Using ROC Curves to Determine Risk with Direct Behavior Rating

Austin H. Johnson[1], Faith G. Miller[1], Megan E. Welsh[1], Sandra M. Chafouleas[1],
T. Chris Riley-Tillman[2], & Gregory A. Fabiano[3]
University of Connecticut[1], University of Missouri[2], & University at Buffalo[3]

Introduction

The use of multi-tiered service delivery frameworks necessitates the proactive implementation of screening instruments in order to identify students who would benefit from intervention services before acute academic or behavioral problems arise. As a result, a number of screening instruments for use in schools have been developed to measure student risk for both academic and behavioral problems. The hallmarks of effective assessment tools within a multi-tiered framework include (a) efficiency, (b) defensibility, (c) repeatability, and (d) flexibility (Chafouleas, 2011). Direct Behavior Rating (DBR) is unique in that it provides the benefits of both a rating scale and systematic direct observation of behavior (Chafouleas, Christ, Riley-Tillman, Briesch, & Chanese, 2007). Direct Behavior Rating has been previously demonstrated to be sensitive to change over time, efficient to utilize, and flexible in implementation. However, it is unclear if DBR can be utilized as a screener with cut points that would accurately detect when behavioral risk is present or not present. This poster presentation reviews the first year of data from a multi-year, multi-site study examining the characteristics of cut points when using DBR to screen for student risk for behavior problems.

Method

Participants. Approximately 1800 public-school students enrolled in 192 lower elementary (1st and 2nd), upper elementary (4th and 5th grade), and middle school (7th and 8th) classrooms across three states (Missouri, New York, and Connecticut) were enrolled in this study. After site, teacher, and student consent was obtained, ten students were randomly selected from each teacher's roster to participate. As identified at the Fall time point, 52.2% of student participants were male. The racial identity of a majority of participants was identified by school staff as White (82.5%), with 13.0% of the participants identified as African-American and 1.7% as Asian. Most participants were identified as non-Hispanic (92.6%). 13% of students were identified as receiving special education supports as part of a formal special education identification.

Procedures. Prior to data collection, teachers were provided training on assessment procedures. The BASC-BESS (Kamphaus & Reynolds, 2007) and DBR Single Item Scales (DBR-SIS) were completed by teachers at three time points during the Fall, Winter, and Spring of the 2011-12 academic year. DBR-SIS forms consisted of 11-point scales (0-10) measuring three behaviors (Academically Engaged, Disruptive, and Respectful). In order to control for potential order effects, presentation order of the assessments was counterbalanced. Direct Behavior Rating observations were structured such that five students were rated twice-daily for five days. Upon completion of the first group of DBR ratings, the teacher subsequently rated a second group of students for five days. All students were rated on three DBR Single Item Scales: academic engagement, respectful behavior, and disruptive behavior.

Results and Discussion

Figure 1. *Values and 95% Confidence Intervals for Area Under the Curve (AUC) Statistics for Performance of Average Academically Engaged, Disruptive, and Respectful Ratings by Grade Group and Time Point.*
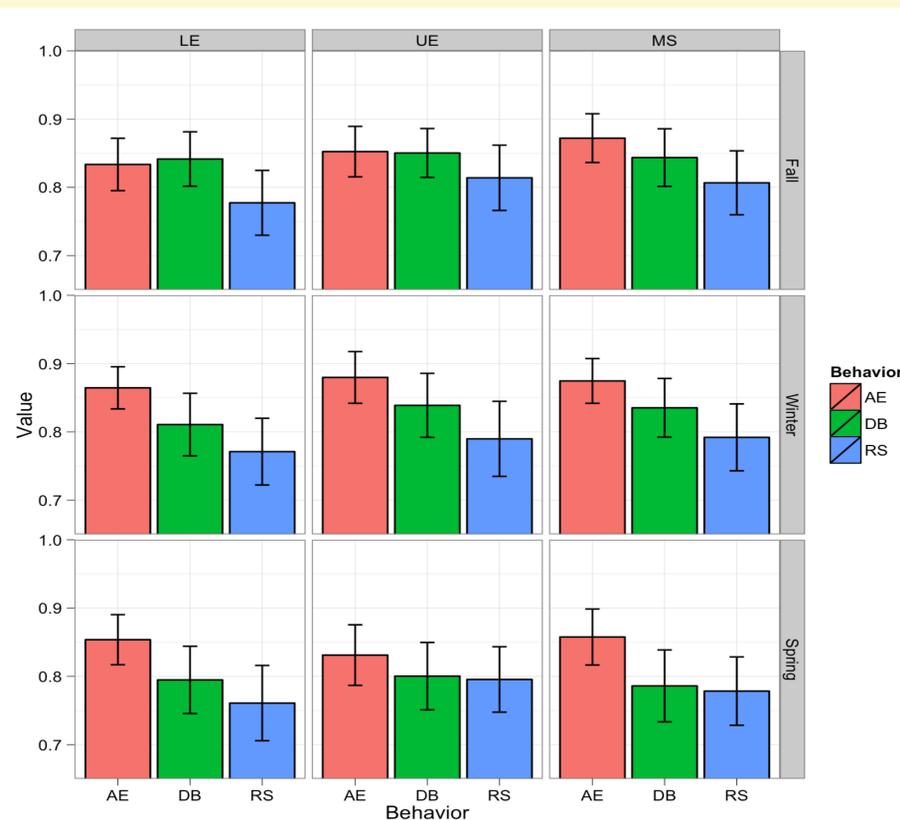

Figure 2. *Values and 95% Confidence Intervals for Sensitivity and Specificity Statistics for Performance of Average Academically Engaged, Disruptive, and Respectful Ratings by Grade Group and Time Point.*
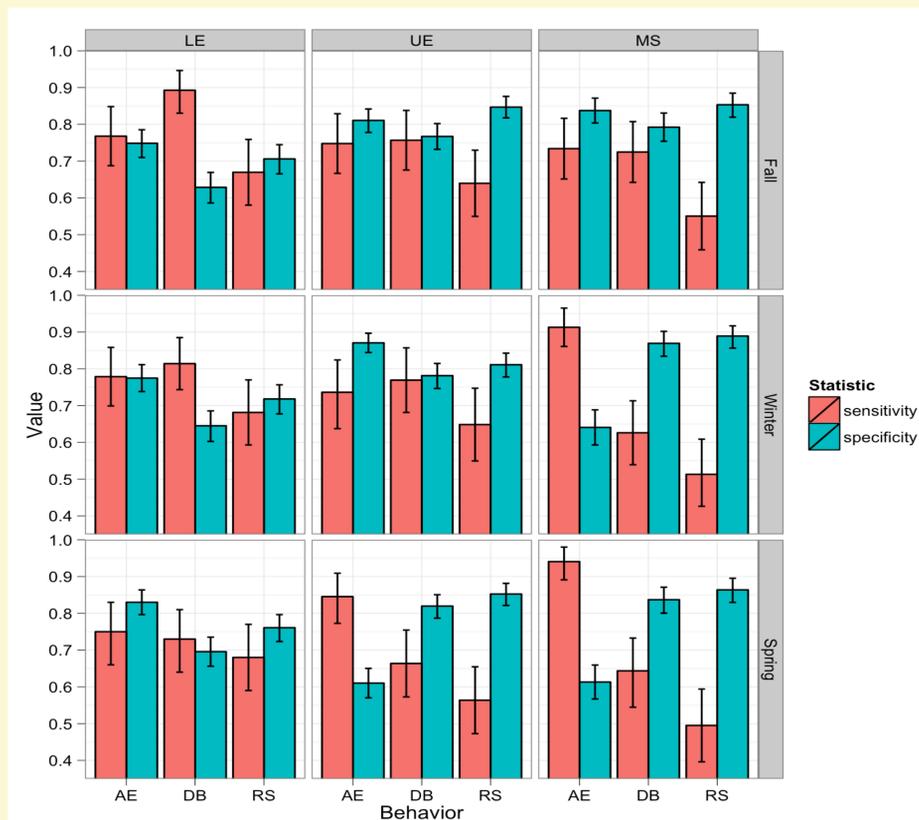
Table 1. *Descriptive Statistics for Sample Size, BESS Risk Status, and DBR Scores by Grade Group and Time Point.*

		BESS				DBR AE	DBR DB	DBR RS
	T (n)	S (n)	No risk	Risk	% Risk	M (SD)	M (SD)	M (SD)
Fall								
LE	60	585	482	102	17.47	8.42 (1.43)	1.27 (1.43)	9.04 (1.25)
UE	70	676	566	108	16.02	8.62 (1.35)	0.89 (1.30)	9.39 (1.08)
MS	61	558	452	106	19.00	8.57 (1.55)	0.86 (1.34)	9.40 (1.13)
Winter								
LE	58	565	453	112	19.82	8.48 (1.34)	1.22 (1.34)	9.06 (1.19)
UE	68	656	562	94	14.33	8.82 (1.17)	0.75 (1.08)	9.39 (1.00)
MS	58	532	415	117	21.99	8.66 (1.50)	0.72 (1.20)	9.46 (1.04)
Spring								
LE	58	566	469	97	17.14	8.71 (1.25)	1.00 (1.20)	9.16 (1.20)
UE	68	656	547	108	16.49	8.90 (1.18)	0.73 (1.24)	9.45 (1.03)
MS	59	538	421	105	19.96	8.75 (1.37)	0.72 (1.17)	9.38 (1.15)

Results indicate that different behaviors may perform with differential effectiveness as diagnostic screeners for behavioral risk as determined by scores on the BASC-BESS depending on (a) the time point at which screening took place, and (b) the grade level of the target student. In Lower Elementary, Academic Engagement ratings generally provided the most optimal balance of diagnostic accuracy statistics across time points. In Upper Elementary, Disruptive ratings generally provided a desirable balance during the Fall and Winter time points. However, this was less true during the Spring, when both Academic Engagement and Disruption were skewed in their ability to correctly specify either risk or no-risk. For Middle School students, Academic Engagement performed at balanced levels in the Fall, with differences between sensitivity and specificity observed for Academic Engagement and Disruptive Behavior in the Winter and Spring.

Across time points and grades, ratings across all behaviors performed significantly better than chance when predicting behavioral risk (p < .05). Academic Engagement ratings generally demonstrated the highest AUC statistics, followed by Disruptive behavior and finally Respectful behavior.