Direct Behavior Rating as a screener of student behavioral risk

Austin H. Johnson[1], Stephen P. Kilgus[2], Sandra M. Chafouleas[1], T. Chris Riley-Tillman[3], Theodore J. Christ[4], Megan E. Welsh[1], Shannon Brooks[2], & Jessica Tomasula[2]

University of Connecticut[1], East Carolina University[2], University of Missouri[3], & University of Minnesota[4]

Introduction

Assessment tools like Curriculum-Based Measurement possess both sufficient evidence and efficiency in their use for supporting data-based decision making for student academics, as these tools can be used for both screening and progress-monitoring purposes. Unfortunately, an equally defensible and efficient tool has yet to be established for behavior. The purpose of this study was to investigate how Direct Behavior Rating – Single Item Scales (DBR-SIS), found by research to be technically adequate behavioral progress monitors, could perform in screening assessment of school-based behavior. Specific research questions were as follows:

1. Which DBR-SIS disruptive behavior (DB), academic engagement (AE), and respect (RS) cut scores allow for the best balance of diagnostic accuracy statistics in determining risk for behavioral difficulty? Do those cut scores vary as a function of grade level and/or research site?
2. Does the use of multiple DBR-SIS targets in a multiple gating fashion result in improved decision making relative to single targets in the determination of risk?
3. Are DBR-based diagnostic decisions more accurate than those associated with chance or the *Student Risk Screening Scale* (Drummond, 1993)?
4. Are DBR-SIS targets concurrently valid indicators of behavioral functioning as measured by other common behavior screeners?

Method

Participants included 1st, 4th, and 7th grade teachers and their students across three US research sites. Participants included 79 1st, 4th, and 7th grade teachers who rated 1,110 students (1st = 411, 4th = 355, 7th = 344) during the winter portion of the 2009-2010 academic year. All students were rated across three assessment methods, including the three DBR-SIS targets, the *Behavior Assessment System for Children – 2, Behavioral and Emotional Screening System* (BESS; Kamphaus & Reynolds, 2007), and the *Student Risk Screening Scale* (SRSS; Drummond, 1993). The BESS served as the criterion against which DBR-SIS targets were compared in determination of DBR-SIS diagnostic accuracy and optimal cut scores. DBR-SIS targets were considered relative to each other and the SRSS in comparisons of diagnostic accuracy.

Students randomly selected by the researchers were divided into 2-3 groups (depending on teacher preference). Teachers completed DBR-SIS ratings of each Group 1 student twice a day for five days across Week 1. Once all DBR-SIS data had been collected, teachers completed both the BESS and SRSS for all Group 1 students by the end of Week 2. Ratings on both measures were to correspond to the behavior displayed by each student during Week 1 only. DBR-SIS ratings of each Group 2 student began on the first day of Week 2, and continued for five days. Prior to the end of Week 3, both the BESS and SRSS were completed for Group 2 students. This process continued until all randomly selected students had been rated on each measure.

Preparation of this poster was supported by a grant from the Institute for Education Sciences (IES), U.S. Department of Education (R324B060014). For additional information, please direct all correspondence to the Project Director, Dr. Sandra M. Chafouleas at sandra.chafouleas@uconn.edu.

Results

Bivariate correlations between all measures were in the expected direction and statistically significant at the $p < .001$ level. Highest correlations between the BESS and all other measures were observed in the 7th grade, with the 1st grade evidencing the second highest. Strongest correlations with the BESS were observed in the SRSS, followed by AE, DB, and RS.

Results of receiver operating characteristic (ROC) curve analyses indicated that all DBR-SIS targets were statistically significantly better than chance in predicting student risk on the BESS. Differences in diagnostic accuracy between DBR-SIS targets were predominantly non-statistically significant. In contrast, across all grades, the SRSS was found to consistently outperform both DB and RS, but not AE to a statistically significant degree. See Table 1.

Results indicate that across each grade, scores of DB ≥ 1, AE ≤ 8, and RS ≤ 9 were best suited for use as screening cut scores. Each score maximized sensitivity, while maintaining adequate levels of specificity. Please see Table 2 for a summary of conditional probability statistics associated with each DBR-SIS target by grade. Finally, findings suggested that increased specificity was gained by requiring students to be at-risk on multiple DBR-SIS targets. Although found to be associated with inappropriately low sensitivity in 4th grade, the DB+AE multiple gating procedure maintained adequate SN (Kettler & Feeney-Kettler, 2011), while achieving a more optimal balance with specificity. See Table 3.

Table 1

AUC Results

	1st Grade				4th Grade				7th Grade			
Scale	AUC	SE	CI-95	*p*	AUC	SE	CI-95	*p*	AUC	SE	CI-95	*p*
DB	.85	.03	.81-.88	<.001	.74	.03	.69-.78	<.001	.83	.03	.79-.87	<.001
AE	.84	.02	.80-.88	<.001	.84	.03	.80-.88	<.001	.87	.02	.83-.91	<.001
RS	.77	.03	.73-.81	<.001	.68	.03	.63-.73	<.001	.79	.03	.74-.83	<.001
SRSS	.90	.02	.87-.93	<.001	.92	.02	.88-.94	<.001	.92	.02	.88-.94	<.001

Note: AUC = area under the curve, SE = standard error, and CI-95 = 95% confidence interval.

Table 2

Conditional Probability Statistics

		1st Grade				4th Grade				7th Grade			
Grade	Score	SN	SP	PPP	NPP	SN	SP	PPP	NPP	SN	SP	PPP	NPP
DB	1	.90	.52	.25	.97	.74	.61	.33	.90	.87	.65	.37	.95
	2	.79	.82	.43	.96	.50	.88	.51	.87	.66	.86	.54	.91
	3	.60	.93	.60	.93	.38	.95	.67	.85	.42	.94	.62	.87
	4	.52	.94	.62	.92	.20	.98	.71	.82	.39	.98	.81	.87
AE	6	.48	.92	.51	.91	.38	.97	.76	.86	.48	.96	.76	.88
	7	.71	.81	.39	.94	.59	.91	.63	.90	.66	.89	.59	.91
	8	.90	.58	.28	.97	.84	.71	.43	.94	.87	.73	.44	.96
	9	1.00	.27	.20	1.00	.92	.27	.26	.97	.97	.36	.27	.98
RS	6	.23	.99	.82	.88	.04	1.00	.75	.80	.15	.99	.83	.83
	7	.35	.96	.61	.89	.11	.98	.57	.81	.21	.99	.78	.84
	8	.52	.91	.49	.91	.30	.93	.54	.83	.40	.97	.79	.87
	9	.69	.79	.37	.94	.53	.80	.41	.87	.69	.84	.52	.92
SRSS	3	.97	.52	.26	.99	.97	.58	.38	.99	.94	.56	.34	.98
	4	.95	.66	.33	.99	.93	.70	.45	.98	.94	.70	.43	.98
	5	.92	.73	.38	.98	.86	.77	.50	.96	.90	.80	.52	.97
	6	.85	.79	.42	.97	.82	.83	.56	.95	.81	.88	.63	.95

Note: SN = sensitivity, SP = specificity, PPP = positive predictive power, and NPP = negative predictive power.

Table 3

Comparison of Single and Combined Scale Screeners

Grade	Scale (cut score)	SN	SP	PPP	NPP	BR	CC	κ
1st	DB (1)	.90	.52	.25	.97	.54	.58	.21
	AE (8)	.90	.58	.28	.97	.49	.63	.25
	RS (9)	.69	.79	.37	.94	.28	.77	.35
	SRSS (5)	.92	.73	.38	.98	.36	.76	.41
	DB+AE	.84	.70	.33	.96	.38	.72	.33
	DB+RS	.68	.81	.39	.93	.26	.79	.37
	AE+RS	.66	.83	.40	.93	.25	.80	.39
	DB+AE+RS	.66	.83	.41	.93	.24	.81	.39
4th	DB (1)	.74	.61	.33	.90	.47	.63	.24
	AE (8)	.84	.71	.43	.94	.41	.74	.41
	RS (9)	.53	.80	.41	.87	.27	.74	.30
	SRSS (4)	.93	.70	.45	.98	.43	.75	.46
	DB+AE	.66	.80	.47	.90	.29	.77	.41
	DB+RS	.53	.83	.44	.87	.25	.76	.33
	AE+RS	.51	.88	.53	.87	.20	.80	.40
	DB+AE+RS	.51	.88	.54	.87	.20	.81	.40
7th	DB (1)	.87	.65	.37	.95	.45	.69	.34
	AE (8)	.87	.73	.44	.96	.38	.76	.44
	RS (9)	.69	.84	.52	.92	.26	.81	.47
	SRSS (5)	.90	.80	.52	.97	.34	.82	.54
	DB+AE	.82	.79	.49	.95	.33	.80	.49
	DB+RS	.66	.88	.56	.91	.23	.83	.50
	AE+RS	.63	.90	.60	.91	.20	.85	.52
	DB+AE+RS	.63	.90	.61	.91	.20	.85	.52

Note: SN = sensitivity, SP = specificity, PPP = positive predictive power, NPP = negative predictive power, BR = base rate, CC = correct classification rate, and κ = Cohen's kappa statistic.

Summary and Conclusions

Consistent with prior DBR-SIS screening research (Kilgus et al., 2012), moderate to strong correlations between DBR-SIS targets and the BESS supported the concurrent validity of DBR as screeners. Resulting AUCs and conditional probability indices suggested DBR-SIS targets were more accurate overall in 1st and 7th grade, and less so in 4th. The best DBR-SIS-based approach to screening varied by grade, with AE found to be best in 4th grade, and DB+AE best in 1st and 4th. DBR-SIS targets were not associated with high levels of all conditional probability indices. Rather, cut scores considered optimal for universal screening offered higher SN and NPP, and low to moderate SP and NPP.

These results are consistent with prior behavioral screening research, which has suggested that most screening measures are not high across all indices (Levitt et al., 2007). Similar to past DBR-related screening research (Kilgus et al., 2012), the use of DBR-SIS multiple gating procedures lead to improved decision-making. Finally, it should be noted that while DBR-SIS single targets and multiple gating procedures yielded adequate diagnostic accuracy, the SRSS was found to consistently outperform DBR-SIS targets with regard to overall diagnostic accuracy and the balance between high sensitivity and adequate specificity.