Journal of School Psychology 64 (2017) 1–16

Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

The utility of single-item readiness screeners in middle school☆

Crystal G. Lewis*, Keith C. Herman, Francis L. Huang, Melissa Stormont, Caroline Grossman, Colleen Eddy, Wendy M. Reinke

University of Missouri, 201 London Hall, Columbia, MO 65211, United States

ARTICLE INFO

Action Editor: Craig Albers

Keywords:
Readiness
Middle school
Conditional probability indices
Screener
Odds ratios

ABSTRACT

This study examined the benefit of utilizing one-item academic and one-item behavior readiness teacher-rated screeners at the beginning of the school year to predict end-of-school year outcomes for middle school students. The Middle School Academic and Behavior Readiness (M-ABR) screeners were developed to provide an efficient and effective way to assess readiness in students. Participants included 889 students in 62 middle school classrooms in an urban Missouri school district. Concurrent validity with the M-ABR items and other indicators of readiness in the fall were evaluated using Pearson product-moment correlation coefficients, with the academic readiness item having medium to strong correlations with other baseline academic indicators ($r = \pm 0.56$ to 0.91) and the behavior readiness item having low to strong correlations with baseline behavior items ($r = \pm 0.20$ to 0.79). Next, the predictive validity of the M-ABR items was analyzed with hierarchical linear regressions using end-of-year outcomes as the dependent variable. The academic and behavior readiness items demonstrated adequate validity for all outcomes with moderate effects ($\beta = \pm 0.31$ to 0.73 for academic outcomes and $\beta = \pm 0.24$ to 0.59 for behavioral outcomes) after controlling for baseline demographics. Even after controlling for baseline scores, the M-ABR items predicted unique variance in almost all outcome variables. Four conditional probability indices were calculated to obtain an optimal cut score, to determine ready vs. not ready, for both single-item M-ABR scales. The cut point of "fair" yielded the most acceptable values for the indices. The odd ratios (OR) of experiencing negative outcomes given a "fair" or lower readiness rating (2 or below on the M-ABR screeners) at the beginning of the year were significant and strong for all outcomes (OR = 2.29 to OR = 14.46), except for internalizing problems. These findings suggest promise for using single readiness items to screen for varying negative end-of-year student outcomes.

1. Introduction

Successfully navigating school contexts requires a level of readiness to meet the academic and behavior expectations that exist within the school (Snow, 2006). Middle school, typically defined as 6th through 8th grade (U.S. Department of Education, 2001), represents a challenging transition for many students as they attempt to meet the varied expectations of multiple teachers as they rotate classrooms each day. Not surprisingly, students frequently experience waning academic engagement and increased behavioral troubles during middle school (Marks, 2000; Pianta & Allen, 2008; U.S. Department of Education, 2003). Success in middle school requires a variety of academic and social skills, and middle school teachers are in a prime position to identify students at risk of failing

☆ The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A130143 to the University of Missouri (PI: Keith Herman). The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

* Corresponding author at: Department of Educational, School & Counseling Psychology, University of Missouri, Columbia, MO 65211, United States.
E-mail address: hamptoncg@missouri.edu (C.G. Lewis).

http://dx.doi.org/10.1016/j.jsp.2017.04.003
Received 25 January 2016; Received in revised form 25 January 2017; Accepted 9 April 2017
Available online 28 April 2017

during this critical developmental period. Given the internal, social, and structural transformations that occur during the middle school years, readiness for the academic and social expectations of middle school may be an especially salient, and understudied, construct. Consequently, the purpose of this study was to explore the utility of single-item screeners with domains of academics and behavior to identify students with low academic and behavioral middle school readiness.

Historically, school readiness has been defined as "the state of child competencies at the time of school entry that are important for later success" (Snow, 2006, p. 9). Given the emphasis on school entry, most research on readiness has focused on the transition into kindergarten. This clearly is a demanding transition as children enter school with diverse experiences and a range of preparation and must adapt to the academic, behavior, and social expectations of formal schooling for the first time (Rimm-Kaufman, Pianta, & Cox, 2000). Despite the traditional focus on school entry at kindergarten, the construct of readiness is also applicable to other developmental periods, particularly school transitions such as entry into middle school, and even the transition from one grade to the next, as students are expected to transition back from a summer break and demonstrate more advanced academic knowledge, behavior, and social skills (Lohaus, Elben, Ball, & Klein-Hessling, 2004). In particular, the middle school years represent an important time of transition where new social and academic expectations emerge (Seidman, Allen, Aber, Mitchell, & Feinman, 1994).

1.1. Transition to middle school

The middle school years are a time of transition between elementary school and high school. It is a period marked by significant changes, including larger schools and rotating classes involving shifting sets of peers, adults, and expectations (Ryan, Shim, & Makara, 2013; Wigfield, Eccles, Mac Iver, Reuman, & Midgley, 1991). In what is often a child's first major school change, students' projected academic growth can be disrupted during the transition to middle school, as large, negative transition effects are observed across students for both reading and math achievement (Akos, Rose, & Orthner, 2015; Bellmore, 2011; Randall & Engelhard, 2009). Shifts in social and self-regulation expectations accompany these structural changes (Kim, Schwartz, Cappella, & Seidman, 2014; Madjar & Cohen-Malayev, 2016; Rudolph, Lambert, Clark, & Kurlakowsky, 2001) and include stricter grading and behavior policies, increased use of peer-referenced evaluation and comparison changing social groups, and an increased proportion of peers experimenting with at-risk behaviors (Eccles & Midgley, 1990; Wigfield et al., 1991). Research suggests that these social disruptions are systematically related to declines in self-esteem and motivation, as various students lose confidence in their abilities (Simmons & Blyth, 1987) and experience decreased engagement in school (Galván, Spatzier, & Juvonen, 2011). All of this takes place concurrently with the significant physical and emotional changes associated with adolescent pubertal development (Steinberg, 2005).

Drawing on person-environment fit theory (Booth & Gerard, 2014; Eccles & Roeser, 2009), several researchers have suggested that these widespread environmental changes contribute to negative outcomes in middle school because these new environments fail to match the changing needs of students (Eccles & Midgley, 1990; Seidman et al., 1994). As students develop more complex peer relationships, they are placed in a large environment that involves frequent classroom and peer group change throughout the day, disrupting peer relationship development (Akos et al., 2015; Wigfield et al., 1991). As they face higher academic expectations, students receive less support from individual teachers (Martinez, Aricak, Graves, Peters-Myszak, & Nellis, 2011), with this mismatch leading to various negative outcomes, including lower grade-point averages, decreased motivation and self-esteem, and increased experimentation with at-risk behavior, such as substance use and criminal activity (National Middle School Association and National Association of Elementary School Principals, 2002; Ryan et al., 2013; Seidman et al., 1994; Wigfield et al., 1991). These negative outcomes are common across ethnic groups, suburban and urban schools, and socioeconomic status, and these outcomes may be worse for culturally and linguistically diverse populations (Gutman & Midgley, 2000; Seidman et al., 1994). Middle school students face an increased risk of depression (Goodwin, Mrug, Borch, & Cillessen, 2012), increased instances of disciplinary action (Theriot & Dupper, 2009), and increased peer victimization (Williford, Brisson, Bender, Jenson, & Forrest-Bank, 2011). Middle school transition seems an apt place to measure readiness; if students are not sufficiently prepared for these major environmental and social changes, they will be challenged and face an increasing risk for failure. It is also important to determine if specific student-level demographic characteristics are associated with increased vulnerability for negative outcomes, especially during critical transitional periods.

1.2. Student characteristics

Sociodemographic characteristics including gender, socioeconomic status (SES), and race are important factors to consider in research on school readiness. Gender is associated with disparities in academic achievement and the development of social and emotional skills. Boys and girls often differ in academic outcomes in math and reading, and teachers may have differing perceptions of academic skills based on gender (Akos et al., 2015; Robinson & Lubienski, 2011). Boys in middle school are also more likely to have externalizing behaviors and receive special education services (Coutinho & Oswald, 2005; Farmer et al., 2015). SES is also known to impact academic achievement; however, the extent of the impact depends on moderating factors (Sirin, 2005; White, 1982). Stormont, Herman, Reinke, King, and Owens (2015) found that students who were eligible for free and reduced lunch were more likely to be rated as having poor academic and overall readiness. Socioeconomic status may also play a role in the development of regulation and prosocial behaviors (Bradley & Corwyn, 2002). Lastly, an achievement gap exists between the average academic achievement of white students and that of many students in racially, culturally and linguistically diverse groups, and this gap may widen from elementary school to middle school (Burchinal et al., 2011). Race, ethnicity, and culture may also impact teachers' perception of students and may result in varying levels of support and expectations for students (Burchinal, Roberts, Rowley, & Zeisel,

2008; Grissom & Redding, 2016). Black students frequently are at risk for experiencing racism at both individual and systemic levels, which may have a negative impact on physical and mental health (Akos et al., 2015; Brondolo, Ver Halen, Pencille, Beatty, & Contrada, 2009). There is evidence that the negative academic impact of the middle school transition particularly impacts low-income and non-white students, as their academic growth is disrupted at a magnitude larger than their white or high-income peers (Akos et al., 2015).

1.3. Existing readiness screeners

Middle school screening is designed to quickly and efficiently identify students who may be at risk for developing or currently experiencing more significant difficulties so that these students can be targeted for more intensive follow-up evaluation and, if necessary, intervention. Given the potential changes in risk experienced during the middle school transition, it is important to establish screening processes in middle school, and it is best to intervene early before more severe problems, such as mental illness, substance use, or school dropout, develop later in adolescence (Goodwin et al., 2012; McIntosh, Flannery, Sugai, Braun, & Cochrane, 2008).

When selecting a screener, predictive validity and practical usability are crucial. Predictive validity indicates a screener's ability to accurately identify risk and predict students' outcomes, often indicated by a screener's consistency with an outcome measure at a future time point. Practical usability includes things such as will a teacher be willing to complete several ratings over multiple days, or will they feel they only have time for a few ratings? Currently, there are several tools available to screen for risk in middle school; however, in our review of existing screeners, there were no quick single-item readiness screeners available for middle school students. The Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997), a systematic screener of students' sociobehavioral strengths and difficulties is free, psychometrically sound (Goodman, 2001; Lane, Parks, Kalberg, & Carter, 2007), and has been validated for use across grades K-12. Despite these advantages, at 25 items for each student, the SDQ may represent a burden on middle school teachers, requiring significant time and effort, and is considered too cumbersome by many secondary teachers (Lane, Wehby, Robertson, & Rogers, 2007). The Social Skills Improvement System-Rating Scales (SSIS-RS; Gresham & Elliott, 2008) is a measure used to screen for social skills, problem behaviors, and academic competences using three raters (teacher, parent, and student) and costs approximately $5 per student to administer and score. SSIS-RS provides information about the student from multiple informants in various areas of functioning, but inter-rater agreement between the informants is moderate to weak, with correlations ranging from $r = 0.21$ to $r = 0.30$ (Gresham, Elliott, Cook, Vance, & Kettler, 2010). Each parent, teacher, and student form takes between 15 and 25 min to complete (Gresham et al., 2010), which, like the SDQ, may require too much time for screening each student in the classroom. The universal screening component of the SSIS, referred to as Performance Screening Guides (PSGs), is a more efficient tool designed to identify youth who would benefit from further assessment. The PSGs assess four domains (math, reading, motivation, and prosocial) intended to summarize several weeks of teacher observations and interactions with all students in their classroom. Existing evidence suggests that teachers find the PSGs to be feasible and useful, and initial evidence suggests the predictive validity of these tools, especially when used alongside curriculum-based measurement tools (Kettler & Albers, 2013). However, many schools may find the cost of the tool to be prohibitive (Lane, Oakes, & Menzies, 2010).

The Student Risk Screening Scale (SRSS; Drummond, 1994) is another free and psychometrically sound measure that consists of seven Likert-type items and was developed for universal screening of students in grades K-6 at risk of developing antisocial behavior. Its utility has been further validated for use at the middle school level (Lane, Parks, et al., 2007; Lane, Wehby, et al., 2007), and its diagnostic accuracy matches or exceeds levels found with the SDQ (Lane et al., 2010). While the SRSS may be useful for screening for behavioral difficulties specifically, an additional screener would be necessary to assess students' academic risk, requiring additional time and research from school professionals. Direct Behavior Ratings (DBR; Chafouleas, Riley-Tillman, & Christ, 2009) represent another tool to assess efficiently academic engagement and disruptive behaviors in children and are validated for use at the middle school level (Chafouleas et al., 2013). DBR is usually conducted by the teacher and is directly connected to a specific period of observation (such as a 15 or 30-minute period) rather than a broad time-period the rater is asked to remember (such as “the past few months”). Thus, this measure has been conceptualized as a hybrid of direct observations and traditional rating scales (Chafouleas et al., 2009). DBR can be used to monitor one to four operationally defined behaviors, and thus requires less teacher effort during each isolated measurement. However, DBR ratings were developed primarily for use in behavioral progress monitoring, and must be completed five to ten times to obtain representative data on students' behavioral trends. Although DBR items have also been evaluated for their utility in screening for emotional and behavioral risk (Chafouleas et al., 2013; Kilgus, Chafouleas, Riley-Tillman, & Welsh, 2012), researchers have still recommended screening procedures include collection of five to ten data points (Kilgus, Riley-Tillman, Chafouleas, Christ, & Welsh, 2014). Additionally, because of the hybrid nature of the DBR tool, teachers are asked to observe groups of students over specified time periods before completing ratings (typically no more than ten students at a time), so although the ratings themselves do not represent a significant burden on teachers, the screening procedure for an entire classroom is conducted over multiple days, as multiple data points are needed for both screening and progress monitoring.

Overall, one clear disadvantage of many screening options (e.g., SDQ, SSIS-RS, SRSS) is the amount of time they require, given that teachers are often overburdened with other demands. Additionally, teachers may not view screening procedures as worthwhile if they do not often observe a clear link between assessment and corresponding intervention provision (Glover & Albers, 2007; Slonski-Fowler & Truscott, 2004). The Kindergarten Academic Behavior Readiness screeners (K-ABR; Stormont, Reinke, & Herman, 2011) may serve as a model for readiness screening in middle school. The K-ABR is an efficient way for kindergarten teachers to assess readiness in students at the beginning of the school year. Stormont et al. (2015) examined the use of single items asking teachers to identify a child's readiness on academic, behavior, and overall domains in comparison to other students in the school. Academic

readiness ratings were predictive of end-of-year academic outcomes such as reading and math scores, teacher-rated competence, and concentration problems, even when controlling for student sex and free and reduced lunch status. Behavior readiness ratings also predicted outcomes including disruptive behaviors, prosocial behaviors, and concentration problems. Specifically, students rated as having poor academic readiness were eight times more likely to have low reading performance at the end of the year, and students rated as having poor behavior readiness were significantly more likely to have lower rates of prosocial behaviors at the end of the year (Stormont et al., 2015). Stormont, Thompson, Herman, and Reinke (2016) further studied the K-ABR and examined the utility of the single-item measure of overall readiness in a population of 893 kindergarten students with 18 teachers. Concurrent validity was found with one overall K-ABR readiness item and other measures of academic (Kindergarten Peer-Assisted Literacy Strategies [K-PAL], $r = 0.57$) and behavior skills (Social-Emotional School Readiness [SER] from Teacher Observation of Child Adaptation, Revised [TOCA-R], $r = 0.69$).

The K-ABR items demand very little time and effort from teachers (one Likert-type response per student), and research has suggested moderate to strong predictive utility when predicting end-of-year outcomes (Stormont et al., 2015; Stormont et al., 2016). Though multi-item measures are often seen as the standard in psychological assessment, single-item measures in fields such as occupational health psychology have well-established construct validity and predictive abilities, rendering them quite useful and extremely efficient (Fisher, Matthews, & Gibbons, 2015). If these single-item measures can reduce the screening burden placed on teachers without sacrificing the diagnostic accuracy and predictive abilities of their multi-item counterparts, single-item screeners would be extremely useful. While more research into the K-ABR is needed, available studies suggest that single-items screeners may be a quick and practical way to identify easily at-risk students in order to intervene early and prevent poor outcomes. However, the screener has not been explored for other important school transitions (e.g., transition from elementary school to middle) or if it is useful for screening readiness at each grade level.

1.4. Purpose

Several recent studies examining the K-ABR screeners showed that effective screening can be accomplished quickly and easily in elementary schools (Stormont et al., 2015; Stormont et al., 2016). The purpose of the current study was to investigate the predictive utility of similar single-item readiness screeners with domains of academics and behavior for a middle school population using the two-item Middle School Academic and Behavior Readiness (M-ABR; Stormont, Reinke, & Herman, 2013) screener. Further, for exploratory purposes, this study examined the relative predictive utility of the readiness ratings in 6th grade versus other middle school grade levels given the prediction from social field theory that entry into middle school may be the most important transition in predicting long-term adaptation problems. Our research questions, which applied to all middle school grade levels (i.e., 6th alone and 7th and 8th grade combined) included: (a) Are the M-ABR items correlated with other measures of student readiness in the beginning of the school year, including previous spring Missouri Assessment Program (MAP; Missouri Department of Elementary and Secondary Education, 2015) scores, observed fall disruptions, or teacher ratings of academic engagement, disruption, concentrations, etc.? (b) Do the M-ABR items predict end-of-year academic and behavioral outcomes after controlling for demographics (e.g., gender, race/ethnicity, grade) and the baseline levels of the outcome? (c) Does the predictive validity of each M-ABR item vary depending on grade level (6th grade vs. 7th and 8th grade combined)? (d) What are optimal cut scores for the practical utility of the M-ABR items and do the odds of having poor academic and behavioral outcomes increase if a student is rated below that cut score?

2. Method

2.1. Participants

This study included 889 students in 62 middle school classrooms in an urban Missouri school district. Teachers in this study were 84% female, 63% white, and 32% black. The student sample was composed of 49% male and 65% eligible for free and reduced lunch (Table 1). The largest percentage of students were black (83%), with most other students identifying as white (13%). There were 337 students in 6th grade, 320 in 7th grade and 232 in 8th grade. The sample was roughly representative of the district, which in 2015 was composed of 73% black students and 22% white students (Missouri Department of Elementary and Secondary Education, 2016). This study utilized data collected as part of a randomized trial evaluating the effects of a classroom management program (CHAMPS) in middle school math and reading classrooms (further information regarding the intervention is provided in the procedures). Teachers were recruited in August of each school year and asked to participate in this study. The current sample represented the first

Table 1
Descriptive information on middle school sample.

		N	Eligible FRL %	Male %	White %	Black %	Other %
Grade	6	337	66%	51%	14%	81%	5%
	7	320	61%	45%	13%	84%	3%
	8	232	71%	52%	11%	86%	3%
	Overall	889	65%	49%	13%	83%	4%

Note. FRL = free and reduced lunch.

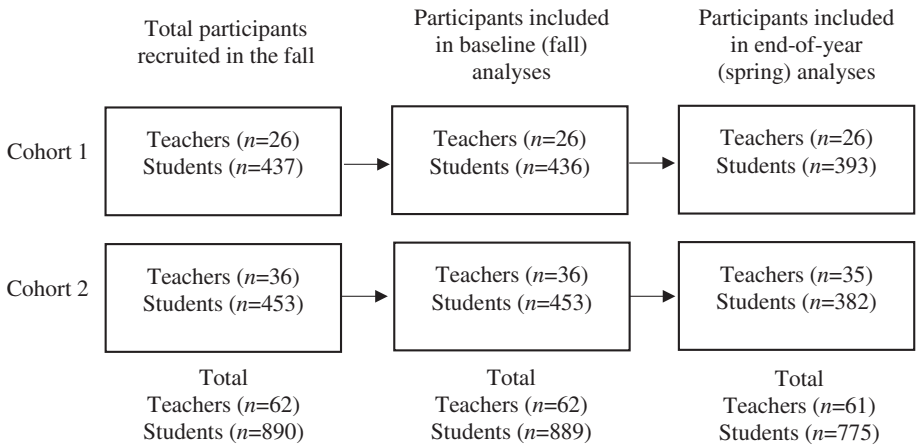

Fig. 1. Flow of participant data utilized in study. In baseline analyses one student was removed as a result of being an outlier based on regression diagnostics. In end-of-year analyses one teacher dropped out of the study and therefore we also did not collect follow up data on her students. The majority of the remaining missing student data was a result of student mobility.

two of four annual cohorts that will be recruited by the end of the project. Sixty-two middle school teachers and their classrooms were recruited in these first two cohorts (see Fig. 1 for the flow of participant data included in the study). All math and reading teachers in these buildings were invited to participate and approximately 73% of those teachers were recruited. The 27% who declined cited lack of time as the reason for not participating. Each teacher selected one of their classrooms as the target of the study and all students in that classroom were invited to participate. Teachers only provided data on that classroom and only on students who provided parent consent and child assent. Eighty percent of parents consented for their child to be in the study, and 100% of those students assented to participate.

2.2. Procedures

This study was completed with approval from the University of Missouri Institutional Review Board. Baseline data were collected in October, with outcome data collected in May of each school year. Half of the teachers were randomly assigned to receive the CHAMPS classroom management training after baseline data collection; the other half received business as usual continuing education. Contamination was minimized by asking intervention teachers to sign contracts stating they will not share CHAMPS information and emphasizing the importance of this. Teachers completed online survey ratings of student performance and online self-assessments, whereas students completed paper self-assessments. Direct observations (i.e., Brief Student-Teacher Classroom Interaction Observation code) were completed by trained independent observers who were not told which teachers were in the intervention in order to remain unbiased (details about observer trainings are described below). Additionally, a standardized achievement test (i.e., Stanford Achievement Test, 10th edition, Abbreviated Battery) was administered to students by the trained research team in May of each school year.

2.3. Measures

2.3.1. Middle School Academic and Behavior Readiness screeners (M-ABR; Stormont et al., 2013)

All teachers in the study completed one-item rating on student academic and one-item rating on behavior readiness in October via an online survey. The questions were scored on a 5-point Likert-type scale (*poor, fair, good, very good, and excellent*). The items included: (a) Compared to other students in this school, how was this child's academic readiness for middle school? (b) Compared to other students in this school, how was this child's readiness for the behavioral expectations of middle school? A study on a similar measure, the K-ABR, found the K-ABR items had excellent test-retest reliability and strong concurrent and predictive validity (Stormont et al., 2015; Stormont et al., 2016).

2.3.2. Demographics

Student free and reduced lunch status and race/ethnicity were both obtained from district-level data. Student and teacher gender were gathered from self-report surveys. Teacher race/ethnicity was gathered from teacher surveys. Grade level was coded as 1 for 6th grade and 0 for 7th or 8th grade.

2.3.3. Teacher Observation of Classroom Adaptation-Checklist (TOCA-C; Koth, Bradshaw, & Leaf, 2009)

Teachers completed pre- and post-ratings, via an online survey, regarding student behaviors in the TOCA-C, a checklist version of the TOCA-Revised interview. These ratings, including 55 items on a 6-point Likert-type scale (*never, rarely, sometimes, often, very often, almost always*), were completed in October and May of the school year. The TOCA-C subscale summed scores included in this study consisted of disruptive behavior, concentration problems, internalizing problems, prosocial behavior problems, and emotion

dysregulation, as well as one standalone academic item. In the current study, internal consistency for the subscales, computed using Cronbach's alpha, ranged from 0.81 to 0.98. In addition, prior studies support the TOCA's internal consistency, consistent factor structure over time, predictive and current validity, and sensitivity to change (Bradshaw, Waasdorp, & Leaf, 2012; Koth et al., 2009; Stormshak et al., 1999).

2.3.4. Stanford Achievement Test, 10th edition, Abbreviated Battery (SAT-10; Harcourt Assessment, Inc., 2004)

Students completed a multiple choice, grade level standardized assessment of math and reading achievement administered at the end of the school year. Scaled scores of the problem solving and reading comprehension subtests are used in analysis. The SAT-10 has evidence of test-retest reliability of 0.87 (Harcourt Assessment, Inc., 2004) and based on the current analytic sample, has been shown to have modest correlations ($r = 0.57$ to $r = 0.63$) with other measures of achievement such as the Missouri Assessment Program (MAP; Missouri Department of Elementary and Secondary Education, 2015) standardized test. Since the SAT-10 was only administered in the spring, the MAP test was used as the baseline covariate for the end-of-year achievement scores.

2.3.5. Direct Behavior Rating (DBR; Chafouleas et al., 2009)

Teachers completed a single-item rating of academic engagement on a scale from 0 to 10 based on an observation period, which was the most recent class period that the teacher observed the student. Teachers were instructed to *Select the button that best reflects the percentage of total time the student exhibited each target behavior: Actively or passively participating in classroom activity*. This measure was shown to have concurrent validity with other measures, including the Behavioral and Emotional Screening System (BESS) and the SRSS (Chafouleas et al., 2013). Using a middle school sample, correlations between the mean value of the DBR academic engagement item and the BESS t-score was $r = -0.55$ and was $r = -0.49$ with the SRSS sum score. Furthermore, no matter the time point (fall, winter, or spring), area under the curve (AUC) values for the DBR academic engagement item ranged from 0.86 to 0.87, suggesting that the DBR items classified students at-risk on the BESS better than chance (Johnson et al., 2016).

2.3.6. Student Teacher Classroom Interaction Observation code (ST-CIO; Reinke & Newcomer, 2010)

Independent observers conducted real-time, direct observations of student disruptive behavior using the ST-CIO code uploaded onto hand held devices using the Multi-Option Observation System for Experimental Studies software program (MOOSES; Tapp, Wehby, & Ellis, 1995). One five-minute observation was completed on every student in October and then again in May of the school year. Student disruptions were coded as a frequency count of any behavior that disrupted instruction. Two weeks prior to each data collection period, observers were trained in the code and practiced in the field to obtain 85% reliability with a master coder. The MOOSES program calculates reliability for frequency variables, such as the frequency of disruptions used in this study, by determining a match between observers within a 5 second window. Variables that are matched are counted as an agreement and variables that are not matched are counted as disagreements. An agreement ratio is then reported for each variable (agreements divided by the sum of agreements plus disagreements × 100%); 80% reliability is considered acceptable (Tapp, 2004). During data collection inter-observer agreement was completed on an average of 35% of observations across all time periods. Observations selected for reliability checks were distributed randomly. The overall mean percentage agreement was 93% for October and 96% for May. Research supports the concurrent and predictive validity of the ST-CIO. A recent study using the ST-CIO found significant relations between observed disruptive behaviors and teacher ratings of disruptive behaviors concurrently and over time. The study also reported that observed positive-to-negative teacher interactions with each student in the fall predicted increases in observed disruptive behaviors in the spring even after controlling for baseline disruptive behaviors (Reinke, Herman, & Newcomer, 2016). In the current study sample, the count of observed disruptions in the fall was significantly correlated with other measures of disruptions in the fall including the teacher-rated DBR and TOCA disruption scales.

2.4. Intervention

Although the effectiveness of the intervention was not a primary focus of this study, an intervention dummy code was added, 0 for control group and 1 for intervention, to control possible intervention effects on end-of-year outcomes. The intervention included a multi-day training in the CHAMPS program as well as on-going consultation and coaching in behavior management techniques. Teacher M-ABR ratings were independent of treatment/control status as intervention was randomly assigned at the teacher level after baseline surveys were completed by teachers. To rule out the possibility that intervention status differentially affected the intercorrelations among study variables, we first conducted analyses separately for each intervention condition. Visually, all correlations between the two conditions were very similar. Next, we converted the Pearson correlation coefficients to Fisher's z and subtracted the control group z-scores from the treatment group z-scores to compare differences in z-scores. The magnitude of differences ranged from $z = -0.19$ to 0.11. Since differences were negligible, we used the entire sample for subsequent analyses to maximize power. Additionally, we controlled for intervention condition in all regression analyses to eliminate any potential contribution of interventions status on the unique relations among study variables.

2.5. Data analyses

There were no missing data for any baseline analyses.[1] The rate of missingness for end-of-year outcomes was 13% and approximately 85% of missing data was a result of student mobility during follow-up.[2] This missingness level is similar to other studies with middle school samples (Chafouleas et al., 2013; Ryan et al., 2013). Missing data were handled using listwise deletion in

SAS. We also tested models for violations of assumptions. Residual vs. predicted plots indicated that the models were linear and there was no heteroskedasticity. QQ plots showed that errors were normally distributed and skewness and kurtosis were < 2 for all outcomes except observed disruptions. The Durbin-Watson statistic fell between 1.5 and 2.5 for all models supporting the assumption of independence of errors and all Variance Inflation Factors were < 10 showing absence of multicollinearity. Lastly, we removed outliers and influential observations using the following criteria: Cook's D > 4 / (n-k-1), Leverage > 2(k + 1) / n, Studentized Residuals were outside + 3/− 3, and the outlier impacted the estimate of the readiness variable (Fox, 1991; Kleinbaum, Kupper, Nizam, & Muller, 2008).

After finalizing our sample, descriptive statistics were run to obtain means of variables and to identify percent of students who were categorized as not ready for middle school based on the M-ABR cut point of "fair" (a score of 2 or below on the scale). Also, chi-square tests were calculated to see if there were any differences in middle school readiness associated with multiple demographic variables. Next, concurrent validity with the M-ABR items and other indicators of readiness in the fall were evaluated using Pearson product-moment correlation coefficients.

Next, predictive validity was measured between the M-ABR items and end-of-year outcomes using Pearson product-moment correlations. Then the M-ABR items were analyzed with hierarchical regressions using end-of-year outcomes as the dependent variable: (a) baseline covariates [gender, free and reduced lunch status, race/ethnicity, grade, intervention status, study year, school, teacher gender, and teacher race/ethnicity] were entered; (b) baseline M-ABR items were added as continuous predictors to determine the unique contributions of the readiness item; and (c) baseline scores on the dependent variable were added as a covariate to determine if M-ABR items predicted the outcome beyond its baseline value, essentially assessing whether the M-ABR could predict the emergence of new problems that were not present at baseline. Data were modeled using proc surveyreg in SAS with Taylor Series Linearization variance estimation to account for the multilevel nature of the data together with school fixed effects (Huang, 2014). All continuous variables were standardized and effect size of the standardized β were measured as small, moderate, or strong based on β = 0.20, 0.50, and 0.80, respectively (Ferguson, 2009). For the observed disruptions outcome, negative binomial regression was used to account for the highly skewed nature of the data and to account for overdispersion (i.e., where $\sigma^2 > \mu$), which also may bias standard errors (Huang & Cornell, 2012). No variables were standardized for this model to allow for easier interpretation. Last, we adjusted our alpha level using a Bonferroni correction (α/13 or $p < 0.004$) to account for inflated likelihood of Type I errors due to multiple planned analyses.

Next, we explored the best cut-point to dichotomize the readiness screener scores into ready vs. not ready groups to evaluate its utility as a tool for making real-world, categorical decisions about which students need additional services and who to divert time and resources to. Therefore, four primary conditional probability indices were derived from contingency tables to explore an optimal cut score for both M-ABR items. We calculated sensitivity (i.e., the probability that the screener correctly identified students with adverse outcomes), specificity (i.e., the probability that the screener correctly identified students who did not have adverse outcomes), positive predictive values (PPV; i.e., the probability that a student rated as not ready for middle school have adverse outcomes), and negative predictive values (NPV; i.e., the probability that a student rated as ready for middle school did not have adverse outcomes; Chafouleas et al., 2013; Glover & Albers, 2007; Stormont et al., 2016). When selecting an optimal cut score, we focused on maximizing levels of sensitivity as sensitivity tends to be most important when the purpose of the cut score is to screen for individuals at risk for a specific condition (Jenkins, Hudson, & Johnson, 2007; Myers, Gross, & McReynolds, 2014). However, we also looked for a cut-point that balanced all four conditional probability indices. Sensitivity and specificity values falling below 0.60 were considered low, values between 0.60 and 0.80 were considered moderate, and those above 0.80 were considered high (Kettler & Feeney-Kettler, 2011). Furthermore, we looked for a cut score with the lowest value of *d* (Yovanoff & Squires, 2006), where,

$$d = \sqrt{(1 - sensitivity)^2 + (1 - specificity)^2}$$

To determine overall diagnostic accuracy of the single readiness item for the academic domain and the single readiness item for the behavior domain, we calculated the area under the curve (AUC) statistic from a Receiver Operating Characteristic (ROC) Curve using the interval level M-ABR items to determine their ability to differentiate between students identified as being ready versus not being ready. Values for the AUC range from 0.50 (indicating classification no better than chance) to 1.00 (perfect classification capability), with values falling between 0.50 and 0.70 providing low accuracy, values between 0.70 and 0.90 providing moderate accuracy, and anything above 0.90 providing high accuracy (Chafouleas et al., 2013; Fischer, Bachmann, & Jaeschke, 2003; Lasko, Bhagwat, Zou, & Ohno-Machado, 2005; Streiner & Cairney, 2007).

Last, we calculated the odds ratios of having negative end-of-year outcomes (being at the 15th percentile or below) based on being rated as unready at baseline using logistic regression. Based on the prevalence within each M-ABR rating, described below, we defined unready as a rating of "fair" or "poor" (a readiness score of 2 or below) on the M-ABR. For end-of-year outcomes, the 15th percentile is a common cut score for identifying at-risk students in school-based tiered models of support (Stormont, Reinke, Herman, & Lembke, 2012) and was the cut score used in previous K-ABR studies. For ease of interpretation we also converted odds ratios into Cohen's d effect sizes, where d = ln (OR)($\sqrt{3}/\pi$) (Chinn, 2000) with small, medium, and large effect sizes falling at d = 0.20, 0.50, and 0.80 (Cohen, 1992). Analyses with categorical outcomes were intended to supplement the analyses with continuous readiness ratings by examining the utility of the readiness items as they would be used in school settings where categorical decisions often are made (e.g., Does this student need support for an academic or behavior problem?). The odds ratios and

[1] Baseline analyses include descriptives, χ^2 analyses, correlations, and concurrent conditional probability analyses.

[2] End-of-year outcome analyses include hierarchical regressions, predictive conditional probability analyses, and logistic regressions.

Table 2
Crosstabs of demographics and readiness items.

Category	Statistic	Not ready academic readiness $n = 258$ (*n*) %	Not ready behavior readiness $n = 244$ (*n*) %
Male		154(35)	152(35)
Female		104(23)	92(20)
	χ^2	16.49**	23.63**
	phi	− 0.14	− 0.16
Not eligible FRL		64(21)	61(20)
Eligible FRL		194(33)	183(31)
	χ^2	15.21**	13.52**
	phi	0.13	0.12
Black		229(31)	221(30)
Asian/Pacific Islander		1(9)	1(9)
Hispanic		4(19)	3(14)
White		24(21)	19(17)
	χ^2	8.07*	12.62**
	phi	0.10	0.12
6th grade		84(25)	87(26)
7th/8th grade		174(32)	157(28)
	χ^2	4.42*	0.72
	phi	− 0.07	− 0.03

Note. Not ready in academic or behavior readiness is scored at "fair" or below. FRL = free and reduced lunch.
* $p < 0.05$.
** $p < 0.01$.

classification results provide information on how useful the readiness items would be in those decisions.

3. Results

3.1. Descriptive and demographic analyses

In our combined sample of 6th through 8th grade students, 8% were rated "poor" on academic readiness at baseline and 21% were rated at "fair" on academic readiness at baseline. Thus 29% of students were classified as being not ready ("fair" or below) on academic readiness at baseline. Eight percent of students were rated "poor" on behavioral readiness at baseline and 19% were rated "fair"; therefore 27% of students were rated not ready ("fair" or below) on behavioral readiness at baseline. Approximately 38% of students were rated not ready on either academic readiness or behavior readiness at baseline.

Table 2 displays the results of our chi-square tests of student demographic differences in middle school readiness ratings. As expected there were significant differences in how students were rated at baseline based on demographics. Males were significantly more likely to be rated not ready ("fair" or "poor" on the scale) compared to females on academic readiness, $\chi^2(1) = 16.49$; $phi = -0.14$; $p < 0.01$, and behavior readiness, $\chi^2(1) = 23.63$; $phi = -0.16$; $p < 0.01$. Similarly, students who qualify for free and reduced lunch were more likely to be rated not ready on academic readiness, $\chi^2(1) = 15.21$; $phi = 0.13$; $p < 0.01$, and behavior readiness, $\chi^2(1) = 13.52$; $phi = 0.12$; $p < 0.01$. Race/ethnicity yielded significant differences for both academic readiness, $\chi^2(3) = 8.07$; $phi = 0.10$; $p < 0.05$ and behavior readiness as well, $\chi^2(3) = 12.62$; $phi = 0.12$; $p < 0.01$, with the proportion of black students being rated as not ready being significantly higher than their non-black peers. Last a chi-square test between 6th grade and the combined 7th and 8th grade group showed that teachers were more likely to rate 7th and 8th graders as not ready in terms of academic readiness, $\chi^2(3) = 4.42$; $phi = -0.07$; $p < 0.05$, but there were no differences in terms of behavioral readiness.

3.2. Concurrent validity

To assess concurrent validity, correlations were calculated between the M-ABR items and other baseline indicators of readiness. Table 3 provides the intercorrelations between baseline (October) M-ABR items and other baseline readiness indicators. Evidence of concurrent validity was found with the academic readiness item having medium to strong correlations with other baseline academic indicators including the MAP test scores, the TOCA concentration problems scale score, the TOCA academic performance item and the DBR academic engagement item ($r = \pm 0.56$ to 0.91). The strongest correlations at baseline for the academic readiness item were with TOCA academic performance ($r = 0.91$) and concentration problems ($r = -0.71$) and with the DBR Academic item ($r = 0.64$). This suggests confirmatory evidence for the academic readiness item as these are the three items we would expect to have the highest correlations. It also suggests that teachers are drawing on their perceptions of student academic performance and concentration problems when making judgements about student readiness. Notably, the baseline correlation between the academic readiness item and the baseline standardized achievement test scores ($r = 0.56$ and $r = 0.57$) was nearly as strong as the correlation between the baseline and follow-up reading and math test scores ($r = 0.63$ and $r = 0.57$). The behavior readiness item also had medium to strong correlations ($r = \pm 0.63$ to 0.79) with other baseline behavior indicators including the TOCA emotion

Table 3
Correlations among study variables at baseline and follow-up.

		1	2	3	4	5	6	7	8	9	10	11	12
1	Academic readiness	–	–	0.73	− 0.40	− 0.38	− 0.62	− 0.29	0.44	0.57	0.45	0.45	− 0.13
2	Behavior readiness	0.65	–	0.54	− 0.59	− 0.62	− 0.63	− 0.24	0.55	0.56	0.27	0.35	− 0.19
3	TOCA academic	0.91	0.64	(0.73)	− 0.40	− 0.39	− 0.63	− 0.27	0.44	0.60	0.41	0.47	− 0.10
4	TOCA emotion	− 0.42	− 0.69	− 0.44	(0.75)	0.72	0.51	0.35	− 0.55	− 0.42	− 0.22	**− 0.10**	0.19
5	TOCA disruptive	− 0.37	− 0.68	− 0.39	0.84	(0.79)	0.49	0.26	− 0.54	− 0.40	− 0.19	− 0.20	0.20
6	TOCA concentration	− 0.71	− 0.79	− 0.75	0.63	0.58	(0.75)	0.28	− 0.57	− 0.68	− 0.28	− 0.34	0.16
7	TOCA internalizing	− 0.31	− 0.23	− 0.31	0.41	0.26	0.29	(0.61)	− 0.34	− 0.24	**− 0.07**	**− 0.08**	**0.01**
8	TOCA prosocial	0.47	0.63	0.49	− 0.66	− 0.62	− 0.67	− 0.43	(0.74)	0.48	0.20	**0.06**	− 0.13
9	DBR academic	0.64	0.68	0.69	− 0.54	− 0.51	− 0.84	− 0.28	0.58	(0.69)	0.23	0.37	− 0.15
10	Reading test	0.56	0.33	0.55	− 0.26	− 0.24	− 0.37	**− 0.10**	0.22	0.33	(0.63)	–	− 0.14
11	Math test	0.57	0.38	0.59	− 0.22	− 0.26	− 0.43	− 0.15	0.14	0.38	–	(0.57)	− 0.14
12	Observed disruptions	− 0.12	− 0.20	− 0.13	0.18	0.25	0.21	**0.04**	− 0.12	− 0.22	− 0.18	**− 0.07**	**(0.08)**

Note. Upper diagonal refers to correlations between variables at baseline (October) and follow-up (May) and lower diagonal refers to correlations between variables at baseline. TOCA = teacher observation of classroom adaptation-checklist, DBR = Direct Behavior Rating. For #10 and #11, MAP scores from the previous spring were used as baseline scores and SAT-10 scores from May are used in follow-up. TOCA emotion, disruptive, concentration and internalizing were all negatively scored scales so a decrease in that scale is a positive outcome. All correlations in bold were not significant at $p < 0.01$.

dysregulation, prosocial behavior, concentration problems, and disruptive behavior. The behavior item had small correlations with internalizing symptoms and observed disruptions. It also had moderate correlations with DBR academic engagement ($r = 0.68$) and with the TOCA academic performance ($r = 0.64$) suggesting that teachers were drawing on both perceptions of academic and behavior items when rating student behavior readiness.

3.3. Predictive validity

3.3.1. Intercorrelations

Correlations were calculated between M-ABR items and follow-up (May) outcome variables. The academic readiness item had medium to strong positive correlations with academic outcomes ($r = \pm 0.45$ to 0.73). Again, the baseline correlation between the academic readiness item and the follow-up standardized achievement test scores ($r = 0.45$ for both) was nearly as strong as the correlation between the baseline and follow-up reading and math test scores ($r = 0.63$ and $r = 0.57$). The behavior readiness item had moderate correlations with most behavioral outcomes ($r = \pm 0.55$ to 0.63) and low correlations with TOCA internalizing ($r = -0.24$) and observed disruptions ($r = -0.19$).

3.4. Hierarchical regressions

To further assess the utility of the M-ABR items, we conducted hierarchical regressions to predict continuous end-of-year outcomes based on baseline readiness items, while controlling for covariates including gender, free and reduced lunch, grade level, race/ethnicity, intervention status, study year, school, teacher gender, teacher race/ethnicity and baseline scores on the dependent variable. To reduce the number of analyses conducted we only ran models with academic readiness predicting academic outcomes and behavioral readiness predicting behavioral outcomes except where indicated by promising concurrent validity findings. We first ran models with only demographic variables included and then added the respective M-ABR item (i.e., either the academic readiness item or the behavior readiness item) to see if the readiness item predicted variance beyond the demographic variables. Table 4 displays the results of all the models. The M-ABR items predicted improvements in variance beyond the demographic variables, ranging from change in $R^2 = 0.05$ for TOCA internalizing problems to change in $R^2 = 0.46$ for TOCA academic performance. This increase in explained variance shows that M-ABR can predict future outcomes above and beyond readily available demographic information.

The academic readiness item significantly predicted all outcomes with small to moderate effects ranging from $\beta = -0.31$ for internalizing problems to $\beta = 0.73$ for academic performance. The behavior readiness item also significantly predicted all outcomes with small to moderate effects ranging from $\beta = -0.24$ for internalizing problems to $\beta = -0.59$ for disruptive behavior. Additionally, observed disruptions decreased by a factor of 0.66 (or by 34%) with a one-unit increase in behavior readiness.

We repeated analyses while controlling for baseline rating of the outcome variable (e.g., fall academic performance was used as a baseline covariate in models with spring academic performance as the outcome) to determine if the readiness items could significantly predict changes over time. Even after controlling for the baseline score of the outcome variable, adding in readiness predicted unique variance in almost all outcome variables. Academic readiness continued to significantly predict all outcomes and behavior readiness remained significant in all models except for concentration problems and prosocial behavior, which did not meet our Bonferoni adjusted alpha of 0.004.

Lastly, we tested the interaction between the readiness items and the grade level variable. The interaction variables were not significant in any of the models. This finding suggests that the predictive validity of the readiness items do not vary depending on grade level.

Table 4
Regression analyses predicting follow-up outcomes from baseline covariates and readiness items.

Follow-up outcome - Baseline predictor	Models with covariates	R^2 (ΔR^2)[c]	Semi-partial r	Standard β	Standard error
SAT10 reading -	1: Model A[a]	0.32 (0.16)	0.40	0.50*	0.06
Academic readiness	2: Model B[b, d]	0.49 (0.02)	0.14	0.18*	0.06
SAT10 math -	1: Model A[a]	0.45 (0.16)	0.40	0.47*	0.05
Academic readiness	2: Model B[b, d]	0.52 (0.07)	0.26	0.37*	0.07
TOCA academic performance -	1: Model A	0.56 (0.46)	0.68	0.73*	0.03
Academic readiness	2: Model B	0.58 (0.02)	0.14	0.40*	0.07
DBR academic engaged -	1: Model A	0.36 (0.26)	0.48	0.51*	0.04
Academic readiness	2: Model B	0.52 (0.02)	0.14	0.20*	0.03
TOCA concentration problems -	1: Model A	0.44 (0.31)	0.56	− 0.60*	0.03
Academic readiness	2: Model B	0.60 (0.01)	0.10	− 0.16*	0.04
Toca internalizing problems -	1: Model A	0.22 (0.08)	0.28	− 0.31*	0.04
Academic readiness	2: Model B	0.44 (0.02)	0.14	− 0.13*	0.04
TOCA internalizing problems -	1: Model A	0.19 (0.05)	0.22	− 0.24*	0.05
Behavior readiness	2: Model B	0.44 (0.02)	0.14	− 0.12*	0.03
TOCA disruptive behavior -	1: Model A	0.45 (0.30)	0.55	− 0.59*	0.04
Behavior readiness	2: Model B	0.65 (0.01)	0.10	− 0.15*	0.04
TOCA prosocial behavior -	1: Model A	0.38 (0.23)	0.48	0.52*	0.05
Behavior readiness	2: Model B	0.58 (0.01)	0.10	0.13	0.04
TOCA emotion dysregulation -	1: Model A	0.40 (0.29)	0.54	− 0.58*	0.04
Behavior readiness	2: Model B	0.59 (0.01)	0.10	− 0.12*	0.04
TOCA concentration problems -	1: Model A	0.43 (0.30)	0.55	− 0.58*	0.04
Behavior readiness	2: Model B	0.59 (0.00)	0.03	− 0.06	0.05
DBR academic engaged -	1: Model A	0.34 (0.24)	0.49	0.53*	0.05
Behavior readiness	2: Model B	0.51 (0.01)	0.10	0.14*	0.04
Observed disruptions -	1: Model A[e]	n/a	n/a	− 0.42 (0.66)*	0.10
Behavior readiness	2: Model B[e]	n/a	n/a	− 0.41 (0.66)*	0.10

Note. Baseline is October of a school year and follow-up is May of the same school year.
[a] Model A includes the following variables: gender, free and reduced lunch, grade level, race/ethnicity, study year, school, intervention, teacher race/ethnicity, teacher gender, M-ABR readiness.
[b] Model B includes all variables in Model A as well as baseline of the outcome variable.
[c] Change in R^2 from Model A or Model B without M-ABR readiness to the model including M-ABR readiness
[d] Baseline scores for SAT10s are MAP scores from the previous spring.
[e] Negative binomial regression model reporting β (exp(β)).
* $p < 0.004$ based on Bonferroni adjusted alphas.

3.5. Conditional probability indices

We calculated conditional probability indices for two M-ABR cut scores, a rating of "poor" (a 1 on the scale) and a rating of "fair" and below (a rating of 2 or below on the scale). Cut scores for ratings beyond "fair" were not calculated as it did not make theoretical sense to test if a rating of "good" predicted a student being at-risk for negative outcomes. First, conditional probabilities were calculated using baseline outcome variables to determine cut scores to assess concurrent validity, which answered the question if the dichotomized readiness items identified students who are currently in the bottom 15th percentile for other readiness indicators such as academic performance or emotion dysregulation (Table 5). Although specificity was above 0.90 for all outcomes with the "poor" cut off, the sensitivity was unacceptably low (i.e., 0.10–0.67), suggesting that this was not an optimal cut point. Using "fair" or below as a cut point on the readiness items yielded much more acceptable sensitivity and specificity values. Most sensitivity values increased to within the moderate range, from 0.70 to 0.80 (Kettler & Feeney-Kettler, 2011), with values only falling in the low range for TOCA internalizing problems and observed disruptions. Concurrently, specificity had values falling in the moderate to high range. In addition to optimizing our conditional probability indices, the cut score of "fair" also satisfied our minimum *d* value requirement for all outcomes. Next, to further investigate the optimal readiness cut score, we also examined the conditional probabilities using end-of-year outcomes (Table 6). Again, the "poor" cut off produced unacceptable sensitivity values, but the "fair" cut off generated values falling between 0.60 and 0.70 for most outcomes, while also producing acceptable scores for specificity and the lowest value *d*. Furthermore, M-ABR items had overall diagnostic accuracy with AUCs for the academic readiness item falling in the moderate range, between 0.73 and 0.86, for all end-of-year outcomes except internalizing problems and AUCs for the behavior readiness item also falling in the moderate range, between 0.78 and 0.83, for all end-of-year outcomes except internalizing problems and observed disruptions.

3.6. Odds ratios

Finally, odds ratios of negative end-of-year outcomes were calculated using the cut score of "fair" or below (Table 7). A rating of "fair" or lower on the M-ABR academic readiness item was associated with increased odds for being in the bottom 15th percentile on SAT-10 reading and math subtests (odds ratios [OR] = 4.59 and 4.88, respectively), teacher rating of academic performance

Table 5
Concurrent validity: cut scores and associated conditional probability indices.

	Cut off "poor"					Cut off "fair" or below				
Academic readiness[a]	SN	SP	PPV	NPV	*d* value	SN	SP	PPV	NPV	*d* value
TOCA academic	0.67	0.99	0.88	0.96	0.33	0.93	0.79	0.36	0.99	0.22
DBR academic	0.35	0.96	0.56	0.91	0.65	0.76	0.78	0.35	0.95	0.33
TOCA concentration	0.34	0.96	0.63	0.89	0.67	0.77	0.80	0.42	0.95	0.30
TOCA internalizing	0.16	0.93	0.35	0.84	0.84	0.52	0.76	0.32	0.88	0.54
Behavior readiness										
TOCA internalizing	0.10	0.93	0.23	0.83	0.90	0.36	0.75	0.24	0.84	0.68
TOCA disruptive	0.41	0.98	0.81	0.90	0.59	0.78	0.82	0.44	0.95	0.29
TOCA prosocial	0.27	0.96	0.56	0.87	0.73	0.61	0.79	0.36	0.91	0.44
TOCA emotion	0.41	0.98	0.76	0.91	0.59	0.76	0.81	0.41	0.95	0.30
TOCA concentration	0.36	0.97	0.73	0.89	0.64	0.79	0.82	0.45	0.96	0.27
DBR academic	0.35	0.96	0.60	0.91	0.65	0.77	0.80	0.38	0.96	0.30
Observed disruptions	0.20	0.94	0.33	0.89	0.80	0.47	0.75	0.22	0.91	0.58

Note. SN = sensitivity; SP = specificity; PPV = positive predictive power; NPV = negative predictive power. These analyses are exploratory.
[a] M-ABR predictors were gathered in October. All outcomes were gathered in October and represent students rated in the bottom 15th percentile (reverse coded for negative indicators).

(OR = 14.46), and teacher rating of academic engagement (OR = 7.05), as well as being in the bottom 15th percentile (reverse coded) of teacher rating of concentration problems (OR = 8.09). Having a "fair" or lower rating on the behavior readiness item increased the odds of being in the bottom 15th percentile (reverse coded) of teacher ratings of disruptive behavior (OR = 9.93), emotion dysregulation (OR = 9.05), and concentration problems (OR = 9.96), and the bottom 15th percentile of teacher ratings of academic engagement (OR = 6.53) and prosocial behavior (OR = 5.26). Furthermore, it was associated with increased likelihood of observed disruptions (OR = 2.29) at follow-up. Ten out of thirteen Cohen's d effect sizes were considered large, ranging from 0.84 to 1.47 (Cohen, 1992). The only outcome not effectively predicted by the screener was internalizing problems.

4. Discussion

To support children and youth in meeting developmental demands, it is important to investigate key junctures or life events that are associated with increased vulnerability for adverse outcomes. Within the K-12 public school context, transitions are important to target for screening efforts given the negative outcomes associated with unsuccessful navigation through the challenges presented or magnified with new settings, expectations, teachers, and peers (Seidman et al., 1994; Stoep et al., 2005). Middle school marks an important time for such investigation, and the purpose of this study was to investigate the utility of two quick single-item screeners to target students who are struggling with academic and/or behavior expectations.

The two middle school readiness items demonstrated strong concurrent validity as evidenced by significant correlations with

Table 6
Predictive utility: cut scores and associated conditional probability indices.

	Cut off "poor"					Cut off "fair" or below				
Academic readiness[a]	SN	SP	PPV	NPV	*d* value	SN	SP	PPV	NPV	*d* value
SAT10 reading	0.10	0.98	0.43	0.86	0.90	0.48	0.83	0.34	0.90	0.54
SAT10 math	0.27	0.92	0.38	0.87	0.73	0.67	0.70	0.30	0.92	0.44
TOCA academic	0.40	0.98	0.68	0.92	0.60	0.79	0.79	0.34	0.97	0.29
DBR academic	0.25	0.96	0.45	0.90	0.75	0.67	0.78	0.30	0.94	0.40
TOCA concentration	0.24	0.96	0.55	0.87	0.76	0.67	0.80	0.38	0.93	0.39
TOCA internalizing	0.07	0.93	0.15	0.85	0.93	0.34	0.74	0.19	0.86	0.71
Behavior readiness										
TOCA internalizing	0.08	0.93	0.17	0.85	0.93	0.29	0.75	0.17	0.85	0.75
TOCA disruptive	0.29	0.98	0.70	0.87	0.71	0.67	0.83	0.44	0.93	0.37
TOCA prosocial	0.26	0.96	0.52	0.89	0.74	0.58	0.80	0.31	0.92	0.47
TOCA emotion	0.27	0.97	0.61	0.88	0.73	0.66	0.82	0.41	0.93	0.38
TOCA concentration	0.26	0.97	0.59	0.88	0.74	0.68	0.82	0.42	0.93	0.36
DBR academic	0.27	0.96	0.48	0.90	0.73	0.63	0.80	0.30	0.94	0.43
Observed disruptions	0.17	0.94	0.26	0.91	0.83	0.41	0.76	0.17	0.92	0.63

Note. SN = sensitivity; SP = specificity; PPV = positive predictive power; NPV = negative predictive power. These analyses are exploratory.
[a] M-ABR predictors were gathered in October. All outcomes were gathered in May and represent students rated in the bottom 15th percentile (reverse coded for negative indicators).

Table 7
Odds ratios of negative outcomes (15th percentile) given a readiness rating of fair or below.

Follow-up binary outcome - baseline binary predictor	Odds ratio	95% Confidence interval	Cohen's *d*
SAT-10 reading - academic readiness	4.59*	[2.59–8.13]	0.84
SAT-10 math - academic readiness	4.88*	[2.64–9.04]	0.87
TOCA academic performance - academic readiness	14.46*	[8.44–24.78]	1.47
DBR academic engaged - academic readiness	7.05*	[4.45–11.19]	1.08
TOCA concentration problems- academic readiness	8.09*	[5.30–12.36]	1.15
TOCA internalizing problems- academic readiness	1.42	[0.94–2.16]	0.19
TOCA internalizing problems- behavior readiness	1.22	[0.79–1.88]	0.11
TOCA disruptive behavior- behavior readiness	9.93*	[6.99–17.41]	1.27
TOCA prosocial behavior- behavior readiness	5.26*	[3.43–8.08]	0.92
TOCA emotion dysregulation- behavior readiness	9.05*	[5.91–13.85]	1.21
TOCA concentration problems- behavior readiness	9.96*	[6.47–15.31]	1.27
DBR academic engaged- behavior readiness	6.53*	[4.15–10.28]	1.04
Observed disruptions- behavior readiness	2.29*	[1.42–3.67]	0.46

Note. Baseline is October of the school year and follow-up is May of the same school year.
* $p < 0.004$ based on Bonferroni adjusted alphas.

several other baseline indicators, as well as through their sensitivity and specificity in predicting students in the bottom 15th percentile on measures gathered at the same time point. The 15th percentile, a relatively low cut point compared to high-stakes testing, was selected in line with tiered models of support to identify the most at-risk students. Similar to the findings in the K-ABR study (Stormont et al., 2015), the utility of the readiness items was supported as they were able to predict variance in the outcome variables beyond demographic variables and baseline scores. These findings suggest the promise of using single readiness items to assess risk for multiple outcomes rather than the more tedious method of using multiple baseline items to assess each individual outcome of interest. Also, we found that the interaction between readiness and grade level was not significant, suggesting that readiness is a construct that applies equally well across grade levels and not just at sixth grade middle school entry.

We found evidence to suggest that a rating of "fair" or below was theoretically supported as a cut point based on the small percentage of students rated as "poor" (8% on academic and 8% on behavioral). This percentage is lower than we would expect within a tiered approach to service delivery which often suggests 15–20% of students could benefit from further academic or behavioral supports (Stormont et al., 2012). This finding contrasts with the K-ABR studies which used "poor" as the cutoff for determining low readiness (Stormont et al., 2015; Stormont et al., 2016); however, in the kindergarten readiness studies, more children fell in the poor category (15%). It is possible that middle school teachers view readiness on a continuum as opposed to kindergarten teachers who regularly use and hear the term readiness and may more easily be able to think in dichotomous terms of not ready versus ready. It is also possible that pressures for schools to make adequate progress may be an important consideration as middle school teachers complete ratings. Teachers may worry that "poor" readiness reflects poorly on their districts given students have been in school for six years and have had access to a curriculum to support their transitions into each grade. Kindergarten teachers may also be reflecting on a true reality of more children not being prepared for entering structured school settings provided they are coming to school from varied preschool experiences. Finally, readiness also reflects teacher's ability to adapt to the varied experiences students have and the skills they possess as they transition to any given grade. Teacher factors, including knowledge of development, curriculum experience, and teaching experience at the current grade level, are important in considering contributing factors to readiness ratings. More research is needed to examine these possibilities.

Utilizing the cut point of "fair" was further supported in conditional probability analyses and the corresponding *d* values. Conditional probability analyses suggested that the middle school single readiness items have a comparable level of specificity/sensitivity as longer, more time-consuming screening instruments. While the cut score of "poor" did not optimize the four indices, the "fair" cut score provided more acceptable results. In predicting end-of-year outcomes with a cut point of "fair", four out of six sensitivity values for academic readiness fell above 0.60 and specificity fell above 0.70 for all outcomes. Furthermore, for behavioral readiness, the same cut off produced sensitivity values above 0.60 for five out of seven end-of-year outcomes and specificity values above 0.70 for all outcomes. These results are similar to conditional probability analyses calculated for the SRSS at the optimal cut point predicting middle school student risk on the BESS with sensitivity of 0.79 and specificity of 0.84 (Chafouleas et al., 2013).

Importantly, this study also documented that students with specific demographic characteristics were at increased risk for scoring in the "fair" or below category on academic and behavior readiness. As in previous research (Stormont et al., 2015), boys, students with free and reduced lunch status, and black students were at greater risk for being rated in a low readiness category. It is important that school-based decision making teams for at-risk students increase their understanding of these readiness gaps and work to reduce readiness and achievement gaps for individual groups based on extensive research documenting continued gaps that exist despite recent efforts to reduce them (Duncan et al., 2007; Ma, Nelson, Shen, & Krenn, 2015). Using a data-informed method for targeting specific areas, such as the poverty gap, would be prudent. Further, this supports the validity of the middle school screener to differentiate between groups based on specific demographic characteristics. More efforts relating to screening and intervention studies to determine a process for targeting groups at greater risk for drop out and other negative outcomes are needed. This is very salient for middle school readiness as the window for intervention with vulnerable students likely narrows every year that they are in

school.

Last, the practical utility of the M-ABR screeners were also examined in mapping on to real world, categorical decisions by calculating odds ratios of negative outcomes for students given readiness cut scores. Here again the M-ABR fared well as a possible tool for making screening decisions for those who may benefit from further evaluation and intervention. The odds of experiencing negative end-of-year outcomes given a low readiness rating at the beginning of the year were very high; youth rated at "fair" or lower in the M-ABR academic domain in the fall were 14 times more likely to be rated by teachers as having poor academic skills at the end of the year. Odds of having negative behavior outcomes at the end of the year given a "fair" or lower behavior readiness score at the start of the year were particularly strong for disruptive behavior (OR = 9.93), poor emotion regulation (OR = 9.05), and concentration problems (OR = 9.96). In addition to these negative teacher-rated outcomes, students were also found to have increased odds of more observed disruptions (OR = 2.29) at follow-up as rated by independent observers.

One outcome that was not associated with the dichotomized readiness ratings was internalizing problems. Internalizing may be a difficult construct for teachers to rate, and prior research suggests that teachers are often unable to recognize internalizing symptoms in their students (Auger, 2004; Kamphaus & Frick, 2002; Loeber, Green, Lahey, & Stouthamer-Loeber, 1991). However, in contrast, other research has found teachers could identify depressive symptoms in their students by utilizing brief Likert ratings (Auger, 2004). In previous work with the K-ABR screeners (Stormont et al., 2015), internalizing symptoms were predicted by low readiness on either academic or behavior readiness. Elementary school teachers may be in a better position to rate internalizing behaviors than middle school teachers given that they typically spend an entire day with each child rather than a single class period as often occurs in middle school.

4.1. Implications for research and practice

Results from this study supported the use of the M-ABR as a screener for academic and behavioral risk for middle school students. These screeners could be utilized within the context of school-based problem solving teams to determine students in need of follow-up assessments and increased supports. Some middle schools incorporate specific times for Response to Intervention classes for all students. While some students use this time as a study hall, this time could also be used as an unobtrusive time for assessment and intervention.

Follow-up assessments to target specific needs could include using direct observation that targets specific behaviors and follow-up in-depth rating scales and/or academic achievement tests for each student. Importantly, it is critical that more schools adopt screening methods that detect significant academic and behavior problems before they occur. Current practice in many schools is to track needs using documented outcomes such as office discipline referrals (ODRs), absences, and grades. Although these data are appealing given they are readily available, they reflect problems that are already occurring at a significant level rather than screening for behaviors to intervene with before they become more severe and prevalent. In this regard, readiness ratings are appealing in that they rely on teacher perceptions of key skills that are needed for success in that grade level. For instance, Stormont et al. (2016) found that kindergarten teachers were more likely to make readiness judgments based on perceptions of general skills such as self-regulation, problem solving, and gets along well with others, rather than on extreme behaviors. Although this finding will need to be replicated with middle school samples, it suggests that readiness may be an early warning indicator of emerging behavior and academic problems. This idea is supported by the current study and Stormont et al. (2015) in which readiness items predicted end-of-year problems even after controlling for the beginning year rating of those problems. Once specific areas are determined teams can use existing resources for supporting youth with challenging behavior.

Although this study adds to the research that demonstrates the promise of single-item readiness ratings, further research is needed to advance our understanding of these measures and optimize their performance in school settings. Notable limitations of the present study that will require further research to address include questions about generalization (given a single sample in a single school district) and the need for additional measures of common school outcomes beyond teacher report. Additionally, although we took several steps to examine and minimize any influence of intervention on observed correlations, it would be prudent to replicate these findings outside the context of intervention trials.

Given the modest success of readiness measures in predicting internalizing problems, it may be useful to supplement the teacher-rated readiness item with a student self-rating. This may be particularly useful in middle school where students are more independent and spend less time with individual teachers. In addition, more work is needed to determine the optimal response range for the readiness items. In the Stormont et al. (2016) study that examined the K-ABR overall readiness item for kindergarteners, the response range was 1–10 and 15% of kindergarteners fell into the "poor" range of 1–4. It is also possible that a different range is needed for the different ages of students. Here we found evidence to suggest that middle school teachers apply different criterion when assessing the various given response options of readiness. Whereas kindergarten teachers readily rated students as having "poor" readiness and these ratings mapped on to other indicators and outcomes, the optimal cut score for middle school teachers was at "fair" or lower. In previous research (Stormont et al., 2016), kindergarten teachers' overall readiness ratings had large positive correlations with specific skills related to self-regulation, including follows directions, works well alone, works well with others, problem solving, and persistence. There may be a common underlying construct that readiness represents in kindergarten such as self-regulation and organization. Further studies can help elucidate these differences between kindergarten and middle school by asking teachers to describe their decision making and reaction to different response options. Moreover, it would be helpful to understand how teachers perceive readiness and what it represents in terms of student skills and competencies. In the present study and in prior research using the K-ABR, teachers rated student's readiness according to their own perceptions of what readiness meant. That is, perceptions of readiness were assessed rather than responding to a provided definition of readiness to use when rating students. The advantage of

this approach is that a well selected prompt word (e.g., readiness) may elicit an entire string of connected behaviors that map onto the overarching construct.

4.2. Conclusions

Identifying students at risk for academic or social deterioration before the problems take root is a critical challenge for all school personnel. Although multiple measures have been developed for this purpose, many are not considered feasible for screening all school children because of their expense in terms of time, money, or both. A single-item readiness approach may help fill the void of measures that can provide a snapshot of all children on a key developmental construct to quickly identify students who may benefit from further assessment and services. The present study suggests the construct of readiness may have relevance beyond school entry to include students throughout the educational career. Readiness, after all, applies to all grades and new life tasks.

References

Akos, P., Rose, R. A., & Orthner, D. (2015). Sociodemographic moderators of middle school transition effects on academic achievement. *Journal of Early Adolescence*, *35*, 170–198.

Auger, R. W. (2004). The accuracy of teacher reports in the identification of middle school students with depressive symptomatology. *Psychology in the Schools*, *41*, 379–389.

Bellmore, A. (2011). Peer rejection and unpopularity: Associations with GPAs across the transition to middle school. *Journal of Educational Psychology*, *103*, 282–295.

Booth, M. Z., & Gerard, J. M. (2014). Adolescents' stage-environment fit in middle and high school: The relationship between students' perceptions of their schools and themselves. *Youth and Society*, *46*, 735–755.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology*, *53*, 371–399.

Bradshaw, C. P., Waasdorp, T. E., & Leaf, P. J. (2012). Effects of school-wide positive behavioral interventions and supports on child behavior problems. *Pediatrics*, *130*, e1136–e1145.

Brondolo, E., Ver Halen, N. B., Pencille, M., Beatty, D., & Contrada, R. J. (2009). Coping with racism: A selective review of the literature and a theoretical and methodological critique. *Journal of Behavioral Medicine*, *39*, 64–88.

Burchinal, M. R., Roberts, J. E., Rowley, S. J., & Zeisel, S. A. (2008). Social risk and protective factors for African American children's academic achievement in the upper elementary school years. *Developmental Psychology*, *44*, 286–292.

Burchinal, M., Steinberg, L., Friedman, S. L., Pianta, R., McCartney, K., Crosnoe, R., & McLoyd, V. (2011). Examining the black-white achievement gap among low-income children using the NICHD study of early child care and youth development. *Child Development*, *82*, 1404–1420.

Chafouleas, S. M., Kilgus, S. P., Jaffery, R., Riley-Tillman, T. C., Welsh, M., & Christ, T. J. (2013). Direct behavior rating as a school-based behavior screener for elementary and middle grades. *Journal of School Psychology*, *51*, 367–385.

Chafouleas, S. M., Riley-Tillman, T. C., & Christ, T. J. (2009). Direct Behavior Rating (DBR): An emerging method for assessing social behavior within a tiered intervention system. *Assessment for Effective Intervention*, *34*, 195–200.

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, *19*, 3127–3131.

Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology*, *112*, 155–159.

Coutinho, M. J., & Oswald, D. P. (2005). State variation in gender disproportionality in special education: Findings and recommendations. *Remedial and Special Education*, *26*, 7–15.

Drummond, T. (1994). *The student risk screening scale (SRSS)*. Grants Pass, OR: Josephine County Mental Health Program.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., ... Japel, C. (2007). School readiness and later achievement. *Developmental Psychology*, *43*, 1428–1446.

Eccles, J. S., & Midgley, C. (1990). Changes in academic motivation and self-perception during early adolescence. In G. R. Adams, & T. P. Gullotte (Eds.), *Childhood to adolescence: A transitional period?* (pp. 134–155). Newbury Park, CA: Sage.

Eccles, J. S., & Roeser, R. W. (2009). Schools, academic motivation, and stage-environment fit. In R. M. Lerner, & L. Steinberg (Eds.), *Handbook of adolescent psychology* (pp. 404–434). (3rd ed.). Hoboken, NJ: Wiley.

Farmer, T. W., Irvin, M. J., Motoca, L. M., Leung, M., Hutchins, B. C., Brooks, D. S., & Hall, C. M. (2015). Externalizing and internalizing behavior problems, peer affiliations, and bullying involvement across the transition to middle school. *Journal of Emotional and Behavioral Disorders*, *23*, 4–16.

Ferguson, C. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*, 532–538.

Fischer, J. E., Bachmann, L. M., & Jaeschke, R. (2003). A reader's guide to the interpretation of diagnostic test properties: Clinical example of sepsis. *Intensive Care Medicine*, *29*, 1043–1051.

Fisher, G. G., Matthews, R. A., & Gibbons, A. M. (2015). Developing and investigating the use of single-item measures in organizational research. *Journal of Occupational Health Psychology*, *21*, 3–23.

Fox, J. (1991). *Regression diagnostics: An introduction. Vol. 79*. Thousand Oaks, CA: Sage.

Galván, A., Spatzier, A., & Juvonen, J. (2011). Perceived norms and social value to capture school culture in elementary and middle school. *Journal of Applied Developmental Psychology*, *32*, 346–352.

Glover, T. A., & Albers, C. A. (2007). Considerations for evaluating universal screening assessments. *Journal of School Psychology*, *45*, 117–135.

Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Child Psychology & Psychiatry & Allied Disciplines*, *38*, 581–586.

Goodman, R. (2001). Psychometric properties of the strengths and difficulties questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 1337–1345.

Goodwin, N. P., Mrug, S., Borch, C., & Cillessen, A. H. N. (2012). Peer selection and socialization in adolescent depression: The role of school transitions. *Journal of Youth and Adolescence*, *41*, 320–332.

Gresham, F. M., & Elliott, S. N. (2008). *Social skills Improvement system-rating scales*. Minneapolis, MN: Pearson Assessments.

Gresham, F. M., Elliott, S. N., Cook, C. R., Vance, M. J., & Kettler, R. (2010). Cross-informant agreement for ratings for social skill and problem behavior ratings: An investigation of the Social Skills Improvement System-Rating Scales. *Psychological Assessment*, *22*, 157–166.

Grissom, J. A., & Redding, C. (2016). Discretion and disproportionality: Explaining the underrepresentation of high-achieving students of color in gifted programs. *AERA Open*, *2*, 1–25.

Gutman, L. M., & Midgley, C. (2000). The role of protective factors in supporting the academic achievement of poor African American students during the middle school transition. *Journal of Youth and Adolescence*, *29*, 223–248.

Harcourt Assessment, Inc (2004). *Stanford achievement test series, tenth edition technical data report.* San Antonio, TX: Author.

Huang, F. L. (2014). Analyzing group level effects with clustered data using Taylor Series linearization. *Practical Assessment, Research & Evaluation*, *19*, 1–9.

Huang, F. L., & Cornell, D. G. (2012). Pick your poisson: A tutorial on analyzing counts of student victimization data. *Journal of School Violence*, *11*, 187–206.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review*, *36*, 582–600.

Johnson, A. H., Miller, F. G., Chafouleas, S. M., Welsh, M. E., Riley-Tillman, T. C., & Fabiano, G. (2016). Evaluating the technical adequacy of DBR-SIS in tri-annual behavioral screening: A multisite investigation. *Journal of School Psychology*, *54*, 39–57.

Kamphaus, R. W., & Frick, P. J. (2002). *Clinical assessment of child and adolescent personality and behavior*(2nd ed.). Boston, MA: Allyn & Bacon.

Kettler, & Albers (2013). Predictive validity of CBM and teacher ratings of academic achievement. *Journal of School Psychology*, *51*, 499–515.
Kettler, R. J., & Feeney-Kettler, K. A. (2011). Screening systems and decision-making at the preschool level: Application of a comprehensive validity framework. *Psychology in the Schools*, *48*, 430–441.
Kilgus, S. P., Chafouleas, S. M., Riley-Tillman, T. C., & Welsh, M. E. (2012). Direct behavior rating scales as screeners: A preliminary investigation of diagnostic accuracy in elementary school. *School Psychology Quarterly*, *27*, 41–50.
Kilgus, S. P., Riley-Tillman, T. C., Chafouleas, S. M., Christ, T. J., & Welsh, M. E. (2014). Direct behavior rating as a school-based behavior universal screener: Replication across sites. *Journal of School Psychology*, *52*, 63–82.
Kim, H. Y., Schwartz, K., Cappella, E., & Seidman, E. (2014). Navigating middle grades: Role of social contexts in middle grade school climate. *American Journal of Community Psychology*, *54*, 28–45.
Kleinbaum, D. G., Kupper, L. L., Nizam, A., & Muller, K. E. (2008). *Applied regression analysis and other multivariable methods: An introduction*(4th ed.). Belmont, CA: Thompson Brooks/Cole.
Koth, C. W., Bradshaw, C. P., & Leaf, P. J. (2009). Teacher observation of classroom adaptation—Checklist: Development and factor structure. *Measurement and Evaluation in Counseling and Development*, *42*, 15–30.
Lane, K. L., Oakes, W., & Menzies, H. (2010). Systematic screenings to prevent the development of learning and behavior problems: Considerations for practitioners, researchers, and policy makers. *Journal of Disability Policy Studies*, *21*, 160–172.
Lane, K. L., Parks, R. J., Kalberg, J. R., & Carter, E. W. (2007a). Systematic screening at the middle school level: Score reliability and validity of the Student Risk Screening Scale. *Journal of Emotional and Behavioral Disabilities*, *15*, 209–222.
Lane, K. L., Wehby, J., Robertson, E. J., & Rogers, L. (2007b). 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.
Lasko, T. A., Bhagwat, J. G., Zou, K. H., & Ohno-Machado, L. (2005). The use of receiver operating characteristic curves in biomedical informatics. *Journal of Biomedical Informatics*, *38*, 404–415.
Loeber, R., Green, S. M., Lahey, B. B., & Stouthamer-Loeber, M. (1991). Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. *Journal of Abnormal Child Psychology*, *19*, 75–95.
Lohaus, A., Elben, C. E., Ball, J., & Klein-Hessling, J. (2004). School transition from elementary to secondary school: Changes in psychological adjustment. *Educational Psychology*, *24*, 161–173.
Ma, X., Nelson, R. F., Shen, J., & Krenn, H. Y. (2015). Effects of preschool intervention strategies on school readiness in kindergarten. *Educational Research for Policy and Practice*, *14*, 1–17.
Madjar, N., & Cohen-Malayev, M. (2016). Perceived school climate across the transition from elementary to middle school. *School Psychology Quarterly*, *31*, 270–288.
Marks, H. M. (2000). Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. *American Educational Research Journal*, *37*, 153–184.
Martínez, R. S., Aricak, O. T., Graves, M. N., Peters-Myszak, J., & Nellis, L. (2011). Changes in perceived social support and socioemotional adjustment across the elementary to junior high school transition. *Journal of Youth and Adolescence*, *40*, 519–530.
McIntosh, K., Flannery, K. B., Sugai, G., Braun, D. H., & Cochrane, K. L. (2008). Relationships between academics and problem behavior in the transition from middle school to high school. *Journal of Positive Behavior Interventions*, *10*, 243–255.
Missouri Department of Elementary and Secondary Education. Available at: http://www.dese.mo.us. Accessed November 2015.
Missouri Department of Elementary and Secondary Education. Available at: http://www.dese.mo.us. Accessed May 2016.
Myers, C. L., Gross, A. D., & McReynolds, B. M. (2014). Broadband behavior rating scales as screeners for autism? *Journal of Autism and Developmental Disorders*, *44*, 1403–1413.
National Middle School Association & National Association of Elementary School Principals (2002). Supporting students in their transition to middle school. Retrieved from http://wwnatiw.nppsd.org/vimages/shared/vnews/stories/525d81ba96ee9/Tr%20-%20Supporting%20Students%20in%20Their%20Transition%20to%20Middle%20School.pdf.
Pianta, R. C., & Allen, J. P. (2008). Building capacity for positive youth development in secondary school classrooms: Changing teachers' interactions with students. In M. Shinn, & H. Yoshikawa (Eds.), *Toward positive youth development: Transforming schools and community programs* (pp. 21–39). Oxford, UK: Oxford University Press.
Randall, J., & Engelhard, G. (2009). Differences between teachers' grading practices in elementary and middle schools. *The Journal of Educational Research*, *102*, 175–186.
Reinke, W. M., Herman, K. C., & Newcomer, L. (2016). The brief student-teacher classroom interaction observation: Using dynamic indicators of behaviors in the classroom to predict outcomes and inform practice. *Assessment for Effective Intervention*, *42*, 32–42.
Reinke, W. M., & Newcomer, L. (2010). *Student teacher classroom interaction observation (STCIO)*. Columbia, MO: University of Missouri.
Rimm-Kaufman, S. E., Pianta, R. C., & Cox, M. J. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly*, *15*, 147–166.
Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal*, *48*, 268–302.
Rudolph, K. D., Lambert, S. F., Clark, A. G., & Kurlakowsky, K. D. (2001). Negotiating the transition to middle school: The role of self-regulatory processes. *Child Development*, *72*, 929–946.
Ryan, A. M., Shim, S. S., & Makara, K. A. (2013). Changes in academic adjustment and relational self-worth across the transition to middle school. *Journal of Youth and Adolescence*, *42*, 1372–1384.
Seidman, E., Allen, L., Aber, J. L., Mitchell, C., & Feinman, J. (1994). The impact of school transitions in early adolescence on the self-system and perceived social context of poor urban youth. *Child Development*, *65*, 507–522.
Simmons, R. G., & Blyth, D. (1987). *Moving into adolescence: The impact of pubertal change and school context*. Hawthorne, NY: Aldine de Gruyter.
Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, *75*, 417–453.
Slonski-Fowler, K. E., & Truscott, S. D. (2004). General education teachers' perceptions of the prereferral intervention team process. *Journal of Educational and Psychological Consultation*, *15*, 1–39.
Snow, K. L. (2006). Measuring school readiness: Conceptual and practical considerations. *Early Education and Development*, *17*, 7–41.
Steinberg, L. (2005). Cognitive and affective development in adolescence. *Trends in Cognitive Sciences*, *9*, 69–74.
Stoep, A. V., McCauley, E., Thompson, K. A., Herting, J. R., Kuo, E. S., Stewart, D. G., ... Kushner, S. (2005). Universal emotional health screening at the middle school transition. *Journal of Emotional and Behavioral Disorders*, *13*, 213–223.
Stormont, M., Herman, K. C., Reinke, W. M., King, K. R., & Owens, S. (2015). The kindergarten academic and behavior readiness screener: The utility of single-item teacher ratings of kindergarten readiness. *School Psychology Quarterly*, *30*, 212–228.
Stormont, M., Reinke, W. M., & Herman, K. C. (2011). *The kindergarten academic behavior readiness screener (K-ABR)*. Columbia, MO: University of Missouri.
Stormont, M., Reinke, W. M., & Herman, K. C. (2013). *The middle school academic behavior readiness screener (M-ABR)*. Columbia, MO: University of Missouri.
Stormont, M., Reinke, W., Herman, K., & Lembke, E. (2012). *Tier two interventions: Academic and behavior supports for children at risk for failure*. New York, NY: Guilford Press.
Stormont, M., Thompson, A., Herman, K. C., & Reinke, W. M. (2016). The social and emotional dimensions of a single item overall school readiness screener and its relations with academic outcomes. *Assessment for Effective Intervention*, 1–10.
Stormshak, E. A., Bierman, K. L., Bruschi, C., Dodge, K. A., Coie, J. D., & Conduct Problems Prevention Research Group (1999). The relation between behavior problems and peer preference in different classroom contexts. *Child Development*, *70*, 169–182.
Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristic curves. *Canadian Journal of Psychiatry*, *52*, 121–128.
Tapp, J. (2004). Multi-option observation system for experimental studies (MOOSES). Retrieved from http://kc.vanderbilt.edu/mooses/mooses.html.
Tapp, J., Wehby, J., & Ellis, D. (1995). A multiple option observation system for experimental studies: MOOSES. *Behavior Research Methods, Instruments, & Computers*,

27, 25–31.
Theriot, M., & Dupper, D. (2009). Student discipline problems and the transition from elementary to middle school. *Education and Urban Society*, *42*, 205–222.
U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2001). *Digest of educational statistics 2001*. Washington, DC: Author.
U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2003). *The condition of education 2003*. Washington, DC: Author.
White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin*, *91*, 461–481.
Wigfield, A., Eccles, J., Mac Iver, D., Reuman, D., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology*, *27*, 552–565.
Williford, A. P., Brisson, D., Bender, K. A., Jenson, J. M., & Forrest-Bank, S (2011). Patterns of aggressive behavior and peer victimization from childhood to early adolescence: A latent class analysis. *Journal of Youth and Adolescence*, *40*, 644–655.
Yovanoff, P., & Squires, J. (2006). Determining cutoff scores on a developmental screening measure: Use of receiver operating characteristics and item response theory. *Journal of Early Intervention*, *29*, 48–62.