**Introduction**

The growing adoption of multi-tiered systematic decision making models in schools, such as Response to Intervention (RTI), has resulted in an increased need for quality methods of assessment. It has been recommended that these approaches be not only psychometrically sound (e.g., valid, reliable, diagnostically accurate), but also efficient, useful, usable, and acceptable (Glover & Albers, 2007) given that the large volume of data collected through these models necessitates that data be obtained at low cost to all involved. Within the area of school-based academic assessment, curriculum-based measurement (CBM) has been identified as one such approach, with research supportive of both its technical adequacy, as well as its efficiency and usability (Wayman, Wallace, Ritch, & Espin, 2007). A body of research has supported the use of CBM within both screening and progress monitoring applications is also noteworthy, as this (a) limits the amount of training that must be provided to users, and (b) creates a link between data collected at baseline (e.g., during screening) and intervention (e.g., during progress monitoring), allowing for an evaluation of child responsiveness to intervention. Unfortunately, research within the area of behavioral assessment has not as of yet identified a CBM equivalent. Although technically defensible approaches have been identified, these have rarely been found to also possess the characteristics of efficiency and flexibility across assessment purposes. As such, the vast majority of schools have yet to adopt screening practices for the purpose of identifying individuals at risk for behavioral difficulty (Romer & McIntosh, 2005). It has been suggested that such limited adoption is due to the lack of viable options. In other words, high screening methods that have high technical defensibility are also efficient, low-cost, and acceptable to users and key stakeholders. As such, the field of behavioral screening is in need of expansion (Chafouleas, Kilgus, & Lovaas, 2010). Recent research has suggested that Direct Behavior Rating – Single Item Scales (DBR-SIS) may be a viable approach to universal screening for behavioral difficulty (Chafouleas, Kilgus, & Hernandez, 2009; Kilgus, Chafouleas, Riley-Tillman & Welsh, 2011). As such, the purpose of the research described herein was to examine the diagnostic accuracy of DBR-SIS as a behavioral screening tool.

**Method**

Overall, 1,016 K-8 students attending schools in Connecticut were enrolled in the current study. Prior to analysis, all data were screened with regard to multiple inclusion criteria. Subsequent to review, a total of 83! students remained (female N=418; 50.3%). Overall, 13.2 percent of students within the current sample were at-risk for behavioral difficulty as determined by the BESS. Of the 66 teachers enrolled, eight were removed from analytic consideration as a result of insufficient data collection for all of their enrolled students.

Students were randomly selected for participation by the researchers, with teacher s rating around 15 students in their class. Each teacher’s group of student participants was separated into 2-3 subgroups. Teachers rated all students in the first subgroup twice a day (i.e., morning and afternoon) for five days across three DBR-SIS targets (disruptive behavior [DBR-DB], academic engagement [DBR-AE], and academic performance [DBR-PPF]). Once completed, the teacher rated all students in the group using the Student Risk Screening Scale (Draummond, 1993; SRSS) and Behavioral and Emotional Screening System (Kamphaus & Reynolds, 2007; BESS). Ratings were to correspond to the behavior displayed by the student during DBR-SIS data collection. This process was repeated for all groups until all randomly selected students had been rated across all assessment methods.

**Results**

Differences between AUCs were inconsistent across grades and grade groups. For instance, although DBR-PPF AUC was statistically significantly lower than the AUC associated with all other scales in 4th grade, it was not significantly higher or lower than any other scale in 7th grade. Overall, relative to other grades and groups, fewer significant differences between scales in AUC were found within the 7th grade and Middle group.

**Summary and Conclusions**

Consistent with prior DBR-SIS screening research (Kilgus et al., 2011), moderate to strong correlations between DBR-SIS targets and the BESS supported the concurrent validity of DBR as screens. Resulting AUCs and predictive validity indices suggested DBR-SIS targets were more accurate in lower grades and grade groupings, and less so at middle school grades. The best approach to screening varied by grade and grade grouping, with DBR-DB found to be best in early elementary, DBR-AE best in late elementary, and DBR-Factor best in middle school. DBR-SIS targets were not associated with high levels of all predictive validity indices. Rather, cut scores considered optimal for universal screening offered high SN and NPP, and low to moderate SP and NP. This is consistent with prior behavioral screening research, which has suggested that most screening measures are not high across all indices (Levitt et al., 2007). In contrast to past DBR-related screening research (Kilgus et al., 2011), combining DBR-SIS targets did not result in improved decision-making. Rather, DBR-SIS combinations were associated with a relatively higher proportion of incorrect decisions in the current sample.