# Evaluating Rater Bias With Only One Rater Per Target

Megan E. Welsh[1], Sandra M. Chafouleas[1], Gregory A. Fabiano[2],
T. Chris Riley-Tillman[3], & Faith G. Miller[1]

*University of Connecticut[1], University at Buffalo[2], University of Missouri[3]*

## Introduction

**Background**

Test bias is an important validity concern, one that should be addressed in evaluating rating scales. The measure evaluated in this study, Direct Behavior Ratings with Single Item Scales (DBR-SIS) is especially difficult to evaluate for bias because only one rater scores each student, a small number of items are involved, and the format of the measure does not easily lend itself to subjective bias review.

**Objective**

This study applies Cleary 's (1968) test bias framework to evaluate DBR-SIS using multilevel modeling. We regress BESS t-scores on the DBR-SIS Composite and evaluate whether separate lines can be discerned for a wide array of focal groups after controlling for rater effects. We also explore whether teacher characteristics contribute to bias.

## Method

**Participants**

Table 1. *Participant Characteristics*

| | Teachers (n=202) | Students (n=1,976) |
|---|---|---|
| Male | 13.4% | 52.1% |
| SPED student | -- | 13.1% |
| Supplemental supports | -- | 40.3% |
| Minority | 2.5% | 16.9% |
| Hispanic | 1.0% | 7.3% |
| Secondary | 30.7% | 30.0% |
| PBIS | 56.4% | 56.2% |
| Taught 1-5 years | 18.3% | -- |
| Teach >=50% SPED | 5.0% | -- |
| Teach 100% SPED | 2.0% | -- |
| SPED certified | 12.4% | -- |
| Fall BESS M(SD) | -- | 50.3 (10.6) |
| Winter BESS M(SD) | -- | 50.6 (10.6) |
| Spring BESS M(SD) | -- | 50.4 (10.6) |
| Fall DBR-SIS M(SD) | -- | 26.7 (3.7) |
| Winter DBR-SIS M(SD) | -- | 27.0 (3.4) |
| Spring DBR-SIS M(SD) | -- | 27.2 (3.3) |

**Instruments**

*Direct Behavior Rating – Single Item Scale* (DBR-SIS; Chafouleas, Riley-Tillman & Christ, 2009). Teacher rating scale of the proportion of time a student is academically engaged, respectful, or disruptive. Students were rated twice daily for five days. Mean ratings were summed to form a composite ranging from 0 (poor behavior) to 30 (perfect behavior).

*Behavioral and Emotional Screening System* (BESS; Kamphaus & Reynolds, 2007). A brief rating scale that can be useful in screening for behavioral and emotional strengths and weaknesses in children and adolescents.

**Analysis**

The relationship between DBR-SIS Composite scores and BESS t-scores were investigated at each time point using the general model:

**Level-1 Model**

$$Y_{ij} = \beta_{0j} + \beta_{1j}*(DBRComposite_{ij}) + \beta_{2j}*(Focalgroup_{ij}) + r_{ij}$$

**Level-2 Model**

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{01} + \gamma_{11}*(Focalgroup_{ij}) + u_{1j}$$
$$\beta_{2j} = \gamma_{02}$$

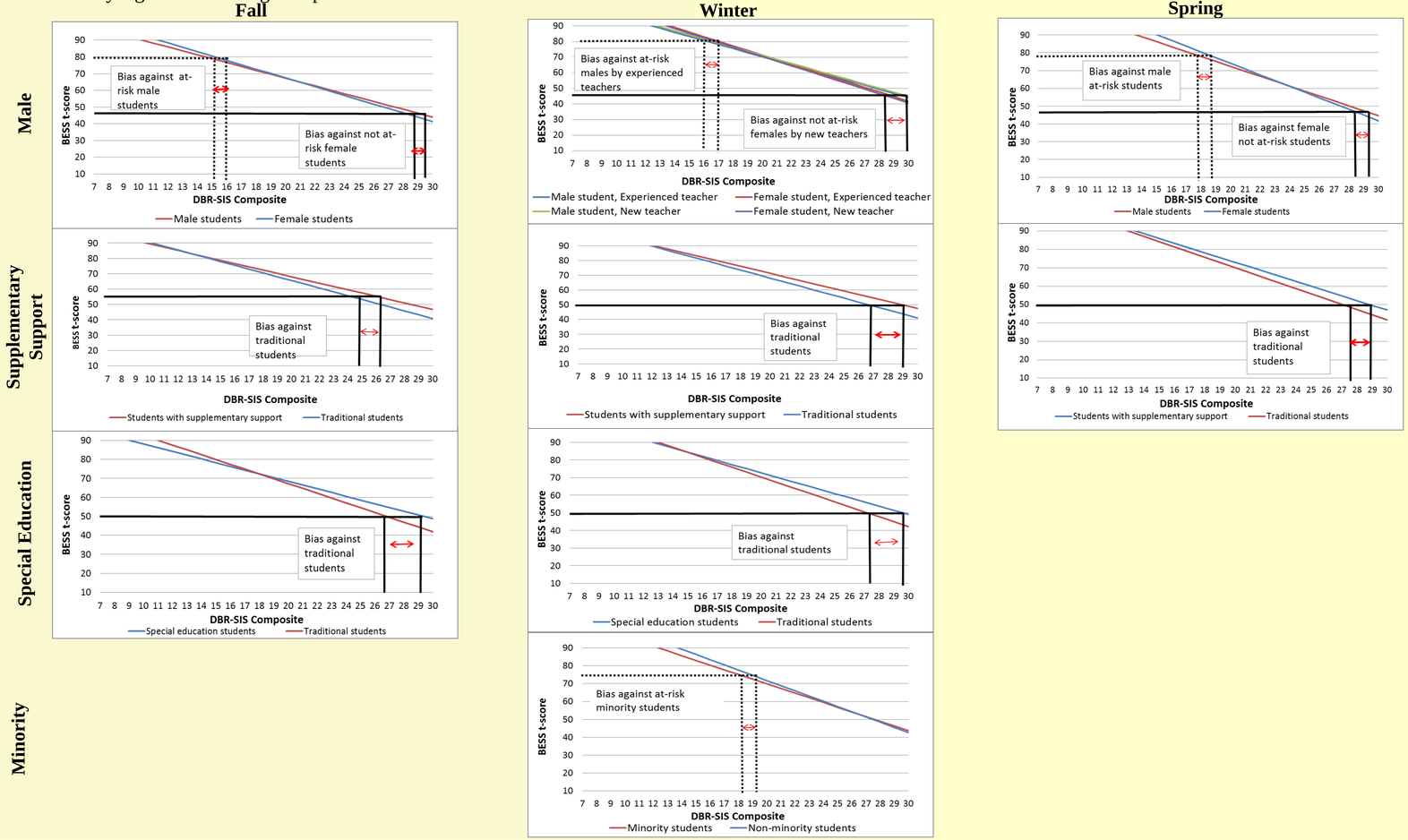
When focal group significantly predicted intercepts or slopes, teacher characteristics that might be associated with bias were added to predict the intercept, the DBR Composite slope, and the focal group slope (e.g., teacher sex predicting sex-related bias; teacher minority status predicting minority-related bias, special education certification, all students disabled, or at least half of students are disabled predicting special education-related bias, teachers with less than six years of experience predicting all forms of bias).

Table 2. *Focal Groups*

| Focal groups examined…. |
|---|
| Males |
| Special education students<br>Supplementary support |
| Minorities |
| Hispanic |
| Secondary student |
| Attend PBIS school |
| Teacher >5 years experience |

## Results

All statistically significant findings are presented below.




## Summary and Conclusions

- This study presents a promising approach to evaluate rating scale test bias when there is only one rater per examinee and scales involve few items.
- We found instances of bias attributable to gender, special education status, and receiving supplementary educational supports at multiple time points and also bias attributable to racial minority status at one time point.
- The direction of bias differed for students at-risk and not at-risk for behavioral difficulty. For example, DBR-SIS scores appear biased against not at-risk girls and also against at-risk boys.
- Finally, after controlling for nesting within rater, we found only one instance in which teacher characteristics helped to explain a finding of bias—new teachers were biased in their ratings of not at-risk girls and experienced teachers were biased against at-risk boys.

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