Project VIABLE – Diagnostic Accuracy of Direct Behavior Rating Single Item Scales

Stephen P. Kilgus[1], Shannon R. Brooks[2], Sandra M. Chafouleas[1], T. Chris Riley-Tillman[2], & Theodore J. Christ[3]

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

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, Wiley, Tichá, & 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 still 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 usable options. That is, few screening methods high in 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 (Cook, Volpe, & Livanis, 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 screener.

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 inclusionary criteria. Subsequent to review, a total of 831 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 rated 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 *respectfulness* [DBR-RS]). Once completed, the teacher rated all students in the group one using the *Student Risk Screening Scale* (Drummond, 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.

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 Sandra Chafouleas at sandra.chafouleas@uconn.edu

Results

A one-way ANOVA and post hoc tests suggested the creation of three grade groups, including "Early Elementary" (Grades K-2), "Late Elementary" (Grades 3-5), and "Middle" (Grades 6-8). All six scales were compared through the calculation of 15 unique *r* coefficients. All bivariate correlations were statistically significant at the .001 level. Cut scores considered best suited for universal screening for behavioral difficulty were those that maximized SN and NPP while maintaining adequate levels of SP. DBR scales considered best suited within each grade and grade grouping included (1) DBR-DB (cut score = 2) in 1st grade and Early Elementary students, (2) DBR-AE for 4th grade and Late Elementary (cut score = 8), and (3) DBR-Factor (cut score = 0) for 7th grade and Middle school. Relative to all other DBR scales, DBR-RS was not considered to be the best indicator of student risk within any grade or group.

DBR-SIS combinations were associated with higher SP levels, and base rates levels more comparable to those associated with the BESS in regards to risk classification. Yet, the use of DBR-SIS combinations also led to low SN and reduced agreement with the BESS. As such, single scales were considered to consistently offer a more suitable approach to universal screening. Across all grades and grade groupings, the AUC associated with all but two scales fell in the moderate or high range of diagnostic accuracy. The difference in AUC between each scale and random decision making (AUC=0.50) was statistically significant at the .01 level. That is, with the exception of DBR-DB in 7th grade, where *p* was equal to .022.

	Grade 1				Grade 4				Grade 7			
Cut Score	SN	SP	PPP	NPP	SN	SP	PPP	NPP	SN	SP	PPP	NPP
DBR-DB												
1	100.00	39.81	17.3	100.0	**91.30**	**61.62**	**35.6**	**96.8**	**50.00**	**82.98**	**22.6**	**94.4**
2	**92.31**	**83.50**	**41.4**	**98.9**	73.91	93.94	73.9	93.9	21.43	92.91	23.1	92.3
3	53.85	96.12	63.6	94.3	56.52	100.00	100.0	90.8	0.00	97.87	0.0	90.8
DBR-AE												
7	61.54	88.35	40.0	94.8	82.61	93.94	76.0	95.9	14.29	95.04	22.2	91.8
8	**76.92**	**66.02**	**22.2**	**95.8**	**95.65**	**74.75**	**46.8**	**98.7**	57.14	85.82	28.6	95.3
9	100.00	22.33	14.0	100.0	100.00	29.29	24.7	100.0	**85.71**	**48.23**	**14.1**	**97.1**
10	100.00	0.00	11.2		100.00	0.00	18.9		100.00	0.00	9.0	
DBR-RS												
8	38.46	96.12	55.6	92.5	39.13	100.00	100.0	87.6	7.14	99.29	50.0	91.5
9	**69.23**	**89.32**	**45.0**	**95.8**	**65.22**	**89.90**	**60.0**	**91.8**	**50.00**	**90.78**	**35.0**	**94.8**
10	100.00	0.00	11.2		100.00	0.00	18.9		100.00	0.00	9.0	
DBR-Factor												
-2	38.46	98.06	71.40	92.70	43.48	100.00	100.00	88.40	0.00	98.58	0.00	90.80
-1	**84.62**	**91.26**	**55.00**	**97.90**	**78.26**	**94.95**	**78.30**	**94.90**	28.57	95.04	36.40	93.10
0	100.00	18.45	13.40	100.00	100.00	27.27	24.20	100.00	**92.86**	**48.23**	**15.10**	**98.6**
1	100.00	0.00	11.20		100.00	0.00	18.90		100.00	0.00	9.00	

	Early Elementary				Late Elementary				Middle			
Cut Score	SN	SP	PPP	NPP	SN	SP	PPP	NPP	SN	SP	PPP	NPP
DBR-DB												
1	88.00	40.08	21.9	94.6	**90.24**	**60.98**	**26.4**	**97.6**	**57.89**	**83.08**	**25.0**	**95.3**
2	**80.00**	**78.63**	**41.7**	**95.4**	63.41	91.29	53.1	94.1	26.32	93.85	29.4	92.9
3	58.00	90.46	53.7	91.9	43.90	98.48	81.8	91.9	0.00	98.46	0.0	91.0
DBR-AE												
7	64.00	84.73	44.4	92.5	68.29	92.80	59.6	95.0	10.53	96.41	22.2	91.7
8	**76.00**	**66.79**	**30.4**	**93.6**	**87.80**	**73.11**	**33.6**	**97.5**	63.16	85.13	29.3	96.0
9	94.00	29.39	20.3	96.2	100.00	36.36	19.6	100.0	**89.47**	**47.18**	**14.2**	**97.9**
10	100.00	0.00	16.0		100.00	0.00	13.4		100.00	0.00	8.9	
DBR-RS												
7	30.00	94.66	51.7	87.6	9.76	99.62	80.0	87.7	5.26	100.00	100.0	91.5
8	36.00	90.84	42.9	88.1	29.27	98.48	75.0	90.0	10.53	99.49	66.7	91.9
9	**64.00**	**81.68**	**40.0**	**92.2**	**51.22**	**90.91**	**46.7**	**92.3**	**47.37**	**90.26**	**32.1**	**94.6**
10	100.00	0.00	16.0		100.00	0.00	13.4		100.00	0.00	8.9	
DBR-Factor												
-2	48.00	92.37	54.50	90.30	31.71	98.86	81.30	90.30	0.00	98.97	0.00	91.00
-1	**78.00**	**82.44**	**45.90**	**95.20**	**65.85**	**94.70**	**65.90**	**94.70**	26.32	95.90	38.50	93.00
0	94.00	21.37	18.60	94.90	100.00	32.58	18.70	100.00	**94.74**	**46.15**	**14.60**	**98.90**
1	100.00	0.00	16.00		100.00	0.00	13.40		100.00	0.00	8.90	

Differences between AUCs were inconsistent across grades and grade groups. For instance, although DBR-RS 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.

	SN	SP	PPP	NPP	Base Rate	Kappa
Grade 1						
SRSS (4)	100.00	73.79	32.50	100.00	34.48	0.39
DBR-DB + DBR-AE	76.92	87.38	43.48	96.77	19.83	0.48
DBR-DB + DBR-RS	69.23	92.23	52.94	95.96	14.66	0.54
DBR-AE + DBR-RS	61.54	89.32	42.11	94.85	16.38	0.42
DBR-DB + DBR-AE+ DBR-RS	61.54	92.23	50.00	95.00	13.79	0.49
Grade 4						
SRSS (4)	91.30	75.76	46.70	97.40	36.89	0.49
DBR-DB + DBR-AE	86.96	83.84	55.56	96.51	29.51	0.58
DBR-DB + DBR-RS	65.22	92.93	68.18	92.00	18.03	0.59
DBR-AE + DBR-RS	65.22	93.94	71.43	92.08	17.21	0.61
DBR-DB + DBR-AE+ DBR-RS	65.22	93.94	71.43	92.08	17.21	0.61
Grade 7						
SRSS (4)	71.43	85.11	32.30	96.80	20.00	0.37
DBR-DB + DBR-AE	50.00	82.98	22.58	94.35	20.00	0.21
DBR-DB + DBR-RS	35.71	96.45	50.00	93.79	6.45	0.37
DBR-AE + DBR-RS	42.86	90.78	31.58	94.12	12.26	0.29
DBR-DB + DBR-AE+ DBR-RS	35.71	96.45	50.00	93.79	6.45	0.37

	Grade 1			Grade 4			Grade 7		
Scale	AUC	*p*[1]	95% CI[2]	AUC	*p*[1]	95% CI[2]	AUC	*p*[1]	95% CI[2]
DBR-DB	.924	**<.001	.86-.97	.896	**<.001	.83-.94	.663	*.022	.58-.74
DBR-AE	.818	**<.001	.74-.88	.937	**<.001	.88-.97	.753	**<.001	.68-.82
DBR-RS	.804	**<.001	.72-.87	.795	**<.001	.71-.86	.705	**.004	.63-.78
DBR-Factor	.904	**<.001	.84-.95	.907	**<.001	.84-.95	.754	**<.001	.68-.82
SRSS	.956	**<.001	.90-.99	.933	**<.001	.87-.97	.798	**<.001	.73-.86
	Early Elementary			Late Elementary			Middle		
DBR-DB	.823	**<.001	.78-.86	.857	**<.001	.81-.89	.707	**.001	.64-.77
DBR-AE	.810	**<.001	.76-.85	.893	**<.001	.85-.93	.780	**<.001	.72-.83
DBR-RS	.740	**<.001	.69-.79	.720	**<.001	.67-.77	.692	**.002	.63-.75
DBR-Factor	.823	**<.001	.78-.86	.864	**<.001	.82-.90	.755	**<.001	.69-.81
SRSS	.928	**<.001	.89-.95	.896	**<.001	.86-.93	.848	**<.001	.79-.89

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 screeners. 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 NPP. 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.