Published in final edited form as:

J Consult Clin Psychol. 2013 August ; 81(4): 588–599. doi:10.1037/a0032366.

The Predictive Utility of a Brief Kindergarten Screening Measure of Child Behavior Problems

Sarah Jensen Racz,

Department of Mental Health, Johns Hopkins Bloomberg School of Public Health

Kevin M. King,

Department of Psychology, University of Washington

Johnny Wu,

Department of Statistics, University of Florida

Katie Witkiewitz,

Department of Psychology, Washington State University

Robert J. McMahon, and

Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada, and the Child & Family Research Institute, Vancouver, British Columbia, Canada

The Conduct Problems Prevention Research Group

Abstract

Objective—Kindergarten teacher ratings, such as those from the Teacher Observation of Classroom Adaptation–Revised (TOCA-R), are a promising cost- and time-effective screening method to identify children at risk for later problems. Previous research with the TOCA-R has been mainly limited to outcomes in a single domain measured during elementary school. The goal of the current study was to examine the ability of TOCA-R sum scores to predict outcomes in multiple domains across distinct developmental periods (i.e., late childhood, middle adolescence, late adolescence).

Method—We used data from the Fast Track Project, a large multisite study with children at risk for conduct problems ($n = 752$; M age at start of study = 6.55 years; 57.7% male; 49.9% Caucasian, 46.3% African American). Kindergarten TOCA-R sum scores were used as the

© 2013 American Psychological Association

Correspondence concerning this article should be addressed to Sarah Jensen Racz, Johns Hopkins Bloomberg School of Public Health, Department of Mental Health, 624 North Broadway, HH Room 808, Baltimore, MD 21205. sracz@jhsph.edu.

Members of the Conduct Problems Prevention Research Group, in alphabetical order, include Karen L. Bierman, Department of Psychology, Pennsylvania State University; John D. Coie, Duke University; Kenneth A. Dodge, Center for Child and Family Policy, Duke University; Mark T. Greenberg, Department of Human Development and Family Studies, Pennsylvania State University; John E. Lochman, Department of Psychology, The University of Alabama; Robert J. McMahon, Department of Psychology, Simon Fraser University, and the Child & Family Research Institute; and Ellen E. Pinderhughes, Department of Child Development, Tufts University.

Karen L. Bierman, John D. Coie, Kenneth A. Dodge, Mark T. Greenberg, John E. Lochman, and Robert J. McMahon are the developers of the Fast Track curriculum and have a publishing agreement with Oxford University Press. Mark T. Greenberg is an author on the PATHS curriculum and has a royalty agreement with Channing-Bete, Inc. He is a principal in PATHS Training, LLC. Robert J. McMahon is a coauthor of *Helping the Noncompliant Child* and has a royalty agreement with Guilford Publications, Inc.; he is also a member of the Treatments That Work Scientific Advisory Board with Oxford University Press.

We are grateful for the collaboration of the school districts that participated in the Fast Track project as well as the hard work and dedication of the many staff members who implemented the project, collected the evaluation data, and assisted with data management and analyses. For additional information concerning Fast Track, see <http://www.fasttrackproject.org>.

predictor in regression analyses; outcomes included school difficulties, externalizing diagnoses and symptom counts, and substance use.

Results—TOCA-R sum scores predicted school outcomes at all time points, diagnosis of ADHD in 9th grade, several externalizing disorder symptom counts, and cigarette use in 12th grade.

Conclusions—The findings demonstrate the predictive utility of the TOCA-R when examining outcomes within the school setting. Therefore, these results suggest the 10-item TOCA-R may provide a quick and accurate screening of children at risk for later problems. Implications for prevention and intervention programs are discussed.

Keywords

child behavior problems; screening methods; TOCA-R

Early behavior problems (e.g., aggressiveness, disruptiveness, oppositionality) are widely recognized as a risk factor for later violence and antisocial behavior (Bennett & Offord, 2001; Campbell, Shaw, & Gilliom, 2000). These early problematic behaviors are associated with several severe negative outcomes in later adolescence and adulthood, including school dropout and unemployment (Fergusson & Horwood, 1998; Jessor, 1998; Loeber & Dishion, 1983). Additionally, aggressive and disruptive behavior identified as early as kindergarten has been shown to predict later delinquent behavior and substance use (Hill, Lochman, Coie, Greenberg, & the Conduct Problems Prevention Research Group [CPPRG], 2004; Petras, Chilcoat, Leaf, Ialongo, & Kellam, 2004; Petras et al., 2005), suggesting that problem behaviors observed early in childhood can persist across development (Loeber, 1982; Loeber & Hay, 1997; Moffitt, 1993). Thus, early intervention aimed at high-risk children is important (Lochman & CPPRG, 1995), and the use of accurate and reliable screening tools to identify children for these targeted prevention programs is essential (Jones, Dodge, Foster, Nix, & CPPRG, 2002).

Brief screening measures distributed to parents, teachers, and other caregivers are often used as an initial step in identifying children who will benefit from prevention programs targeting early behavior problems (CPPRG, 1999; Feil, Walker, & Severson, 1995). Studies examining the predictive utility of these screening tools showed that combining parent and teacher ratings of behavior problems during kindergarten predicted more difficulties in interactions with peers and teachers (Wehby, Dodge, Valente, & CPPRG, 1993), higher levels of delinquency (Hill et al., 2004), and lower levels of social competence (Lochman & CPPRG, 1995) in first grade. Additionally, parent and teacher ratings of child aggressiveness and hyperactivity in preschool predicted child behavior problems at the end of preschool (Doctoroff & Arnold, 2004) and 5 years later (Stormont, 2000). However, these multireporter screening measures can be costly and time-consuming. It is therefore important to examine the predictive utility of brief single-rater screening measures, as these may represent a cost- and time-efficient method to identify children at risk for later behavior problems and associated negative outcomes.

Teacher Ratings of Early Behavior Problems

Teacher ratings are a promising resource for the implementation of brief, single-informant screening procedures. Focusing on teacher ratings in the school context provides a highly efficient and cost-effective method of identifying early behavior problems, as compared with gathering information from parents or peers, as a teacher is able to rate an entire classroom in one sitting (Petras et al., 2005). Furthermore, teacher ratings of aggressive and disruptive child behavior in first grade predicted difficulties in classroom behavior, academic achievement, and social adjustment in third grade (Flanagan, Bierman, Kam, &

CPPRG, 2003). Additionally, teacher nominations were more accurate than parent nominations in predicting which children would develop behavior problems 1 year later (Dwyer, Nicholson, & Battistutta, 2006). This research suggests that teachers may be an ideal source for identifying children who are likely to show later behavior problems and may therefore benefit from participation in preventive interventions. Additionally, several large-scale intervention programs use teacher ratings as the first step in a multiple-gating assessment procedure (e.g., the Fast Track project; Lochman & CPPRG, 1995). Parents of children identified as higher risk based on these teacher ratings can then be contacted to provide their own ratings of their children's behavior. Given the wide use of teacher ratings, it is important to determine the accuracy and predictive utility of these screening measures.

The Teacher Observation of Classroom Adaptation–Revised (TOCA-R; Werthamer-Larsson, Kellam, & Wheeler, 1991) is a commonly used teacher rating screening tool. The Authority Acceptance (AA) scale of this measure includes 10 items asking teachers to rate the frequency of their students' behavior problems in the classroom. Scores on the TOCA-R collected during first through fifth grade have been shown to predict later (through age 18) violent and antisocial behavior in both boys and girls as identified in court records (Petras et al., 2004, 2005). In a recent study (Bradshaw, Schaeffer, Petras, & Jalongo, 2010), children were classified into one of three early starter aggressive-disruptive behavior trajectory groups on the basis of their TOCA-R scores during first through fifth grades: chronic high, low-moderate (for girls), or increasing (for boys). As compared with children who did not display behavior problems in elementary school, children in these early starter trajectory groups were found to be at risk for a greater number of negative nonaggressive life outcomes at age 19–20, including early pregnancy and unemployment (for girls) and high school dropout (for boys). Thus, the TOCA-R has demonstrated predictive validity to later problematic outcomes when examined across multiple assessment periods, but its utility as a screener at a single time point early in children's development (e.g., prior to first grade) is less well established.

Only a few studies have examined the ability of screening measures like the TOCA-R to predict behavior beyond elementary school (cf. Bradshaw et al., 2010; Petras et al., 2004, 2005). Additionally, many of these studies have only examined outcomes in one particular domain or context (e.g., violent behavior, behavior in the classroom, nonaggressive life outcomes) and at one particular point in time. Thus, it is unknown how well early teacher ratings of aggressive and disruptive behaviors, assessed at a single time point, predict behaviors both within and external to the school context, and how these predictions may extend across different developmental periods. Furthermore, it remains unclear how broadly or narrowly this screening measure could be used. For instance, it could be that this measure is a broad predictor, capturing multiple outcomes (e.g., clinical diagnoses of externalizing disorders, substance use, school problems). Conversely, it could be that this measure is a narrow predictor, only addressing school outcomes. As a result of this uncertainty, the utility of this measure to identify children who would benefit from particular prevention programs is still unknown.

Goal of the Current Study

Previous studies examining the predictive validity of early screening measures of behavior problems such as the TOCA-R have been limited in that they have mostly examined outcomes in one domain or one context (e.g., behavior problems at school), and at one point in time (e.g., third grade). Few studies have examined how well ratings of early behavior problems collected at a single time point predict later negative outcomes and adjustment problems across multiple domains and across distinct developmental periods (i.e., late childhood, middle adolescence, and late adolescence). By examining a variety of outcomes

across several developmental periods, it is possible to suggest more targeted interventions for at-risk children. For instance, it may be that high scores on the TOCA-R in kindergarten would predict more school-based behavior problems during late childhood and more delinquency and substance use problems during adolescence. These findings might suggest that for children identified as at risk based on TOCA-R ratings in kindergarten, school-based interventions addressing behavioral and emotional regulation in the classroom would be needed during childhood, whereas more community-based interventions addressing delinquency and substance use would be appropriate during adolescence. To date, no studies using early screening measures have addressed this issue regarding the timing and type of interventions indicated for at-risk children.

Therefore, the goal of the current study was to examine the ability of the TOCA-R measured during kindergarten to predict outcomes at the end of elementary (sixth grade), middle (eighth or ninth grade), and high (11th or 12th grade) school in the following domains: (a) school (behavior problems, social competence, and academic and disciplinary difficulties); (b) clinical diagnoses of externalizing disorders (i.e., attention-deficit/hyperactivity disorder [ADHD], conduct disorder [CD], and oppositional defiant disorder [ODD]); and (c) substance use (at the end of middle and high school only). In sum, the goal of the current study was to comprehensively examine the prediction of multiple outcomes across multiple domains and developmental periods from the TOCA-R, administered at one time point in kindergarten. To date, no studies have provided this comprehensive evaluation of the predictive utility of the TOCA-R.

Method

Participants

Fast Track project—Participants came from a community-based sample of children drawn from the Fast Track project, a longitudinal multisite investigation of the development and prevention of childhood conduct problems (CPPRG, 1992, 2000). Schools within four sites (Durham, NC; Nashville, TN; Seattle, WA; and rural Pennsylvania) were identified as high risk based on crime and poverty statistics of the neighborhoods that they served. Within each site, schools were divided into sets matched for demographics (size, percentage free or reduced lunch, ethnic composition), and the sets were randomly assigned to control and intervention groups. Using a multiple-gating screening procedure that combined teacher and parent ratings of disruptive behavior, 9,594 kindergarteners across three cohorts (1991–1993) from 55 schools were screened initially for classroom conduct problems by teachers using the AA score of the TOCA-R (Werthamer-Larsson et al., 1991; see also Lochman & CPPRG, 1995, for more details regarding screening procedures). The AA scale of the TOCA-R includes 10 items asking teachers to rate the frequency of their students' behavior problems in the classroom. Those children scoring in the top 40% on the TOCA-R within cohort and site were then solicited for the next stage of screening for home behavior problems by their parents, using items from the Child Behavior Checklist (Achenbach, 1991a) and similar scales, and 91% agreed to participate ($n = 3,274$). The teacher and parent screening scores were then standardized and summed to yield a total severity-of-risk screen score. Children were selected for inclusion into the high-risk sample on the basis of this screen score, moving from the highest score downward until desired sample sizes were reached within sites, cohorts, and groups. Deviations were made when a child failed to matriculate in the first grade at a core school ($n = 59$) or refused to participate ($n = 75$) or to accommodate a rule that no child would be the only girl in an intervention group. The outcome was that 891 children (control = 446, intervention = 445) participated.

In addition to the high-risk sample of 891 children, a stratified normative sample of 387 children was identified to represent the population-normative range of risk scores and was

followed over time. This normative sample was selected from the control schools, such that 100 kindergarten children were selected at each site (except for Seattle, WA, where only 87 children were selected). Participants in the normative sample were stratified to represent the population according to race, gender, and level of teacher-reported behavior problems (10 children at each decile of the distribution of scores from the TOCA-R). The normative sample included a portion of high-risk control group children to the proportional degree that they were represented in the school population. Written consent from parents and verbal assent from children were obtained. Parents were paid \$75 for completing the summer interviews, and teachers were compensated \$10 per child for completing the measures. The Institutional Review Boards of the participating universities approved all study procedures.

Sample description—The current study used data from the high-risk control and normative groups. Participants from the high-risk intervention sample were not included in this study. Because 79 of those recruited for the high-risk control group were also included as part of the normative sample, the total sample included 754 participants. However, two children in this sample were missing TOCA-R scores and were therefore excluded from the current analyses, yielding a final sample of 752 children. Children were, on average, 6.55 years old ($SD = .43$) at the start of the Fast Track project. As would be expected given the higher prevalence of conduct problems documented among boys as compared with girls (Hinshaw & Lee, 2003), 57.7% of the sample was male. Reflecting the ethnic diversity in the populations at the four sites, the majority of the sample was either Caucasian (49.9%) or African American (46.3%), with 3.8% of the sample representing other ethnic groups (e.g., Hispanic, Asian). Due to the multisite sampling design of the Fast Track project, race and urban/rural status were confounded, as nearly all of the African American participants lived in urban areas. In fact, less than 1% of the entire sample consisted of African Americans living in rural communities. Thus, for the current study, a race/urban status variable was examined representing three groups: urban African Americans (46.0%) urban Caucasians (24.2%), and rural Caucasians (25.7%). For analyses examining ethnicity and race/urban status, other ethnic minorities were not included due to the small sample sizes in these groups. For this final sample of 752 children, 147 teachers provided TOCA-R ratings in kindergarten.

Procedure

Annual home interviews were conducted with primary caregivers (typically mothers) and children. Interviews began during the summer before children's entry to first grade and concluded 2 years after the child completed (or would have completed) 12th grade. Caregivers and children completed the interviews separately with two different interviewers over the course of approximately 2 hr. Measures given during these interviews assessed several domains, including parenting behaviors, child behavior problems, family functioning, parent-child relationship quality, peer relationships, academic achievement, and characteristics of the broader neighborhood. Measures included in the current study are described below.

We chose the specific grades included in this study because they aligned with assessments from the Fast Track project that fell within the developmental periods we aimed to address. The timing of assessments during the Fast Track project also took into account the length of the assessment battery at each year. For instance, in years when the lengthy Computerized Diagnostic Interview Schedule for Children (CDISC; Shaffer & Fisher, 1997) was administered, the remainder of the assessment battery was trimmed significantly to reduce participant burden. For this reason, not all measures were administered at all years, leading to the uneven timing of assessments.

Measures

The TOCA-R—The AA scale of the TOCA-R (Werthamer-Larsson et al., 1991) assessed during kindergarten was used as the predictor in all analyses. The entire TOCA-R is a 43-item questionnaire designed for teachers to assess authority and acceptance behavior, concentration problems, and shy behavior relevant to the child's behavior in a classroom situation on a 6-point Likert-type scale (0 = *never*, 1 = *rarely*, 2 = *sometimes*, 3 = *often*, 4 = *very often*, and 5 = *always*). The AA scale includes 10 items representing aggressive and disruptive behavior problems (e.g., “breaks rules,” “harms others,” and “takes others’ property”). Scores on the TOCA-R are commonly summed to create a composite sum score (e.g., Petras et al., 2004, 2005). Therefore, for the current study, sum scores were created on the basis of the 10-item AA scale of the TOCA-R (hereafter referred to as *TOCA-R sum scores*).

School outcomes—Several school-related outcomes were measured at sixth, eighth, and 11th grades, including behavior problems in school, social competence, and academic and disciplinary difficulties. Teacher-rated child behavior problems in school were measured in sixth and eighth grade with the T-score of the Externalizing subscale of the Teacher's Report Form (TRF; Achenbach, 1991b). The Externalizing subscale of the TRF is a 34-item measure that asked teachers to report on the child's level of multiple problematic behaviors (e.g., disobedient, disruptive, physically aggressive, explosive, stubborn, truant). Items were scored on a 3-point Likert scale, ranging from 0 (*not true [as far as you know]*) to 2 (*very true or often true*). Cronbach's alpha coefficients for this subscale were .96 in sixth grade and .97 in eighth grade, indicating strong internal consistency.

Teacher-rated social competence in sixth and eighth grade was measured with the Teacher Social Competence (TSC) scale, which was developed by the Fast Track project (CPPRG, 1995). The TSC is a 17-item measure that assessed child competence in academic behavior, prosocial skills, and emotional regulation (e.g., performs academically at grade level, handles disagreements in a positive way, cooperates with others, initiates interactions in a positive manner, recognizes and labels feelings, stops and calms down when excited). Items on this measure asked teachers to rate the frequency of these social behaviors on a 6-point scale, ranging from 0 (*almost never*) to 5 (*almost always*). The total score on the TSC was calculated as the mean of all 17 items. Internal consistency of the TSC was strong (Cronbach's α = .94 in sixth grade and .95 in eighth grade).

Parent- and child-reported academic and disciplinary difficulties in sixth, eighth, and 11th grade were measured with the School Adjustment scale, which was developed by the Fast Track project (CPPRG, 1997a, 1997b). The parent report version of this measure included 18 items that evaluated the child's past school year in terms of academic performance, disciplinary problems, and general worries about school (e.g., school year difficult for child, school work was really hard for child, other kids tried to make child do bad things, child got into trouble by breaking rules). Two items related to the parent's contact with the school and teachers were not included in this study. This measure used a 5-point response scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The total academic and disciplinary difficulties-parent report score was calculated as the mean of all 16 items. Cronbach's alpha coefficients were .90 in sixth and eighth grades and .84 in 11th grade, indicating adequate internal consistency.

The child report version of the School Adjustment scale included 20 items related to the child's academic and disciplinary difficulties, relationships with other students, and general aspects about the school and teachers. The subscale pertaining to academic and disciplinary difficulties was used for the current study. Example items on this subscale included “the school year was difficult,” “school work was really hard,” “I got into trouble this year,” and

“teachers were on me because I broke rules.” The eight items on this subscale were on a 5-point scale ranging from 1 (*never true*) to 5 (*always true*). The total academic and disciplinary difficulties-child report score was calculated as the mean of all eight items. Internal consistency of this subscale was adequate (Cronbach’s $\alpha = .75$ in sixth grade, $.76$ in eighth grade, and $.74$ in 11th grade).

Externalizing disorders—The externalizing diagnoses considered for this study included ADHD (combined type), ODD, and CD. Diagnoses of externalizing disorders in sixth, ninth, and 12th grade were assessed with the CDISC (Shaffer & Fisher, 1997). The CDISC is widely used to assess *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*, psychiatric symptoms and diagnoses in children and adolescents aged 6–17 years. We analyzed both binary diagnosis variables (i.e., child met diagnostic criteria for a particular disorder or not) and symptom count variables (i.e., a count of how many symptoms of a particular disorder were endorsed). Both parent- and child report versions of this measure were used in the current study, such that a child was considered to have displayed a particular symptom or received a diagnosis in the past year if either the parent or the child or both endorsed that symptom or if the report of the parent or the child or both met criteria for a diagnosis. This approach was used because previous research has shown that parents and children provide unique information regarding ADHD, ODD, and CD diagnostic criteria (Colins, Vermeiren, Schuttyen, Broekaert, & Soye, 2008). For analyses, we examined both diagnoses and symptom counts for each disorder for each grade separately and combined (i.e., a diagnosis of ADHD, ODD, CD, or any externalizing disorder at sixth, ninth, and 12th grade; total symptom counts of ADHD, ODD, CD, or any externalizing disorder at sixth, ninth, and 12th grade).

Substance use—Information regarding children’s substance use was collected with the Tobacco, Alcohol and Drugs measure. This measure was adapted by the Fast Track project from the substance use section of the Self-Administered Youth Questionnaire from the National Longitudinal Survey of Youth 1997 (Bureau of Labor Statistics, 2002; Elliot, Huizinga, & Ageton, 1985). We examined child report of tobacco, alcohol, and marijuana use at eighth and 12th grade. Use of other illicit drugs was not included due to extremely low endorsement of these substances in the sample. For analyses, we examined the number of days children reported smoking cigarettes in the past month (range = 0–30), the number of days they reported drinking alcohol in the past year (range = 0–365), and the number of times children used marijuana in the past month (range = 0–100). Additionally, we examined the number of drinks children reported drinking each time they drank alcohol as well as the number of days children engaged in binge drinking (had 5+ drinks in a row) in the past year (range = 0–365). We also multiplied the number of days children drank alcohol with the number of drinks children had each time they drank alcohol to obtain a measure of intensity of alcohol use. All substance use variables were square-root transformed for analyses due to nonnormality and skewness in the data (Tabachnick & Fidell, 2001).

Missing data—The current analyses examined school outcomes measured at the end of sixth, eighth, and 11th grades; externalizing diagnoses measured at the end of sixth, ninth, and 12th grades; and substance use measured at the end of eighth and 12th grades. For school outcomes, 228 (30.3%) children were missing data in sixth grade, 259 (34.4%) in eighth grade, and 282 (37.5%) in 11th grade. For externalizing diagnoses, 119 (15.8%) were missing data in sixth grade, 185 (24.6%) in ninth grade, and 245 (32.6%) in 12th grade. For substance use, 179 (23.8%) were missing data in eighth grade and 205 (27.3%) in 12th grade. Children with missing data on school outcomes were more likely to be in the high-risk control group than in the normative group: sixth grade, $\chi^2(1, N = 752) = 6.16, p < .05$; eighth grade, $\chi^2(1, N = 752) = 6.30, p < .05$; 11th grade, $\chi^2(1, N = 752) = 6.39, p < .05$. For

externalizing diagnoses, children with missing data in 12th grade were more likely to be urban African American, $\chi^2(2, N = 752) = 9.39, p < .01$. Additionally, for substance use, children with missing data in 12th grade were more likely to be in the high-risk control group, $\chi^2(1, N = 752) = 16.08, p < .001$; urban African American, $\chi^2(2, N = 752) = 10.43, p < .01$; and male, $\chi^2(1, N = 752) = 3.94, p < .05$. There were no other significant differences in attrition by group (high-risk control vs. normative), race/urban status, gender, or TOCA-R sum score at any of the other time points.

Analysis Plan

Analyses were conducted in SPSS version 14.0 (for descriptive statistics) and Mplus version 6.0 (Muthén & Muthén, 2010). To account for the oversampling of high-risk children in the Fast Track project and to increase generalizability to the population, we used a probability weight based on group (normative vs. high-risk control) that had been previously calculated for all normative and high-risk control group participants (see Jones et al., 2002, for a description of the creation and calculation of this weighting variable). Gender, race/urban status, and risk group (normative vs. high-risk control) were included as covariates in all regression analyses. The research design of the Fast Track project involved children who were nested within classrooms. For example, the 891 high-risk children recruited for this project were nested within 401 first-grade classrooms. However, by Grade 3, these same children were nested in 527 classrooms due to transfers and relocations, and some intervention and control children were in the same classrooms at that time. Therefore, we determined that it was not appropriate or possible to account for nesting within classrooms or schools in the outcome data examined for this study (see CPPRG, 2002, for further details on the nested structure of the data in the Fast Track project).

A series of linear regression analyses were conducted to test the prediction of school outcomes (behavior problems, social competence, and adjustment) from the TOCA-R sum scores. Covariates and TOCA-R sum scores were entered simultaneously as predictors of the various outcomes. Full information maximum likelihood was used to handle missing data (amount of missing data ranged from 15.8% to 37.5% across the variables). We used a maximum likelihood estimator that calculated robust standard errors, which provides valid standard error estimates when variables are nonnormal (Asparouhov & Muthén, 2006). For the prediction of externalizing diagnoses (i.e., ADHD, ODD, and CD), we conducted a series of logistic regressions for the binary diagnostic variables as well as linear regression analyses for the continuous symptom count variables. To predict substance use outcomes, we used a series of censored linear regressions. We used censored regression because the distributions of the substance use variables were continuous, positively skewed with an abundance of zeroes, and theoretically left-censored at zero. Censored regressions are used commonly with variables representing behaviors that do not occur frequently in the general population (e.g., substance use among children, serious delinquency), as data transformations that attempt to normalize the distribution of variables are ineffective in managing an abundance of zeros (Long, 1997). The proportion of variance explained (R^2) was used as a measure of the effect sizes for the TOCA-R sum scores in the prediction of outcomes (Cohen, Cohen, West, & Aiken, 2003).

Results

Descriptive Statistics

Table 1 provides sample sizes, means, and standard deviations for all continuous outcomes for each year measured. For the binary diagnosis outcomes, the percentage and number of children receiving that diagnosis are presented for each year measured. The zero-order correlations between the TOCA-R sum scores and the school outcomes and between the

TOCA-R sum scores and externalizing disorder outcomes are presented in Table 2 and Table 3, respectively. As seen in these tables, the TOCA-R sum scores were significantly correlated with almost all outcomes at all time points. Specifically, higher TOCA-R sum scores were associated with more teacher-reported behavior problems, lower teacher-reported social competence, and lower parent- and child-reported school adjustment. In terms of externalizing diagnoses, higher TOCA-R sum scores were related to higher diagnosis rates (with the exception of diagnoses of CD in 12th grade) and symptom counts across all disorders.

Covariates

Many of the covariates included in the regression models were significant predictors of several of the outcomes included in the current study. The following is a summary of the most consistent covariate effects observed in the analyses (the full results are available from the first author). For teacher-rated school outcomes, children living in urban areas (urban Caucasians and urban African Americans) were rated as higher on behavior problems as compared with rural Caucasians ($\beta = .19, p < .001$; $\beta = .30, p < .001$, for sixth and eighth grade, respectively) as were urban African Americans when compared with urban Caucasians ($\beta s = -.22, p < .001$ for both sixth and eighth grade). Additionally, children living in urban areas were rated as lower on social competence as compared with rural Caucasians ($\beta = -.18, p < .001$; $\beta = -.25, p < .001$, for sixth and eighth grade, respectively), as were urban African Americans when compared with urban Caucasians ($\beta = .28, p < .001$; $\beta = .26, p < .001$, for sixth and eighth grade, respectively). For child-rated school outcomes, children living in urban areas were lower on school adjustment than rural Caucasians ($\beta = -.15, p < .01$; $\beta = -.21, p < .001$; $\beta = -.15, p < .05$, for sixth, eighth, and 11th grade, respectively), and females were higher on school adjustment than males ($\beta = .20, p < .001$; $\beta = .18, p < .01$; $\beta = .21, p < .01$, for sixth, eighth, and 11th grade, respectively).

Males were also consistently higher on symptom counts of ADHD and CD as compared with females (βs range from $-.12$ to $-.30$, all $p s < .05$). Additionally, children living in urban areas had higher symptom counts of CD than rural Caucasians ($\beta = .18, p < .001$; $\beta = .21, p < .001$; $\beta = .11, p < .05$, for sixth, ninth, and 12th grade, respectively), as did urban African Americans when compared with urban Caucasians ($\beta = -.11, p < .05$; $\beta = -.12, p < .05$; $\beta = -.10, p < .05$, for sixth, ninth, and 12th grade, respectively). For substance use in 12th grade, males reported higher levels of use (except for number of drinks consumed each time they drank alcohol, where there was no gender difference) than females (βs range from $-.14$ to $-.26$, all $p s < .05$).

School Outcomes

Table 4 presents the results from the linear regression analyses predicting school outcomes at the end of elementary, middle, and high school from the TOCA-R sum scores. Above and beyond the covariates, teacher-reported behavior problems during kindergarten were prospectively associated with almost all teacher-, parent- and child-rated school outcomes at sixth, eighth, and 11th grades; however, child-rated school adjustment in eighth and 11th grades were not significantly associated. For example, a one standard deviation increase in the TOCA-R sum score predicted a 0.32-standard deviation increase in teacher-rated behavior problems in sixth grade and a 0.29-standard deviation unit increase in teacher-rated behavior problems in eighth grade. As seen in Table 4, higher kindergarten TOCA-R sum scores predicted higher levels of teacher-rated child behavior problems and lower teacher-rated child social competence at sixth and eighth grades. Higher TOCA-R sum scores also predicted lower parent- and child-rated school adjustment at sixth, eighth, and 11th grades (however, the school adjustment coefficients at eighth and 11th grades were significant according to parent report only). Across reporters and time points, early teacher ratings of

behavior problems using the TOCA-R explained approximately 2%–5% of the variance in later behavior problems at school, social competence, and school adjustment, after controlling for the covariates.

Diagnosis of Externalizing Disorders (ADHD, ODD, and CD)

Of the externalizing disorders, the TOCA-R sum scores predicted only diagnosis of ADHD in ninth grade ($\beta = .30$, $SE = .12$, $p < .01$, $OR = 1.09$, 95% CI for OR [1.02, 1.16]). In other words, the odds of receiving an ADHD diagnosis in ninth grade increased between 1.02 and 1.16 times for every one-unit increase in the TOCA-R sum score. The TOCA-R sum scores did not predict any other diagnosis at any grade, nor did they predict any cumulative diagnosis of ADHD, CD, ODD, or any externalizing disorder (ADHD, CD, and ODD combined).

As expected, utilizing symptom counts, compared with predicting dichotomous diagnoses, provided better predictive utility (MacCallum, Zhang, Preacher, & Rucker, 2002). As seen in Table 5, after controlling for the covariates, TOCA-R sum scores in kindergarten prospectively predicted the number of ADHD symptoms endorsed by the parent or the child or both in sixth grade, such that a one-standard deviation increase in TOCA-R sum scores predicted a 0.14-standard deviation unit increase in ADHD symptoms in sixth grade. Additionally, a one-standard deviation increase in TOCA-R sum scores predicted a 0.14-standard deviation unit increase in ninth grade CD symptoms, a 0.12-standard deviation unit increase in total ADHD symptoms (sixth, ninth, and 12th grades combined), and a 0.12-standard deviation unit increase in total externalizing symptoms (ADHD, CD, and ODD combined) across all three grades combined.

Substance Use

Results from the censored linear regressions of the substance use variables indicated that, after controlling for the covariates, TOCA-R sum scores in kindergarten were unrelated to any of the substance use outcomes at either eighth or 12th grade (all regression coefficients were not significant at the .05 level), with the exception of cigarette use in 12th grade. This finding indicated that higher TOCA-R sum scores predicted more days smoked cigarettes in the past month ($\beta = .20$, $p < .05$, 95% CI [.03, .36]).

Replication of Findings

It is important to note that sum scores, as reflective of classical test theory, provide equal weight to each behavior on a particular screening measure. Therefore, sum scores may not accurately capture the true severity of a child's aggressive and disruptive behaviors, as some behaviors (e.g., being stubborn and disobedient) are considered less severe than others (e.g., fighting and harming others). Several researchers have argued that item response theory (IRT), which explicitly assesses differential severity across items (and thus behaviors), may be a more appropriate method for modeling the items used in brief screening scales for behavior problems (Embretson & Reise, 2000). In the current study, the findings from the TOCA-R sum scores were replicated with IRT-based TOCA-R scores (see Wu et al., 2012, for a description of the creation of the IRT TOCA-R scores). The IRT scores demonstrated a similar pattern of prediction to later school outcomes, externalizing diagnoses and symptom counts, and substance use outcomes as seen with the TOCA-R sum scores. Furthermore, the standard errors and confidence intervals were relatively similar between these two sets of analyses. However, the IRT TOCA-R scores tended to account for slightly more variance in the majority of the outcome variables as compared with the TOCA-R sum scores.

We also ran post hoc regression analyses without controlling for the effects of the covariates. A similar pattern of findings was observed in these analyses, as higher TOCA-R

sum scores predicted more behavior problems at school, lower social skills, and poorer school adjustment. Additionally, in these analyses higher TOCA-R sum scores predicted more cigarette use in eighth and 12th grades. Higher TOCA-R sum scores also predicted higher externalizing disorder symptom counts (for ADHD, CD, and ODD) as well as higher odds of receiving an externalizing diagnosis. Lastly, although it was not possible for us to account for nesting within classrooms or schools in the outcome data examined for this study, we conducted post hoc regression analyses accounting for clustering in the TOCA-R sum scores collected during kindergarten (clusters were based on classrooms). A similar pattern of findings was observed in these analyses as was seen in the initial regression analyses, with similar beta coefficients obtained for all outcomes.

Discussion

The goal of the current study was to examine the predictive utility of the TOCA-R (specifically, the AA scale), a commonly used teacher screening measure, administered during kindergarten. We extended prior research by examining the TOCA-R at a single time point at an earlier age than examined in previous studies. Additionally, we examined a broader range of outcomes across a broader range of developmental periods than has been documented in the extant literature. Our results suggested that higher TOCA-R kindergarten scores were associated with more behavior problems at school, lower social skills, and poorer school adjustment reported by multiple informants (teacher, parent, and child) at the end of elementary, middle, and high school. The TOCA-R sum scores were also related, although somewhat more inconsistently, to the odds of an ADHD diagnosis, as well as ADHD, CD, and externalizing disorder symptom counts, but not to an ODD diagnosis or symptoms or any substance use outcomes.

The findings from the current study indicated that early teacher ratings consistently predicted later school outcomes, including school adjustment and social competence, as late as the end of high school. The prediction of ADHD diagnosis in ninth grade and symptoms in the sixth and ninth grades further supports the ability of the TOCA-R to predict outcomes mainly within the school setting, as teachers frequently play an important role in initial screenings for ADHD symptoms (Snider, Busch, & Arrowood, 2003; Snider, Frankengerger, & Aspen-son, 2000). Additionally, ADHD symptoms are frequently first observed and most troublesome in the classroom where children are required to sustain their attention and refrain from hyperactive or impulsive behaviors (Barkley, 2003). Overall, these findings suggest that the TOCA-R may be most useful when predicting problematic behaviors that occur within the classroom.

A strength of the current study is the inclusion of a racially and regionally diverse sample. Prior research has demonstrated that the TOCA-R administered in kindergarten exhibits group differences, as boys and African Americans have a higher overall mean, and therefore more frequent behavior problems, than girls or Caucasians, respectively (Koth, Bradshaw, & Leaf, 2009). In previous work with the same larger sample as the current study (Wu et al., 2012), item bias by gender was revealed in the TOCA-R, such that at equivalent levels of latent behavior problems, males received more endorsements of overt behaviors (e.g., harming others, fights, breaks things) from teachers, whereas females received more endorsements of nonphysical behaviors (e.g., stubborn, takes property, lies). Moreover, overt behaviors tended to be better at distinguishing among levels of latent behavior problems for males, and covert behaviors tended to be better at distinguishing among levels of latent behavior problems for females. However, given the similarity in the results obtained from the TOCA-R sum scores and the IRT scores, our findings suggest that the gender bias identified by Wu and colleagues (2012) may not affect the predictive utility of this screening measure.

The importance of including diverse samples in studies examining problem behaviors is underscored by the significant covariate effects in the current study. For instance, children living in urban areas were found to be rated as higher in behavior problems, lower in social competence and school adjustment, and higher in endorsed symptoms of CD than children in rural areas. Compared with urban Caucasians, urban African Americans also received the lowest/highest scores, indicating that this population is particularly at risk for the development of these problematic behaviors. Additionally, when compared with females, males were found to be rated as lower in school adjustment and higher in endorsed symptoms of ADHD and CD. These findings are consistent with other studies indicating that males, ethnic minorities, and children living in urban areas are at most risk for behavior problems (Hinshaw & Lee, 2003). It is therefore important that future studies continue to incorporate diverse samples (i.e., in terms of race, gender, and geographical location). Overall, the findings from the current study suggest that in applied settings, such as schools or community mental health centers, or when a quick screening of behavioral problems is needed, the calculation of a sum score on the TOCA-R may be sufficient to determine which children are at risk for later problems.

The post hoc regression analyses without controlling for the effects of the covariates allowed for exploration of the utility of the TOCA-R in applied settings, as it is unlikely that professionals in these settings would have the resources or information needed to consider these demographic factors in their screening procedures. A similar pattern of findings was observed in these analyses, providing support for the utility of the TOCA-R as a kindergarten screener in practical, applied settings (e.g., schools, community mental health centers, private practice offices). However, it is important to note that the findings regarding diagnostic outcomes were not significant in the regression analyses with covariates, suggesting caution in using the TOCA-R and similar brief screening measures to predict whether a child will receive a diagnosis. Such diagnostic decisions require careful, evidence-based assessments (Mash & Hunsley, 2005) along with the inclusion of demographic variables known to account for a significant amount of variance in these disorders (e.g., race, gender, etc.).

Clinical Implications and Future Directions

Generally, studies examining the development of maladaptive and problematic behaviors have focused mainly on elementary school, with little attention to earlier behaviors during preschool and kindergarten (Campbell et al., 2000). The current study shows that early behavior problems identified in kindergarten with the TOCA-R can reliably predict a range of adverse school-related outcomes in late childhood, middle adolescence, and late adolescence. Therefore, the TOCA-R may be most effective in identifying which children will benefit from prevention programs targeting problems within the school environment and across the school years (e.g., elementary, middle, and high school). It may be that classroom-based intervention programs would be best suited to implement with children at high levels of TOCA-R sum scores. Several existing interventions target improvements in school behavior. Examples include Promoting Alternative Thinking Strategies (Kusche & Greenberg, 1994), Coping Power (Lochman & Wells, 2003), and the Good Behavior Game (Barrish, Saunders, & Montrose, 1969).

The identification of children at risk for a range of negative outcomes beyond the school context may require the application of additional, broader screening measures. Specifically, other screening measures may be needed to identify children at risk for later outcomes that occur in contexts outside of the school setting (i.e., diagnoses of externalizing disorders and substance use), which may require information from additional reporters (Hill et al., 2004; Kerr, Lunkenheimer, & Olson, 2007; Lochman & CPPRG, 1995). Cost-benefit analyses are needed to determine whether adding additional raters is beneficial and outweighs the added

cost of both time and resources needed to collect these ratings. Moreover, given previous research demonstrating that the TOCA-R did not adequately cover the lower range of problem behaviors (Wu et al., 2012), it may be that adding additional items to the TOCA-R may improve specificity of prediction to later ages. For example, recent research has suggested that poor self-control observed as early as age 3 predicts substance use disorders at age 32 (Moffitt et al., 2011). Improving prediction to later ages may involve creating measures that blend traditional symptom checklists with measures that tap the underlying processes (e.g., poor self-control) that increase risk for those symptoms. Future research should continue to explore the optimal combination of screening measures to capture the full range of negative outcomes that could be experienced by children and adolescents.

The findings from the current study indicate that early behavior problems that place children at risk for later adverse school-related outcomes can be identified as early as kindergarten. The results from this study therefore suggest that kindergarten may be an appropriate time to begin delivering prevention programs that address aggressive and disruptive behaviors in children. The prompt application of effective prevention programs with these children may interrupt the progression of behavior problems in school before they become entrenched and difficult to change (Lochman & CPPRG, 1995). However, the ability of these programs to effectively address these negative behaviors relies on the accurate identification of children who need these preventive interventions (Hill et al., 2004; O'Connell, Boat, & Warner, 2009). Future studies should therefore continue to explore the predictive validity of methods used in the early identification of at-risk children (Keenan & Wakschlag, 2002; Wakschlag & Keenan, 2001).

The timing of screening measures is an important consideration, and the administration of these measures may need to be altered depending on the purpose of the screening. For instance, previous studies have suggested that prediction to later outcomes is enhanced when TOCA-R ratings collected during first grade are used (Flanagan et al., 2003; Hill et al., 2004). Petras and colleagues (2004, 2005) have reported that the spring of third grade for boys and fifth grade for girls are the optimal times for minimizing both false-negative and false-positive identifications of children at risk for later problems based on TOCA-R sum scores. The findings from the current study suggest that for the general identification of children at risk for a variety of school-based negative outcomes at various developmental stages, administration of the TOCA-R during kindergarten is warranted. Future research should seek to refine conclusions regarding the ideal timing of TOCA-R administration.

Limitations

Limitations of the current study should be noted. First, the early behavior ratings examined in the current study only described a minimal amount of the variance (2%–5%) in later teacher-rated behavior problems, social competence, and school adjustment. More research is needed to identify additional factors that may explain more of the variance in these behavior and adjustment difficulties (e.g., parenting, peer engagement in antisocial behaviors, environmental context). Additionally, although the current study is longitudinal, care should be taken to not assume causal relationships between TOCA-R sum scores and outcomes, as several third variables may influence this relationship (e.g., low or unchanging teacher expectations, negative perceptions of children's classroom behavior, and academic achievement).

Conclusion

In summary, this study illustrates the predictive validity of TOCA-R sum scores when predicting various outcomes, particularly those observed within the school setting, across a range of developmental periods. The ability of the TOCA-R to predict outcomes into late

adolescence speaks to the benefit of using a widely distributed brief teacher report screening instrument. The findings also indicate that children identified as at risk for later behavior problems experience difficulty with school, teachers, and their fellow classmates. Prevention programs working with at-risk children should therefore continue using strategies that promote the development of academic skills, prosocial interactions with peers and teachers, and overall positive attitudes toward school. Taken together, this study, as well as other studies examining teacher-reported screening tools, supports the adage to “catch behavior problems early,” before children are directed into a persistent negative life course trajectory.

Acknowledgments

This work was supported by National Institute of Mental Health (NIMH) Grants R18 MH48043, R18 MH50951, R18 MH50952, R18 MH50953, and by National Institute on Drug Abuse (NIDA) Grant 1 RC1 DA028248-01. The Center for Substance Abuse Prevention and NIDA also provided support for Fast Track through a memorandum of agreement with the NIMH. This work was also supported in part by Department of Education Grant S184U30002, NIMH Grants K05MH00797 and K05MH01027, and NIDA Grants DA16903, DA015226, and DA017589. Preparation of the manuscript was also supported by an Establishment Award from the Child & Family Research Institute awarded to Robert J. McMahon. A preliminary version of this article was presented at the Association for Psychological Science annual meeting, Boston, MA, May 2010.

References

- Achenbach, TM. Manual for the Child Behavior Checklist/4–18 and 1991 profile. Burlington: University of Vermont, Department of Psychiatry; 1991a.
- Achenbach, TM. Manual for the Teacher’s Report Form and 1991 profile. Burlington: University of Vermont, Department of Psychiatry; 1991b.
- Asparouhov, T.; Mut.en, B. Constructing covariates in multilevel regression. 2006. (Mplus Web Notes: No 11.) Available from: <http://www.statmodel.com>
- Barkley, RA. Attention-deficit/hyperactivity disorder. In: Mash, EJ.; Barkley, RA., editors. Child psychopathology. 2nd ed.. New York, NY: Guilford Press; 2003. p. 75-143.
- Barrish HH, Saunders M, Montrose WM. Good behavior game: Effects of individual contingencies for group consequences on disruptive behavior in a classroom. *Journal of Applied Behavior Analysis*. 1969; 2:119–124. [PubMed: 16795208]
- Bennett KJ, Offord DR. Screening for conduct problems: Does the predictive accuracy of conduct disorder symptoms improve with age? *Journal of the American Academy of Child & Adolescent Psychiatry*. 2001; 40:1418–1425. [PubMed: 11765287]
- Bradshaw CP, Schaeffer CM, Petras H, Ialongo N. Predicting negative life outcomes from early aggressive-disruptive behavior trajectories: Gender differences in maladaptation across life domains. *Journal of Youth and Adolescence*. 2010; 39:953–966. [PubMed: 19688587]
- Bureau of Labor Statistics, U.S. Department of Labor. National Longitudinal Survey of Youth 1997 cohort, 1997–2001. Produced by the National Opinion Research Center, the University of Chicago and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, OH: The Ohio State University; 2002.
- Campbell SB, Shaw DS, Gilliom M. Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology*. 2000; 12:467–488. [PubMed: 11014748]
- Cohen, J.; Cohen, P.; West, SG.; Aiken, LS. Applied multiple regression/correlation analysis for the behavioral sciences. 3rd ed.. Mahwah, NJ: Erlbaum; 2003.
- Colins O, Vermeiren R, Schuttyen G, Broekaert E, Soye V. Informant agreement in the assessment of disruptive behavior disorders in detained minors in Belgium: A diagnosis-level and symptom-level examination. *Journal of Clinical Psychiatry*. 2008; 69:141–148. [PubMed: 18312049]
- Conduct Problems Prevention Research Group. A developmental and clinical model for the prevention of conduct disorders: The FAST Track program. *Development and Psychopathology*. 1992; 4:509–527.

- Conduct Problems Prevention Research Group. Teacher–Social Competence Scale. 1995. Available from <http://www.fasttrackproject.org>
- Conduct Problems Prevention Research Group. School Adjustment–Child. 1997a. Available from <http://www.fasttrackproject.org>
- Conduct Problems Prevention Research Group. School Adjustment–Parent. 1997b. Available from <http://www.fasttrackproject.org>
- Conduct Problems Prevention Research Group. Initial impact of the Fast Track prevention trial for conduct problems: II. Classroom effects. *Journal of Consulting and Clinical Psychology*. 1999; 67:648–657. [PubMed: 10535231]
- Conduct Problems Prevention Research Group. Merging universal and indicated prevention programs: The Fast Track model. *Addictive Behaviors*. 2000; 25:913–927. [PubMed: 11125779]
- Conduct Problems Prevention Research Group. Evaluation of the first 3 years of the Fast Track prevention trial with children at high risk for adolescent conduct problems. *Journal of Abnormal Child Psychology*. 2002; 30:19–35. [PubMed: 11930969]
- Doctoroff GL, Arnold DH. Parent-rated externalizing behavior in preschoolers: The predictive utility of structured interviews, teacher reports, and classroom observations. *Journal of Clinical Child and Adolescent Psychology*. 2004; 33:813–818. [PubMed: 15498748]
- Dwyer SB, Nicholson JM, Battistutta D. Parent and teacher identification of children at risk of developing internalizing or externalizing mental health problems: A comparison of screening methods. *Prevention Science*. 2006; 7:343–357. [PubMed: 16850278]
- Elliot, DS.; Huizinga, D.; Ageton, S. Explaining delinquency and drug use. Beverly Hills, CA: Sage; 1985.
- Embretson, SE.; Reise, SP. Item response theory for psychologists. Mahwah, NJ: Erlbaum; 2000.
- Feil EG, Walker HM, Severson HH. The Early Screening Project for young children with behavior problems. *Journal of Emotional and Behavioral Disorders*. 1995; 3:194–202.
- Fergusson DM, Horwood LJ. Early conduct problems and later life opportunities. *Journal of Child Psychology and Psychiatry*. 1998; 39:1097–1108. [PubMed: 9844980]
- Flanagan KS, Bierman KL, Kam C-M. the Conduct Problems Prevention Research Group. Identifying at-risk children at school entry: The usefulness of multibehavioral problem profiles. *Journal of Clinical Child and Adolescent Psychology*. 2003; 32:396–407. [PubMed: 12881028]
- Hill LG, Lochman JE, Coie JD, Greenberg MT. the Conduct Problems Prevention Research Group. Effectiveness of early screening for externalizing problems: Issues of screening accuracy and utility. *Journal of Consulting and Clinical Psychology*. 2004; 72:809–820. [PubMed: 15482039]
- Hinshaw, SP.; Lee, SS. Conduct and oppositional defiant disorders. In: Mash, EJ.; Barkley, RA., editors. *Child psychopathology*. 2nd ed.. New York, NY: Guilford Press; 2003. p. 144-198.
- Jessor, R. New perspectives on adolescent risk behavior. In: Jessor, R., editor. *New perspectives on adolescent risk behavior*. Cambridge, UK: Cambridge University Press; 1998. p. 1-10.
- Jones D, Dodge KA, Foster EM, Nix R. Conduct Problems Prevention Research Group. Early identification of children at risk for costly mental health service use. *Prevention Science*. 2002; 3:247–256. [PubMed: 12458763]
- Keenan K, Wakschlag LS. Can a valid diagnosis of disruptive behavior disorder be made in preschool children? *American Journal of Psychiatry*. 2002; 159:351–358. [PubMed: 11869995]
- Kerr DCR, Lunkenheimer ES, Olson SL. Assessment of child problem behaviors by multiple informants: A longitudinal study from preschool to school entry. *Journal of Child Psychology and Psychiatry*. 2007; 48:967–975. [PubMed: 17914997]
- Koth CW, Bradshaw CP, Leaf PJ. Teacher Observation of Classroom Adaptation—Checklist: Development and factor structure. *Measurement and Evaluation in Counseling and Development*. 2009; 42:15–30.
- Kusche, CA.; Greenberg, MT. The PATHS curriculum. Seattle, WA: Developmental Research and Programs; 1994.
- Lochman JE. the Conduct Problems Prevention Research Group. Screening of child behavior problems for prevention programs at school entry. *Journal of Consulting and Clinical Psychology*. 1995; 63:549–559. [PubMed: 7673532]

- Lochman JE, Wells KC. Effectiveness of the Coping Power program and of classroom intervention with aggressive children: Outcomes at a 1-year follow-up. *Behavior Therapy*. 2003; 34:493–515.
- Loeber R. The stability of antisocial and delinquent child behavior: A review. *Child Development*. 1982; 53:1431–1446. [PubMed: 6756808]
- Loeber R, Dishion T. Early predictors of male delinquency: A review. *Psychological Bulletin*. 1983; 94:68–99. [PubMed: 6353467]
- Loeber R, Hay D. Key issues in the development of aggression and violence from childhood to early adulthood. *Annual Review of Psychology*. 1997; 48:371–410.
- Long, JS. Regression models for categorical and limited dependent variables. Thousand Oaks, CA: Sage; 1997.
- MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. *Psychological Methods*. 2002; 7:19–40. [PubMed: 11928888]
- Mash EJ, Hunsley J. Evidence-based assessment of child and adolescent disorders: Issues and challenges. *Journal of Clinical Child and Adolescent Psychology*. 2005; 34:362–379. [PubMed: 16026210]
- Moffitt TE. Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review*. 1993; 100:674–701. [PubMed: 8255953]
- Moffitt TE, Arseneault L, Belsky D, Dickson N, Hancox RJ, Harrington HL, Caspi A. A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*. 2011; 108:2693–2698.
- Muthén, B.; Muthén, L. Mplus (Version 6.0). Los Angeles, CA: Authors; 2010.
- O'Connell, ME.; Boat, T.; Warner, KE., editors. Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities. Washington, DC: National Academies Press; 2009.
- Petras H, Chilcoat HD, Leaf PJ, Ialongo NS, Kellam SG. Utility of TOCA-R scores during elementary school years in identifying later violence among adolescent males. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2004; 43:88–96. [PubMed: 14691364]
- Petras H, Ialongo NS, Lambert SF, Barrueco S, Schaeffer CM, Chilcoat HD, Kellam SG. The utility of elementary school TOCA-R scores in identifying later criminal court violence among adolescent females. *Journal of the American Academy of Child & Adolescent Psychiatry*. 2005; 44:790–797. [PubMed: 16034281]
- Shaffer, D.; Fisher, P. NIMH-Diagnostic Interview Schedule for Children. New York: New York State Psychiatric Institute; 1997.
- Snider VE, Busch T, Arrowood L. Teacher knowledge of stimulant medication and ADHD. *Remedial and Special Education*. 2003; 24:46–56.
- Snider VE, Frankenberger W, Aspengren M. The relationship between learning disabilities and attention deficit hyperactivity disorder: A national survey. *Developmental Disabilities Bulletin*. 2000; 28:18–38.
- Stormont M. Early child risk factors for externalizing and internalizing behaviors: A 5-year follow-forward assessment. *Journal of Early Intervention*. 2000; 23:180–190.
- Tabachnick, BG.; Fidell, LS. Using multivariate statistics. New York, NY: HarperCollins; 2001.
- Wakschlag LS, Keenan K. Clinical significance and correlates of disruptive behavior in environmentally at-risk preschoolers. *Journal of Clinical Child and Adolescent Psychology*. 2001; 30:262–275.
- Wehby JH, Dodge KA, Valente E Jr. the Conduct Problems Prevention Research Group. School behavior of first grade children identified as at-risk for development of conduct problems. *Behavioral Disorders*. 1993; 19:67–78.
- Werthamer-Larsson L, Kellam SG, Wheeler L. Effect of first-grade classroom environment on shy behavior, aggressive behavior, and concentration problems. *American Journal of Community Psychology*. 1991; 19:585–602. [PubMed: 1755437]
- Wu J, King KM, Witkiewitz K, Racz SJ, McMahon RJ. the Conduct Problems Prevention Research Group. Item analysis and differential item functioning of a brief conduct problem screen. *Psychological Assessment*. 2012; 24:444–454. [PubMed: 22040514]

Table 1
Descriptive Statistics of Outcome Variables Measured at the End of Elementary, Middle, and High School

Variable	6th grade <i>M (SD)/%</i>	8th grade <i>M (SD)/%</i>	9th grade <i>M (SD)/%</i>	11th grade <i>M (SD)/%</i>	12th grade <i>M (SD)/%</i>
Teacher-rated behavior problems	60.38 (10.69) _a	59.22 (11.14) _b			
Teacher-rated social competence	2.80 (1.05) _a	2.80 (1.05) _a			
Parent-rated school adjustment	3.39 (.74) _a	3.39 (.73) _a		3.41 (.60) _a	
Child-rated school adjustment	3.54 (.70) _a	3.62 (.67) _b		3.74 (.62) _c	
ADHD diagnosis	8.8% (<i>n</i> = 66) _a		6.5% (<i>n</i> = 48) _a		3.6% (<i>n</i> = 27) _a
CD diagnosis	7.8% (<i>n</i> = 59) _a		7.7% (<i>n</i> = 58) _a		4.9% (<i>n</i> = 37) _a
ODD diagnosis	12.5% (<i>n</i> = 94) _a		10.0% (<i>n</i> = 75) _a		6.9% (<i>n</i> = 52) _a
ADHD symptom count	5.53 (4.13) _a		3.97 (4.96) _b		2.53 (3.90) _c
CD symptom count	2.83 (2.76) _a		3.75 (3.30) _b		4.14 (3.85) _c
ODD symptom count	4.13 (2.58) _a		4.15 (2.63) _a		3.36 (2.46) _b
Days smoked cigarettes past month ^a		.50 (.05)			1.64 (.10)
Days drank alcohol past year ^a		.52 (.06) _a			1.63 (.11) _b
Number of drinks each time past year ^a		.40 (.04) _a			.99 (.05) _b
Days had 5+ drinks in a row past year ^a		.25 (.04) _a			.84 (.08) _b
Alcohol intensity (Days × Drinks) ^a		1.08 (.15) _a			3.71 (.31) _b
Used marijuana past month ^a		.18 (.03) _a			.69 (.07) _b

Note. Means in the same row with different subscripts differ at $p < .05$ in t -test comparisons (z -test comparisons for percentages). ADHD = attention-deficit/hyperactivity disorder; CD = conduct disorder; ODD = oppositional defiant disorder.

^aVariables are square-root transformed for all analyses.

Table 2
Intercorrelations Between TOCA-R and School Outcomes Measured at the End of Elementary, Middle, and High School

Variable	1	2	3	4	5	6	7	8	9	10	11
1. TOCA-R	—										
2. Bx problems (TR): Grade 6	.41***	—									
3. Bx problems (TR): Grade 8	.35***	.57***	—								
4. Social comp (TR): Grade 6	-.38***	-.78***	-.52***	—							
5. Social comp (TR): Grade 8	-.33***	-.51***	-.76***	.56***	—						
6. School adjust (PR): Grade 6	-.29***	-.46***	-.38***	.57***	.42***	—					
7. School adjust (PR): Grade 8	-.27***	-.40***	-.54***	.44***	.58***	.50***	—				
8. School adjust (PR): Grade 11	-.41***	-.34***	-.36***	.42***	.36***	.36***	.48***	—			
9. School adjust (CR): Grade 6	-.31***	-.44***	-.37***	.49***	.41***	.52***	.40***	.28***	—		
10. School adjust (CR): Grade 8	-.21***	-.38***	-.46***	.39***	.49***	.34***	.52***	.29***	.51***	—	
11. School adjust (CR): Grade 11	-.16***	-.28***	-.25***	.35***	.37***	.25***	.35***	.32***	.42***	.48***	—

Note. TOCA-R = Teacher Observation of Classroom Adaptation-Revised; Bx = Behavior; TR = teacher rated; comp = competence; adjust = adjustment; PR = parent rated; CR = child rated.

 $p < .001$.

Table 3
Intercorrelations Between TOCA-R and Externalizing Disorders Measured at the End of Elementary, Middle, and High School

Variable	1	2	3	4	5	6	7	8	9	10
1. TOCA-R	—	.19***	.23***	.13**	.08*	.13**	.06	.15***	.13**	.09*
2. ADHD: Grade 6	.28	—	.33***	.34***	.18***	.10*	.04	.35***	.21***	.25***
3. ADHD: Grade 9	.26	.51***	—	.40***	.13**	.24***	.16***	.21***	.35***	.25***
4. ADHD: Grade 12	.23	.46***	.54***	—	.14**	.23***	.13**	.21***	.18***	.35***
5. CD: Grade 6	.30	.54***	.39***	.30***	—	.37***	.30***	.26***	.19***	.10*
6. CD: Grade 9	.29	.44***	.49***	.33***	.64***	—	.23***	.19***	.37***	.17***
7. CD: Grade 12	.24	.40***	.44***	.47***	.61***	.65***	—	.20***	.16***	.21***
8. ODD: Grade 6	.15	.63***	.35***	.29***	.56***	.38***	.39***	—	.33***	.23***
9. ODD: Grade 9	.10*	.34***	.52***	.32***	.37***	.61***	.44***	.50***	—	.25***
10. ODD: Grade 12	.09*	.30***	.38***	.53***	.25***	.34***	.54***	.41***	.51***	—

Note. Numbers above the diagonal represent nonparametric correlations (Spearman's) of externalizing diagnoses (yes/no); numbers below the diagonal represent parametric correlations of externalizing disorder symptom counts. TOCA-R = Teacher Observation of Classroom Adaptation-Revised; ADHD = attention-deficit/hyperactivity disorder; CD = conduct disorder; ODD = oppositional defiant disorder.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 4

Linear Regression Analyses of TOCA-R Sum Scores Predicting School Outcomes

Variable	TOCA-R sum scores		
	ΔR^2	β	95% CI for β
Teacher-rated outcomes			
Behavior problems: Grade 6	.05 ***	.32 ***	[.19, .46]
Behavior problems: Grade 8	.04 ***	.29 ***	[.18, .39]
Social competence: Grade 6	.02 **	-.22 **	[-.35, -.09]
Social competence: Grade 8	.04 ***	-.28 ***	[-.39, -.17]
Parent-rated outcomes			
School adjustment: Grade 6	.01 *	-.13 *	[-.24, -.02]
School adjustment: Grade 8	.01 *	-.16 *	[-.29, -.02]
School adjustment: Grade 11	.01 *	-.15 *	[-.29, -.01]
Child-rated outcomes			
School adjustment: Grade 6	.03 ***	-.22 **	[-.34, -.09]
School adjustment: Grade 8	.01 *	-.12	[-.26, .01]
School adjustment: Grade 11	.00	-.06	[-.20, .08]

Note. Covariates entered at Step 1: gender, race/urban status, and group. TOCA-R = Teacher Observation of Classroom Adaptation-Revised; CI = confidence interval.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table 5

Linear Regression Analyses of TOCA-R Sum Scores Predicting Externalizing Disorder Symptom Counts

Variable	TOCA-R sum scores		
	ΔR^2	β	95% CI for β
ADHD: Grade 6	.01 *	.14 *	[.03, .25]
ADHD: Grade 9	.01 *	.11	[−.02, .23]
ADHD: Grade 12	.00	.08	[−.08, .23]
CD: Grade 6	.01	.11	[−.03, .24]
CD: Grade 9	.01 *	.14 *	[.01, .26]
CD: Grade 12	.01	.13	[−.02, .28]
ODD: Grade 6	.00	.03	[−.09, .16]
ODD: Grade 9	.00	.07	[−.06, .19]
ODD: Grade 12	.01	.11	[−.03, .25]
ADHD: all grades	.01 *	.12 *	[.00, .23]
CD: all grades	.01	.11	[−.02, .24]
ODD: all grades	.00	.07	[−.05, .19]
Total symptom count: all grades	.01	.12 *	[.00, .23]

Note. Covariates entered at Step 1: gender, race/urban status, and group. TOCA-R = Teacher Observation of Classroom Adaptation-Revised; ADHD = attention-deficit/hyperactivity disorder; CD = conduct disorder; ODD = oppositional defiant disorder.

* $p < .05$.