

Northeastern University Bouvé College of Health Sciences University of Connecticut

East Carolina University



## **Statement of the Problem**

•Given increased demand for accountability and data-based decision making in schools, the search is underway to identify appropriate measurement tools for social behavior. •Although systematic direct observation (SDO) has typically been regarded as the method of choice when formatively assessing social behavior within research contexts, there exists some concern that the practical utility of SDO data may be limited by its scope in applied settings such as schools. Extant research findings suggest that the sheer number of systematic observations required in order to draw a generalizable picture of student behavior may have the potential to overwhelm existing systems and resources (e.g., Hintze & Matthews, 2004). •Use of direct behavior rating (DBR) may provide a more feasible alternative in terms of frequent data collection; however, some concern has been raised regarding the influence of rater effects (e.g. halo error) on the reliability and accuracy of data. Furthermore, although early evidence appears promising, research regarding the psychometric properties of DBR is still in its infancy. •The purpose of the current study, therefore, was to examine the relative psychometric properties of two extant behavior assessment methods for the purpose of formative assessment: systematic direct observation and direct behavior rating.

## Method

### Data collected over the course of 10 consecutive school days during a 45-minute group lesson (phonemic awareness, phonics, mathematics)

Academica classroom a <u>Exa</u> less <u>Exa</u> inst <u>Exa</u> talk	ally Engaged: Academ activities. amples of Active Engage on amples of Passive Enga ructional materials. amples of Non-Engagen ing/yelling unrelated to	ically engaged behavior is de <u>gement:</u> writing, raising his/he <u>gement</u> : listening to/looking <u>ment</u> : off-task behavior (i.e. s o classroom instruction), leav	efined as actively or p er hand, answering a at the teacher, readir taring into space), dis ing the classroom	passively participating in question, or talking about ng silently, or looking at sruptive behavior (i.e.	a			
:15 :30 :45 1:00		1:15 1:30 1:45 2:00			1 general in an (7	educa n inclu <i>Mean</i>	ation usive age	i + e ki = 5
		Classroom videotaped an coded using m sampling pro	instruction d subsequer nomentary-t ocedures by	ntly ime 2	Plac total ti 0 1	e a mark a me the stu	along the dent was	e line s acad

This study was supported by a grant from the Institute for Education Sciences, U.S. Department of Education (R324B060014). For additional information, please direct all correspondence to Amy Briesch at a.briesch@neu.edu

Briesch, A. M., Chafouleas, S. M., & Riley-Tillman, T. C. (2010, March). Comparing the psychometric properties of behavior assessment methods: Systematic Direct Observation & Direct Behavior Rating. Poster presentation at the National Association of School Psychologists Annual Convention, Chicago, IL.

# Comparing the Psychometric Properties of Behavior Assessment Methods: Systematic Direct Observation & Direct Behavior Rating

50

40

30

20

10

Amy M. Briesch<sup>1</sup>, Sandra M. Chafouleas<sup>2</sup>, & T. Chris Riley-Tillman<sup>3</sup> Northeastern University<sup>1</sup> University of Connecticut<sup>2</sup> East Carolina University<sup>3</sup>

1 special education teacher indergarten classroom years, 11 months)



### Table 1

G Study Full Model Results: Comparison of Variance Components by Rater Type			Rater effects	Table 2							
			•SDO: Both the rater facet (0%) and the persons by rater	Results of Decision Studies with Conditions Varied by Day and Rater Type							
	Teachers	Observers	interaction (0%) contributed negligible variance to the model,								
	% Variance	% Variance	suggesting that <b>neither interobserver agreement nor rater bias</b> were of concern.			1 day	5 days	10 days	15 days	20 days	100 days
Person	47	48	•DBR: One-quarter of the variance in DBR data was attributable to	$1 \text{ observation}/\text{day}^1$							
Day	0	1	rater-related effects.	Researcher-	$E\hat{ ho}^2$	.50	.83	.91	.93	.98	.99
Occasion:Day, Occ x Day	2.5	5	•The fact that overall rating differences between the two teachers were identified, was not surprising, given descriptive	conducted SDO	Φ	.48	.82	.90	.93	.97	.99
Rater	7.5	0	differences noted in teachers' use of the DBR scale (i.e., <i>Mean</i>	Teacher-completed DBR	$E\hat{ ho}^2$	.54	.66	.68	.69	.70	.70
Person x Rater	20	0	rating for Teacher 1 = 90.89%, <i>Mean</i> rating for Teacher 2 =		Φ	.47	.58	.61	.62	.63	.63
Person x Day	0	2	80.32%). •The variance component for the interaction between persons	3 observations/day <sup>2</sup>	<b>D</b> ^2	73	03	96	07	08	00
Rater x Day	2	0	and raters accounted for 20% of rating variance, indicating that	conducted SDO	$E \rho^2$ $\Phi$	.73	.93	.90	.97	.98	.99
Person x Rater x Day	3	1	teachers varied in their perceptions of the relative standing		_					- 0	- 0
Residual	17	44	of particular students (e.g., rater bias effect). •These findings suggest that recordings must either be analyzed	Teacher-completed DBR	$E \hat{ ho}^2 \Phi$	.62 .55	.68 .60	.69 .62	.69 .62	.70 .63	.70 .63
$E\hat{ ho}^2$	.82	.98	within rater, thereby focusing on changes in an individual teacher's perception of the problem, or teachers must engage in	<i>Note.</i> <sup>1</sup> D study results based on collection of one SDO/DBR data point per day. <sup>2</sup> D study results based on collection of three SDO/DBR data points per day. <i>Time effects</i>							
Φ	.77	.97	training in order to minimize reliance on global perception and move closer in line with objective observation								
<i>Note</i> . Residual error comprised <i>Generaliza</i> Jearly 50% (47% teac	of <i>p x o:d, r x o:d, p x</i> <i>bility study resu</i> hers, 48% resea	<i>r x o:d, e</i> <i>elts</i> <b>archers) of the</b>	100 90 90 80 70 60	•In the case of SDO, 30 could be explained by and rating occasion . A cannot be neatly interp SDO data were sensit behavior over time.	J% of the int Althou reted, <b>ive to</b>	the re teracti gh thi it doe <b>fluct</b>	sidua on be s nest s gen uatio	l erro tweer ced in erally <b>ns in</b>	r vari n pers teraci 7 sug <b>stud</b>	ance son, d tion gest tl ent	lay, hat
learly 50% (47% teac riance in scores was a	hers, 48% resea attributable to th	rchers) of the ne facet of		<ul><li>behavior over time.</li><li>In contrast, a much sm</li></ul>	naller	(	13.5°	13.5%) pr	13.5%) proport	13.5%) proportion of	13.5%) proportion of the

variance in scores was attributable to the facet of person, suggesting that both methods were equally sensitive to inter-individual differences in academic engagement.

•The remaining 50% of variance in scores, however, was accounted for in meaningfully different ways depending on the assessment method, thus suggesting different implications for the use of either an SDO or DBR assessment approach.

Results of the current investigation suggest important considerations relevant to the selection of behavior assessment methods across varying assessment purposes. It should not come as a surprise that more dependable estimates of student engagement were obtained using SDO than DBR, given both the precision of the recording procedure (i.e., behavior recorded every 15 seconds) versus once every 15 minutes) and the nature of the rating task (i.e., analogue observation of videotaped behavior versus in vivo rating during regular teaching activities). However, the selection of appropriate assessment tools is rarely based on numbers alone, requiring simultaneous consideration of issues related to defensibility and feasibility. Results of the current study, for example, suggest that a level of reliability sufficient for the purposes of rank-order screening (i.e., .70) may be obtained given the collection of 3 SDO or 20 DBR data points. Although the choice may seem obvious at a surface level, several factors may interact to determine selection. For example, in situations in which decisions need to be made quickly, use of SDO would be the most appropriate choice given that fewer data points would be needed. When time is not a limiting factor, however, school psychologists may find teachers' daily use DBR over the course of a few weeks to be less intrusive (to both the natural classroom ecology and the school psychologist's schedule) than conducting 45-60 minutes worth of targeted observation. Furthermore, of particular importance for those situations in which behaviors occur infrequently or teacher perception of the target behavior is deemed an important outcome variable, it appears that the collection of a sufficient amount of teacher-generated data may serve as a viable alternative to scheduled observations.

## **Results and Discussion**

## Conclusions

Student

Northeastern University Bouvé College of Health Sciences University of Connecticut East Carolina University

### Table 2

•In contrast, a much smaller (13.5%) proportion of the variance in DBR recordings of academic engagement was attributable to time-related facets or interactions.

– 🖛 Researcher 2

--+-- Teacher 1

····∎··· Teacher 2

10 11 12

q

•One potential explanation for the **lower degree of** rating variability noted for DBR is that a general impression halo effect may have been involved. Given the fact that teachers were asked to observe and rate the behavior of 12 students at once, it is possible that the nature of the rating task contributed to a halo effect.