Project VIABLE - Direct Behavior Rating: Evaluating Behaviors with Positive and Negative Definitions

Rose Jaffery[1], Albee T. Ongusco[3], Amy M. Briesch[1], Theodore J. Christ[2], Sandra M. Chafouleas[1], T. Chris Riley-Tillman[3]

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

Introduction

Direct Behavior Rating (DBR) is an assessment method which combines rating scale and direct observation procedures in order to provide a feasibly collected stream of data useful for decision-making in a problem-solving model. For example, using a single-item DBR scale, a teacher might rate the percentage of time disruptive behavior was displayed during math class across occasions in order to evaluate student response to various intervention supports. Although research has begun to document defensibility and usability of single item DBR scales, further research is needed to establish firm guidelines for use.

In an initial study, Riley-Tillman, Chafouleas, Christ, Briesch and LeBel (2009) investigated the impact of alternate definitions of behaviors using single-item DBR with an 11-point scale (0-10; Figure 1). Findings suggested that DBR data of general outcome behaviors (*academically engaged, disruptive*) were more consistent with systematic direct observation (SDO) data than were DBR data of specific behaviors (*hand raising, calling out*). Also, connotative wording (positive, negative) of behaviors might influence the accuracy of DBR data. The purpose of this study was to replicate previous findings to determine which behavior targets yield the most accurate ratings and how to connotatively define those behaviors. The current study also aimed to extend the concept of rating inaccuracy to include both random and systematic inaccuracy, and also evaluate whether or not the base rate at which a behavior occurs during a rating session influences inaccuracy. These evaluations are necessary for establishing recommendations regarding target behaviors in DBR instrumentation.

Method

Participants included 88 undergraduate students who were randomly assigned to one of two conditions (either positive or negative wording of target behaviors) and given a corresponding DBR packet that listed six target behaviors and rating scales (see Table 1). Following brief instruction on using DBR, participants viewed five 2-minute video clips of classroom instruction in an elementary school. Following each clip, participants used the corresponding DBR form to rate each of two target students on the target behaviors.
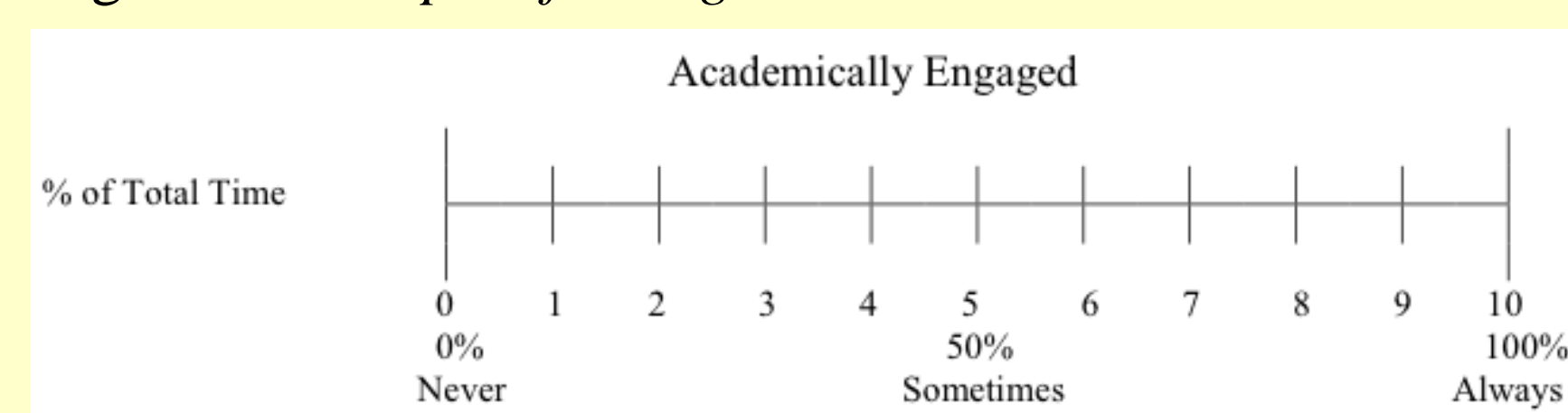
Figure 1. *Example of a single-item DBR scale*

The following week, participants attended a second session, during which they viewed the same six video clips but used the alternately worded rating packet. For example, if at session 1, a participant was given a DBR form with positively worded target behaviors, then that participant used a DBR form with negatively worded behaviors at session 2. Ultimately, every participant rated each of the five video clips using both types of wording, thereby resulting in a fully crossed design. The outcome variable of interest was the rating assigned by the participant to the target student's behavior. Trained graduate student researchers determined the true state of behaviors through second by second coding of each clip (using SDO) to determine the percentage of time the target behavior was exhibited. To assess accuracy, participant ratings were compared to researcher ratings.

Table 1. *Target Behaviors Evaluated*

Positive	Negative
Academically Engaged†	Academically Unengaged†
Well-behaved†	Disruption†
Appropriate Verbal Behavior	Inappropriate Verbal Behavior
Appropriate Interaction with Teacher	Inappropriate Interaction with Teacher
Appropriate Interaction with Peers	Inappropriate Interaction with
Appropriate Motor Behavior	Inappropriate Motor Behavior

† Behaviors also examined in Riley-Tillman, Chafouleas, Christ, Briesch & LeBel (2009).

Results

Base rates for the behaviors *academically engaged* and *disruptive* fell within an acceptable range (Table 2). Correlation coefficients associated with ratings of *academically engaged* and *disruptive* were most robust (Table 3). Differences and absolute differences between SDO and DBR were tested for statistically significant differences within behaviors and across wording condition. Estimates close to zero for the difference scores (DBR – SDO) are preferable to indicate that, on average, DBR would reliably estimate SDO scores (i.e., low systematic inaccuracy). Estimates close to zero for absolute differences scores (|DBR – SDO|) are preferable to indicate that the magnitude of the difference between DBR and SDO is low (i.e., low random inaccuracy).

Table 2. *Descriptive Statistics for SDO, DBR, & Difference Scores*

Connotative Condition	Scaled Scores[a]				Difference			
	SDO		DBR		DBR – SDO		|DBR – SDO|	
	M	SD	M	SD	M	SD	M	SD
Academically Engaged								
Positive	6.20	2.93	5.97	3.49	-.22	2.43	1.79	1.65
Negative	3.90	2.88	4.09	3.61	.19	2.72	2.02	1.82
Disruption								
Positive	8.00	2.06	5.82	3.66	-2.17	2.42	2.44	2.15
Negative	2.10	2.02	3.96	3.81	1.86	2.57	2.34	2.14
Interaction with Teacher								
Positive	1.30	1.15	6.09	3.59	4.79	4.11	5.42	3.23
Negative	.95	1.16	3.59	3.75	2.64	3.5	2.99	3.20
Interaction with Peer								
Positive	.00	.05	6.20	3.55	6.2	3.55	6.2	3.55
Negative	.10	.30	2.90	3.51	2.8	3.24	2.81	3.23
Motor Behavior								
Positive	1.36	1.2	5.23	3.69	3.87	3.85	4.41	3.22
Negative	1.85	1.93	4.37	3.74	2.52	3.62	3.36	2.85
Verbal Behavior								
Positive	.05	.27	6.55	3.58	6.50	3.60	6.51	3.59
Negative	.45	.74	2.61	3.28	2.16	3.08	2.33	2.95

Note. N = 88, All distributions were within acceptable limits for kurtosis and skew (z-score < |2.00|).
[a] Results reported with a scaled score range of 0 to 11; the procedures to convert SDO to DBR scaled score values appear in the method section.

Table 3. *Correlational Analysis of SDO and DBR within Six Behaviors and Across Connotative Wording Conditions*

Target Behavior		Positively Worded		Negatively Worded	
		SDO Scaled Score	SDO Raw Score	SDO Scaled Score	SDO Raw Score
Academically Engaged					
	DBR Scaled Score	.73**	.71**	.67**	.67**
	SDO Scaled Score		.99**		.99**
Disruption					
	DBR Scaled Score	.78**	.76**	.78**	.77**
	SDO Scaled Score		.99**		.99**
Interaction with Teacher					
	DBR Scaled Score	-.33**	-.36**	.37**	.35**
	SDO Scaled Score		.97**		.98**
Interaction with Peer					
	DBR Scaled Score	.01	.04	.28**	.25**
	SDO Scaled Score		.74**		.89**
Motor Behavior					
	DBR Scaled Score	.03	-.01	.32**	.32**
	SDO Scaled Score		.97**		.99**
Verbal Behavior					
	DBR Scaled Score	-.05*	.02	.37**	.39**
	SDO Scaled Score		.94**		.96**

Note. N = 88, coefficients are only presented for within positively and negatively worded conditions and not across Conditions.
** p < .01, * p < .05

Repeated measures ANOVA was used to identify scores significantly different from zero. Results were statistically significant (*p* < .01) within each behavior when SDO and DBR difference scores were used, indicating systematic inaccuracy. Estimates of effect size using Eta-squared exceeded estimates for moderate or large effect sizes for a subset of behaviors: *interaction with peer, disruptive,* and *verbal behavior*. Effects of wording across conditions for systematic inaccuracy were trivial/small for *academically engaged* and *motor behavior*. Results were statistically significant (*p* < .01) within each behavior except *disruptive* when absolute differences (random inaccuracy) were examined. Estimates of effect size were moderate or large for a subset of behaviors: *interaction with teacher, interaction with peers,* and *verbal behavior*. The effects of connotative wording were small or trivial for *academically engaged, motor behavior,* and *disruptive*.

SDO and DBR data were fairly consistent across target students. Data for well-behaved/disruptive and academically engaged/unengaged are presented for one of the target students in Figures 2 and 3, respectively.

The mean difference between DBR and SDO was -2.17 for *well-behaved* and 1.86 for *disruptive*. Those data are indicative of a consistent (systematic) bias among ratings, an example of which is illustrated in Figure 2.

Figure 2. SDO and mean DBR data of *Well-Behaved/Disruptive* for one of the target students for each of the five 2-minute observation/rating periods.
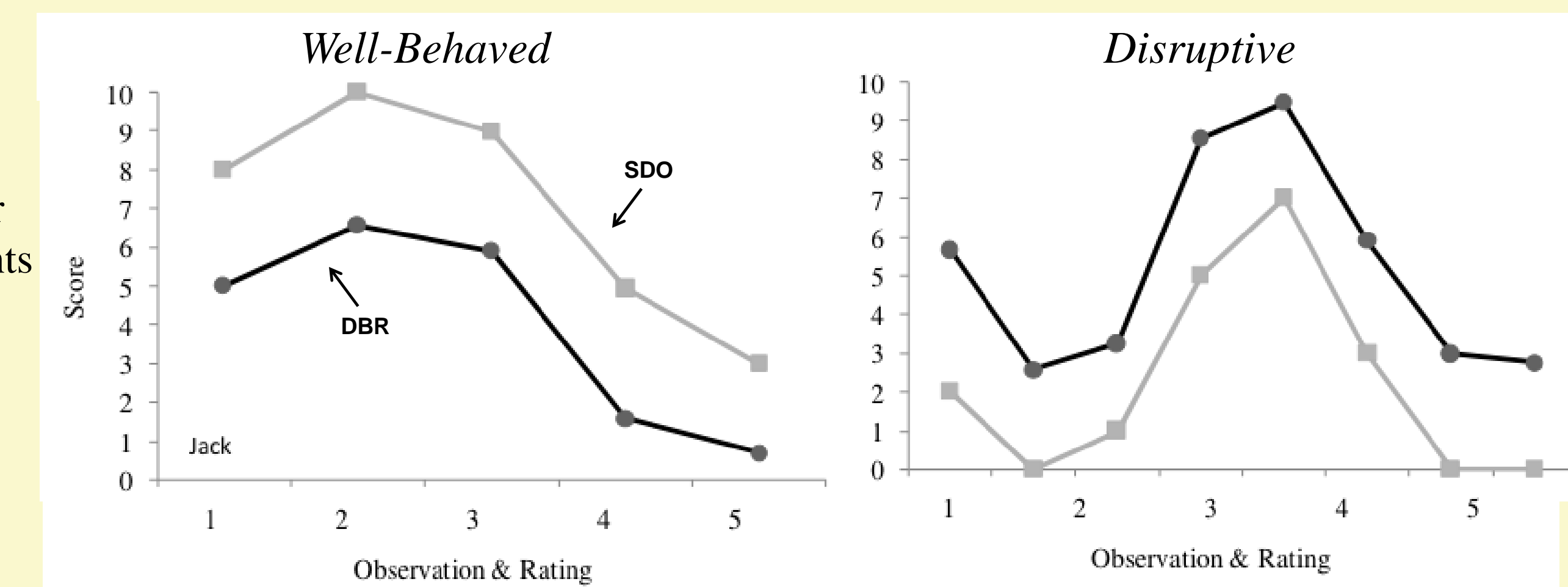
• Notice that the profiles of DBR and SDO data are similar; however, DBR data for *well-behaved* are lower than those of SDO.
• The reverse occurs for *disruptive* so that DBR data are above those for SDO. On average, raters using DBR underestimated *well-behaved* and overestimated *disruptive* with approximately the same magnitude.

No such evidence of bias is apparent in the DBR and SDO data for *academically engaged/unengaged*, an example of which is illustrated in Figure 3.

Figure 3. SDO and mean DBR data of *Academically Engaged/Unengaged* for one of the target students for each of the five 2-minute observation/rating periods.
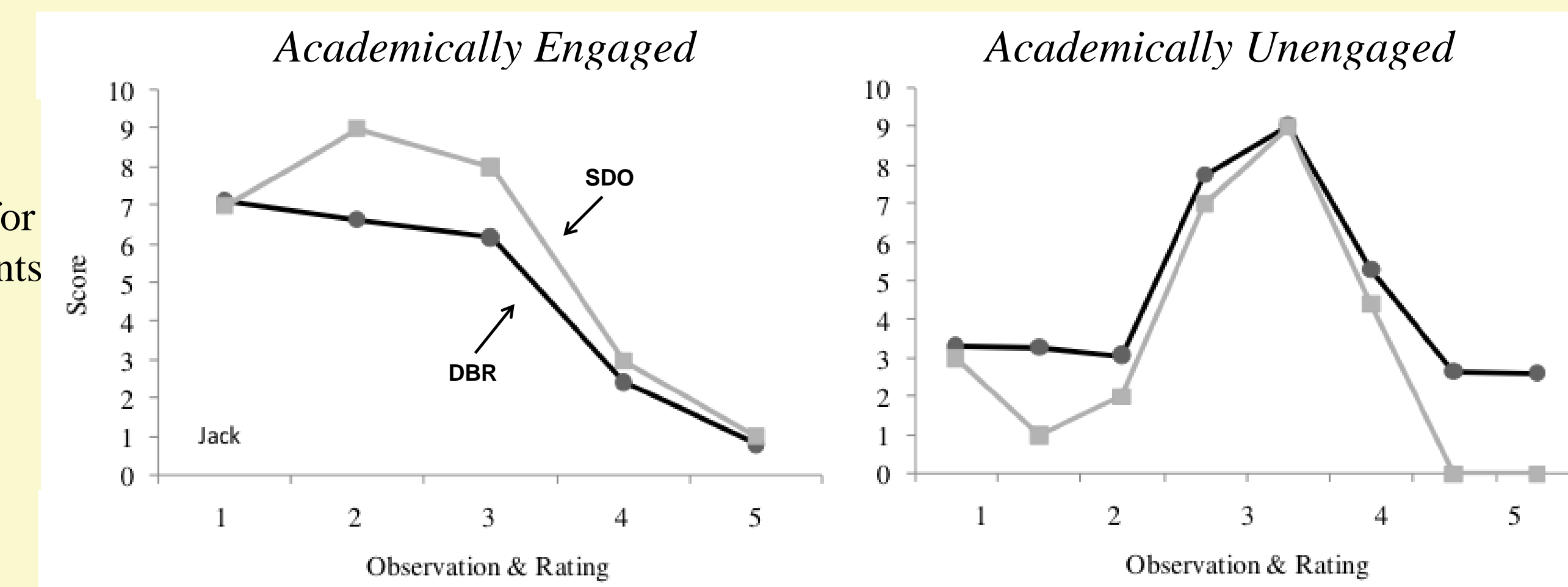
Summary and Conclusions

The general outcome behaviors of *academically engaged* and *disruptive* were the best in terms of criterion related validity and boasted lower magnitudes of random and systematic inaccuracy. This contributes to the accumulating evidence that DBR data can be used in practice to guide a variety of assessment decisions. Although significant differences across connotative wording conditions for *academically engaged* and *disruptive* appeared, effect sizes were trivial except for a bias among raters to underestimate *well-behaved* and overestimate *disruptive*. Results generally provide support for either wording condition for *academically engaged* and *disruptive*.

Consistent with the previous study, ratings of the more specific behaviors (e.g., *motor behavior*) corresponded with less accurate ratings and substantial differences across wording conditions. For these behaviors, negatively worded definitions produced difference scores that were lower in magnitude, suggesting descriptions should be negatively worded if specific behaviors are used. Finally, behaviors with low base rates tended to correspond with lower quality ratings in both this and the prior study. Therefore, it might be the base rate at which behaviors occur that influences rating accuracy rather than level of generality/specificity.