online. doi: 10.1007/s12310-017-9227-2

Robers S, Zhang J, & Truman J (2012). Indicators of school crime and safety: 2011 (NCES 2012–002/NCJ 236021). Washington, DC: National Center for Education Statistics, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice.

Roberts BW, Chernyshenko OS, Stark S, & Goldberg LR (2005). The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires. Personnel Psychology, 58, 103–139. doi:10.1111/j.1744-6570.2005.00301.x

Rumberger RW, & Losen DJ (2016). The High Cost of Harsh Discipline and Its Disparate Impact. Civil Rights Project-Proyecto Derechos Civiles.

Severson H, Walker H, Hope-Doolittle J, Kratochwill T, & Gresham F (2007). Proactive, early screening to detect behaviorally at-risk students: Issues, approaches, emerging innovations, and professional practices. Journal of School Psychology, 45, 193–223. doi:10.1016/j.jsp.2006.11.003

Skiba R, Chung C, Trachok M, Baker T, Sheya A, & Hughes R (2014). Parsing Disciplinary Disproportionality: Contributions of Infraction, Student, and School Characteristics to Out-of-School Suspension and Expulsion. American Educational Research Journal, 51, 640–670. doi:10.3102/0002831214541670

Suh S, Suh J, & Houston I (2007). Predictors of categorical at-risk high school dropouts. Journal of Counseling & Development, 85, 196–203. doi:10.1002/j.1556-6678.2007.tb00463.x

Suh S, & Suh J (2007). Risk factors and levels of risk for high school dropouts. Professional School Counseling, 10, 297–306. doi:10.5330/prsc.10.3.w26024vvw6541gv7

Thapar A, Harrington R, & McGuffin P (2001). Examining the comorbidity of ADHD-related behaviours and conduct problems using a twin study design. The British Journal of Psychiatry, 179, 224–229. doi:10.1192/bjp.179.3.224 [PubMed: 11532799]

Thornberry TP, & Krohn MD (2000). The self-report method for measuring delinquency and crime. Criminal Justice, 4, 33–83.

Walker H, Severson H, Feil E (2014) Systematic screening for behavior disorders (SSBD) technical manual: Universal screening for preK-9, 2, Eugene, OR: Pacific Northwest Publishing.

Wang Y, & Chen H-J (2012). Use of Percentiles and Z -Scores in Anthropometry In Preedy VR(ed.), Handbook of Anthropometry: Physical Measures of Human Form in Health and Disease (pp. 29–48). New York: Springer.

Yong AG, & Pearce S (2013). A beginner's guide to factor analysis: Focusing on exploratory factor analysis. Tutorials in Quantitative Methods for Psychology, 9, 79–94. doi:10.20982/tqmp.09.2.p079

Yuan KH, & Bentler PM (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. Sociological methodology, 30, 165–200. doi:10.1111/0081-1750.00078

Bentler PM,& Bonett DG (1980). Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 88(3), 588–606.

Burke R, Schuchmann LF, & Barnes BA (2006). Common sense parenting trainer guide. Boys Town, NE: Boys Town Press.

De Voe JF, Peter K, Kaufman P, Miller A, Noonan M, Snyder T, & Baum K (2004). Indicators of School Crime and Safety: 2004 (NCES 2005–002/NCJ 205290) U.S. Departments of Education and Justice. Washington, DC: U.S. Government Printing Office.

Fletcher J, & Wolfe B (2009). Long-term consequences of childhood ADHD on criminal activities. The Journal of Mental Health Policy and Economics, 12, 119–138. [PubMed: 19996475]

Hooper D, Coughlin J, & Mullen MR (2008). Structural equation modeling: Guidelines for determining model fit. Electronic Journal of Business Research Methods, 6, 53–60.

Kilgus SP, Chafouleas SM, & Riley-Tillman TC (2013). Development and initial validation of the Social and Academic Behavior Risk Screener for elementary grades. School Psychology Quarterly, 28, 210–226. [PubMed: 23773134]

Kilgus SP, Sims WA, von der Embse NP, & Riley-Tillman TC (2015). Confirmation of models for interpretation and use of the Social and Academic Behavior Risk Screener (SABRS). School Psychology Quarterly, 30(3), 335–352. [PubMed: 25264747]

Kilgus SP, Eklund K, Nathaniel P, Taylor CN, & Sims WA (2016). Psychometric defensibility of the Social, Academic, and Emotional Behavior Risk Screener (SAEBRS) Teacher Rating Scale and multiple gating procedure within elementary and middle school samples. Journal of School Psychology, 58, 21–39. [PubMed: 27586068]

Mason WA, Fleming CB, Gross TJ, Thompson RW, Parra GR, Haggerty KP, & Snyder JJ (2016). Randomized trial of parent training to prevent adolescent problem behaviors during the high school transition. Journal of Family Psychology, 30(8), 944–954. 10.1037/fam0000204.supp [PubMed: 27504751]

Mendez LMR, & Knoff HM (2003). Who gets suspended from school and why: A demographic analysis of schools and disciplinary infractions in a large school district. Education and Treatment of Children, 26, 30–51.

Muthén LK and Muthén BO (1998–2017). Mplus User's Guide. EighthEdition Los Angeles, CA: Muthén & Muthén

Sasser TR, Kalvin CB, & Bierman KL (2016). Developmental trajectories of clinically significant attentiondeficit/hyperactivity disorder (ADHD) symptoms from grade 3 through 12 in a high-risk sample: Predictors and outcomes. Journal of Abnormal Psychology, 125, 207–219. [PubMed: 26854506]

Skiba RJ, Horner RH, Chung CG, Rausch MK, May SL, & Tobin T (2011). Race is not neutral: A national investigation of African American and Latino disproportionality in school discipline. School Psychology Review, 40, 85–107.

Vandenberg RJ, & Lance CE (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. Organizational Research Methods, 3(1), 4–69. 10.1177/109442810031002.
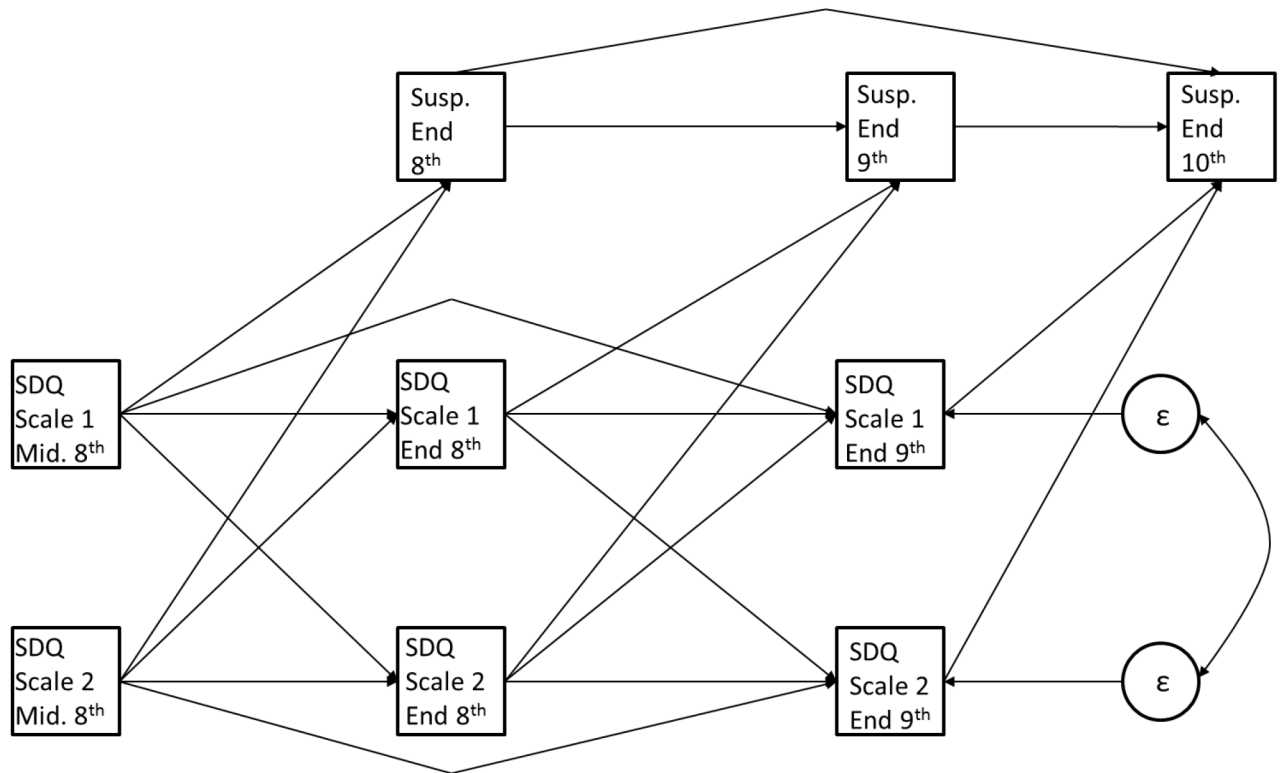
Figure 1.
Logistic regression path analysis model used for the SDQ original and EFA derived scales.
Covariate paths are removed for parsimony. SDQ Scale 1 for the original model = SDQ CP;
SDQ Scale 1 for the EFA model 1 = SDQ EXT; SDQ Scale 1 for the EFA model 2 = SDQ
EMC. SDQ Scale 2 for the original model = SDQ HS; SDQ Scale 2 for the EFA model 1 =
SDQ AG; SDQ Scale 2 for the EFA model 2 = SDQ MOC.

Table 1.

Parent Demographic Information.

	Middle 8th Grade	End 10th Grade
N	321	303
Age in Years *M*(*SD*)	40.70 (7.69)	
	N (%)	**N (%)**
Parent		
Mother	226 (68%)	232 (77)%
Father	53 (17%)	51 (17%)
Other	42 (13%)	20 (7%)
Race		
White or Caucasian	158 (49%)	150 (50%)
Black or African American	87 (27%)	81 (27%)
Asian or Asian American	15 (5%)	14 (5%)
Hispanic, Spanish or Latino	32 (10%)	30 (10%)
Native Indian or Alaskan Native	7 (2%)	7 (2%)
Native Hawaiian or other Pacific Islander	15 (5%)	15 (5%)
Unknown	7 (2%)	6 (2%)
Highest Level of Education		
Some High School	25 (8%)	28 (9%)
High School Diploma or Equivalent	59 (18%)	56 (18%)
Some College or Vocational training	175 (55%)	170 (56%)
Bachelor's or more advanced degree	40 (13%)	47 (16%)
Unknown	26 (8%)	2 (1%)
Household incomes below $24,000	132 (41%)	115 (36%)
Receive Food Stamps	185 (58%)	149 (46%)
Employment		
Full time	140 (44%)	166 (55%)
Part time	48 (15%)	33 (11%)
Unemployed	42 (13%)	29 (10%)
Homemaker	25 (8%)	27 (9%)
Student	23 (7%)	11 (4%)
Disabled	28 (9%)	29 (10%)
Retired	2 (<1%)	4 (1%)
Unknown	13 (4%)	4 (1%)

Table 2.

Youth Demographic Information.

	Middle 8th Grade	End 10th Grade
N	321	301
Age in Years *M*(*SD*)	13.46 (0.53)	15.95 (0.47)
	N (%)	**N (%)**
Gender		
Boy	150 (47%)	140 (47%)
Girl	171 (53%)	161 (53%)
Race		
White or Caucasian	124 (39%)	119 (40%)
Black or African American	111 (35%)	102 (34%)
Asian or Asian American	25 (8%)	23 (8%)
Hispanic, Spanish or Latino	36 (11%)	33 (11%)
American Indian or Alaskan Native	13 (4%)	12 (4%)
Native Hawaiian or other Pacific Islander	12 (4%)	12 (4%)

Table 3

Model Fit Statistics by Number of Factors Included in the Exploratory Factor Analyses

EFA Scale	χ^2	df	RMSEA [C.I. 90%]	RMSR
Symptoms-based				
1 Factor	1056.64 ***	275	.095 [.089, .101]	.098
2 Factors	725.82 ***	251	.077 [.071, .084]	.070
3 Factors	**553.13 *****	**228**	**.067 [.060, .074]**	**.055**
4 Factors	461.43 ***	206	.062 [.055, .070]	.047
5 Factors	357.24 ***	185	.054 [.046, .062]	.040
Strengths-based				
1 Factor	1056.64 ***	275	.095 [.089, .101]	.098
2 Factors	725.82 ***	251	.077 [.071, .084]	.070
3 Factors	**553.13 *****	**228**	**.067 [.060, .074]**	**.055**
4 Factors	461.43 ***	206	.062 [.055, .070]	.047
5 Factors	357.24 ***	185	.054 [.046, .062]	.040

Notes.

*
p .05;

**
p .01;

***
p .001.

Table 4

Exploratory Factor Analysis Strengths and Difficulty Questionnaire (SDQ) Item Loadings on the 3-factor Models from each EFA.

SDQ – Developer Scored				SDQ – Strengths-based Scored			
Item	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3	Item
Steals from home, school or elsewhere	**.35**	-.01	.08	**.35**	-.01	.08	Respects others property
Often fights with other youth or bullies them	**.38**	.04	.09	**.38**	.04	.09	Resolves disputes appropriately
Often lies or cheats	**.51**	.09	.14	**.51**	.09	.14	Often honest
Generally well behaved, usually does what adults request (R)	**.69**	.05	.05	**.69**	.05	.05	Generally well behaved, usually does what adults request
Often loses temper	**.48**	-.02	.20	**.48**	-.02	.20	Often controls temper
Restless, overactive, cannot stay still for long	.28	.03	**.52**	.28	.03	**.52**	Can stay still for appropriate periods of time
Constantly fidgeting or squirming	.28	.06	**.49**	.28	.06	**.49**	Often appears relaxed
Easily distracted, concentration wanders	**.51**	.12	**.49**	**.51**	.12	**.49**	Thinks about things
Thinks things out before acting (R)	**.49**	.13	.19	**.49**	.13	.19	Thinks things out before acting
Good attention span, sees work through to the end (R)	**.50**	.07	**.38**	**.50**	.07	**.38**	Good attention span, sees work through to the end
Many worries or often seems worried	-.02	**.42**	**.52**	-.02	**.42**	**.52**	Often seems calm
Often unhappy, depressed or tearful	.15	**.36**	**.46**	.15	**.36**	**.46**	Often happy
Nervous in new situations, easily loses confidence	.15	**.39**	**.48**	.15	**.39**	**.48**	Confident in new situations
Often complains of headaches, stomach-aches or sickness	-.02	.04	**.38**	-.02	.04	**.38**	Often feels comfortable
Many fears, easily scared	-.03	**.51**	.25	-.03	**.51**	.25	Usually unafraid
Generally liked by other youth (R)	.20	**.62**	.03	.20	**.62**	.03	Generally liked by other youth
Gets along better with adults than with other youth	-.08	**.44**	.04	-.08	**.44**	.04	Gets along better with youth than with other adults
Picked on or bullied by other youth	.15	**.48**	.18	.15	**.48**	.18	Gets along with other youth
Has at least one good friend (R)	.27	**.36**	-.01	.27	**.36**	-.01	Has at least one good friend
Would rather be alone than with other youth	.04	**.55**	.06	.04	**.55**	.06	Would rather be with other youth than alone
Helpful if someone is hurt, upset or feeling ill	**-.36**	-.10	.23	**.36**	-.1	.23	Helpful if someone is hurt, upset or feeling ill
Considerate of other people's feelings	**-.69**	-.15	.04	**.69**	-.15	.04	Considerate of other people's feelings
Shares readily with other youth, for example books, games, food	-.25	-.26	.24	.25	-.26	.24	Shares readily with other youth, for example books, games, food
Kind to younger children	-.25	-.05	-.01	.25	-.05	-.01	Kind to younger children
Often offers to help others (parents, teachers, children)	**-.46**	-.02	-.03	**.46**	-.02	-.03	Often offers to help others
Eigen Values	5.22	2.44	1.83	5.22	2.44	1.83	

Notes. (R) = reverse scored item.

Table 5

EFA scales with Strengths and Difficulty Questionnaire (SDQ) items reflecting symptoms.

	M	*SD*	*α*	*Range*
Misbehavior	3.47	2.54	.75	0 to 11
Fails to do what told	0.38	0.52		
Lies/cheats	0.31	0.58		
Distractible	0.78	0.75		
Inattentive	0.91	0.69		
Impulsive	0.85	0.61		
Fights/bullies	0.14	0.38		
Steals	0.12	0.40		
Often loses temper[a]	2.25	1.08		
Isolation	2.24	2.17	.67	0 to 12
Disliked	0.28	0.52		
Prefers being alone	0.36	0.60		
Easily scared	0.30	0.55		
Bullied	0.40	0.63		
Prefers adults company	0.67	0.68		
Lacks Friends	0.22	0.56		
Agitation	2.87	2.46	.70	0 to 12
Restless	0.43	0.64		
Worried	0.54	0.70		
Fidgeting	0.37	0.64		
Nervous/unsure	0.75	0.72		
Unhappy	0.29	0.52		
Somatic complaints	0.50	0.66		

Notes.

[a] Item removed to improve internal consistency.

Table 6

EFA scales with Strengths and Difficulty Questionnaire (SDQ) items reflecting strengths.

	M	*SD*	*α*	*Range*
Emotional competence	9.12	2.46	.70	0 to 12
Feels comfortable	1.50	0.66		
Calm	1.46	0.70		
Happy	1.71	0.52		
Confident	1.25	0.72		
Still	1.57	0.64		
Relaxed	1.63	0.64		
Social Competence	9.74	2.18	.67	0 to 12
Usually unafraid	1.70	0.55		
Good Friend	1.78	0.56		
Peers like	1.72	0.52		
Prefers peers	1.33	0.68		
Prefers company	1.64	0.60		
Gets along with peers	1.60	0.63		
Moral Competence	16.88	3.62	.79	5 to 22
Well behaved	1.62	0.52		
Controls temper	1.53	0.62		
Respects property	1.89	0.40		
Respects others	1.86	0.38		
Truthful	1.69	0.58		
Thinks ahead	1.15	0.61		
Attentive	1.09	0.69		
Focused	1.22	0.75		
Helps in need	1.72	0.50		
Offers to help	1.58	0.57		
Considerate	1.56	0.53		

Table 7

Intercorrelations of Strengths and Difficulties (SDQ) Scales from the Original Scales, and both Symptoms-based and Strengths-based Exploratory Factor Analysis (EFA) Scales. Cronbach's alpha (α) for each measure is in bold on the diagonal.

SDQ Traditional Scales

	Conduct Problems			Hyperactivity		
	Mid 8th	End 8th	End 9th	Mid 8th	End 8th	End 9th
Conduct Problems						
Mid 8th	**.66**					
End 8th	.58***	**.66**				
End 9th	.50***	.59***	**.68**			
Hyperactivity						
Mid 8th	.51***	.49***	.35***	**.75**		
End 8th	.42***	.57***	.37***	.72***	**.76**	
End 9th	.39***	.53***	.53***	.64***	.67***	**.76**

SDQ Symptoms EFA Scales

	Misbehavior			Isolation			Agitation		
	Mid 8th	End 8th	End 9th	Mid 8th	End 8th	End 9th	Mid 8th	End 8th	End 9th
Misbehavior									
Mid 8th	**.75**								
End 8th	.69***	**.76**							
End 9th	.60***	.67***	**.76**						
Isolation									
Mid 8th	.29***	.23***	.12*	**.67**					
End 8th	.29***	.33***	.15**	.66***	**.61**				
End 9th	.25***	.27***	.20***	.69***	.69***	**.58**			
Agitation									
Mid 8th	.47***	.37***	.33***	.40***	.34***	.34***	**.70**		
End 8th	.41***	.51***	.36***	.30***	.46***	.32***	.64***	**.65**	
End 9th	.32***	.39***	.49***	.34***	.31***	.39***	.55***	.57***	**.66**

SDQ Strengths Scored EFA Scales

	Emotional competence			Social competence			Moral competence		
	Mid 8th	End 8th	End 9th	Mid 8th	End 8th	End 9th	Mid 8th	End 8th	End 9th
Emotional competence									

	Mid 8th	End 8th	End 9th	Mid 8th	End 8th	End 9th	Mid 8th	End 8th	End 9th
Mid 8th	**.70**	.64***	.55***	.40***	.33***	.32***	.43***	.33***	.30***
End 8th		**.65**	.57***	.32***	.46***	.33***	.37***	.46***	.34***
End 9th			**.66**	.34***	.32***	.44***	.28***	.35***	.45***
Social competence									
Mid 8th				**.67**	.65***	.63***	.27***	.24***	.14*
End 8th					**.61**	.70***	.28***	.35***	.19***
End 9th						**.58**	.24***	.30***	.24***
Moral competence									
Mid 8th							**.79**	.70***	.58***
End 8th								**.81**	.68***
End 9th									**.82**

Table 8

Correlations of Suspensions, Gender, and Race with Strengths and Difficulties (SDQ) Scales Across Each Time Point, including Number (Percentage) or Mean and Standard Deviation.

	Suspension			Gender	Race	N	n (%)/M	SD
	8th	9th	10th					
Suspension								
End 8th	--					321	48 (16%)	
End 9th	.35***	--				302	56 (19%)	
End 10th	.32***	.36***	--			284	42 (15%)	
Gender	.18***	.08	.04	--		321	150 (47%)	
Race	.21***	.13*	.19***	.05	--	321	111 (35%)	
SDQ Original Scales								
Conduct Problems								
Mid 8th	.31***	.17**	.21***	.08	-.05	317	1.41	1.65
End 8th	.33***	.19***	.23***	.09	-.07	296	1.28	1.61
End 9th	.24***	.24***	.24***	.10	-.06	304	1.46	1.74
Hyperactivity								
Mid 8th	.24***	.24***	.14*	.18***	-.05	316	3.32	2.34
End 8th	.31***	.25***	.17**	.15*	-.07	295	2.81	2.23
End 9th	.29***	.33***	.21***	.17**	-.06	302	3.25	2.44
SDQ Symptoms-based EFA Scales								
Misbehavior								
Mid 8th	.28***	.23***	.21***	.13*	.002	317	3.47	2.54
End 8th	.33***	.25***	.20***	.14*	-.01	295	3.01	2.45
End 9th	.27***	.32***	.25***	.14*	-.01	304	3.46	2.68
Isolation								
Mid 8th	.05	.03	-.03	.09	-.13*	316	2.24	2.17
End 8th	.13*	.09	.04	.03	-.15**	296	2.04	1.97

	Suspension			**Gender**	**Race**	**N**	**n (%)/ M**	**SD**
	8th	**9th**	**10th**					
Agitation								
End 9th	.09	.11	.002	.04	-.15*	298	2.36	1.94
Mid 8th	.16**	.21***	.07	.04	-.14*	315	2.87	2.46
End 8th	.18**	.15**	.18**	-.01	-.22***	295	2.39	2.21
End 9th	.15**	.30***	.16**	-.10	-.18**	302	2.90	2.34
SDQ Strengths-based EFA Scales								
Emotional competence								
Mid 8th	-.16**	-.20***	-.066	-.04	.14*	318	9.12	2.46
End 8th	-.18**	-.15*	-.18**	.01	.22***	296	9.61	2.21
End 9th	-.15*	-.29***	-.15*	.08	.17**	304	9.08	2.35
Social competence								
Mid 8th	-.05	-.03	.03	-.10	.13*	318	9.74	2.19
End 8th	-.13*	-.09	-.04	-.03	.15**	296	9.96	1.97
End 9th	-.06	-.13*	-.004	-.01	.15**	304	9.62	2.02
Moral competence								
Mid 8th	-.32***	-.22***	-.19**	-.15**	-.01	318	16.88	3.62
End 8th	-.35***	-.26***	-.20***	-.13*	-.01	296	17.44	3.59
End 9th	-.26***	-.31***	-.24***	-.12*	-.01	304	16.81	3.88

Notes. Suspension = any suspensions by the end of the school year. Gender = girl (= 0) or boy (=1). Race = Non-African-American (= 0) of African-American (=1).

*
p .05;

**
p .01;

***
.001.

Table 9

Logistic Regression Results for Any Suspensions in 8th, 9th, and 10th Grade predicted by the Strengths and Difficulties Questionnaire (SDQ) Conduct Problems (CP) and Hyperactivity (HS) Scales.

	8th		9th		10th	
	b (S.E.)	OR	*b (S.E.)*	OR	*b (S.E.)*	OR
Suspension						
8th			1.50 (0.41) ***	4.50	0.90 (0.48)	2.45
9th					1.33(0.42) **	3.78
Gender	.87 (0.37) *	2.38	.01 (0.34)	1.01	−.29 (0.41)	0.75
Race	1.41 (0.36) ***	4.11	.67 (0.36)	1.95	1.11 (0.43) **	3.03
SDQ CP	0.39 (0.11) ***	1.48	0.02 (0.11)	1.02	0.28 (0.12) *	1.32
SDQ HS	.16 (0.08)	1.17	.24 (.09) **	1.27	.05 (0.10)	1.05

Notes. Suspension = any suspensions by the end of the school year. Gender = girl (= 0) or boy (=1). Race = Non-African-American (= 0) of African-American (=1).

*
p .05;

**
p .01;

***
p .001.

Table 10

Logistic Regression Results for Any Suspensions in 8th, 9th, and 10th Grade predicted by the Strengths and Difficulties Questionnaire (SDQ) Exploratory Factor Analysis (EFA) Derived Scales.

	8th		9th		10th	
Symptoms-based EFA						
	b (S.E.)	OR	*b (S.E.)*	OR	*b (S.E.)*	OR
Suspension						
8th			1.51 (0.39) ***	4.52	0.96 (0.46) *	2.61
9th					1.25 (0.44) **	3.49
Gender	0.88 (0.36) *	2.40	0.06 (0.33)	1.06	−0.21 (0.40)	0.81
Race	1.32 (0.36) ***	3.74	0.67 (0.36)	1.95	1.04 (0.45) *	2.82
SDQ MB	.27 (0.07) ***	1.31	.15 (0.07) *	1.16	.17 (0.08) *	1.18
SDQ AG	.09 (0.07)	1.10	.11 (0.07)	1.11	.05 (0.11)	1.05
Strengths-based EFA						
	b (S.E.)	OR	*b (S.E.)*	OR	*b (S.E.)*	OR
Suspension						
8th			1.44 (0.40) ***	4.21	0.90 (0.47)	2.47
9th					1.36 (0.43) **	3.90
Gender	0.87 (0.36) *	2.39	0.06 (0.34)	1.06	−0.21 (0.39)	0.81
Race	1.41 (0.37) ***	4.09	0.65 (0.36)	1.92	0.98 (0.43) *	2.65
SDQ EMC	−.10 (0.07)	0.91	−.13 (0.07)	0.88	−.05 (0.11)	0.95
SDQ MOC	−.23 (0.05) ***	0.80	−.09 (0.07)	0.91	−.10 (0.05)	0.91

Notes. Suspension = any suspensions by the end of the school year. Gender = girl (= 0) or boy (=1). Race = Non-African-American (= 0) of African-American (=1). MB = misbehavior scale. AG = agitation scale. EMC = emotional competence scale. MOC = moral competence scale.

*
  p .05;

**
  p .01;

***
  p .001.

Table 11

Chi-Square Analyses Across Time Periods for Elevated-Risk Score at 75th percentile and 90th percentile using the Misbehavior Factor Scale.

75th Percentile: Elevated Cut-Score = 5

Risk	8th Grade SDQ-MB			9th grade SDQ-MB			10th grade SDQ-MB		
	Low N(%)	Elevated N(%)	Total N	Low N(%)	Elevated N(%)	Total N	Low N(%)	Elevated N(%)	Total N
Suspension									
No	198(74%)	71(26%)	269	183(80%)	47(20%)	230	169(73%)	64(27%)	233
Yes	20(42%)	28(58%)	48	28(57%)	21(43%)	49	18(47%)	20(53%)	38
Total N	218	99	317	211	68	279	187	84	271
χ^2 (df)	19.35 (1) ***			11.02 (1) ***			9.67 (1) **		
Phi	.25 ***			.20 ***			.19 **		
ppv	.28			.31			.24		
npv	.91			.87			.90		

90th Percentile: Elevated Cut-Score = 7

Risk	8th Grade SDQ-MB			9th grade SDQ-MB			10th grade SDQ-MB		
	Low N(%)	Elevated N(%)	Total N	Low N(%)	Elevated N(%)	Total N	Low N(%)	Elevated N(%)	Total N
Suspension									
No	243(90%)	26(10%)	269	213(93%)	17(7%)	230	211(91%)	22(9%)	233
Yes	33(69%)	15(31%)	48	37(76%)	12(24%)	49	27(71%)	11(29%)	38
Total	276	41	317	250	29	279	238	33	271
χ^2 (df)	16.85 (1) ***			12.68 (1) ***			11.62 (1) ***		
Phi	.23 ***			.21 ***			.21 ***		
ppv	.37			.41			.33		
npv	.88			.85			.89		

Notes. SDQ-MB = Misbehavior scale.

*
  p .05,

**
  p .01,

***
p
.001.