Using Direct Behavior Ratings in a Middle School Setting

PRESENTERS:

RIVKAH ROSEN, UNIVERSITY OF CONNECTICUT
NICHOLAS CROVELLO, UNIVERSITY OF CONNECTICUT

CONTRIBUTORS:

DR. MEGAN WELSH, UNIVERSITY OF CONNECTICUT
DR. SANDRA CHAFOULEAS, UNIVERSITY OF CONNECTICUT
DR. FAITH MILLER, UNIVERSITY OF CONNECTICUT
DR. CHRIS RILEY-TILLMAN, UNIVERSITY OF MISSOURI
DR. GREG FABIANO, UNIVERSITY AT BUFFALO

NASP ANNUAL CONVENTION 2014
FEBRUARY 19TH, 2014

Table of Contents

- DBR Overview
- An investigation into reliability of direct observations using extant literature
- Method & Materials
- Analyses
  - 1st vs. 2nd Ratings
    - Combined
    - Separate
  - AM vs. PM Ratings
    - Separate
- Implications

Direct Behavior Ratings and Middle School

- All presented analyses are the result of a multi-site collaborative data collection effort at the University of Connecticut, University of Missouri, and University at Buffalo under the federal grant VIABLE-II

Principal Investigators
- Dr. Sandra Chafouleas
- Dr. Chris Riley-Tillman
- Dr. Greg Fabiano
- Dr. Megan Welsh
- Dr. Hariharan Swaminathan

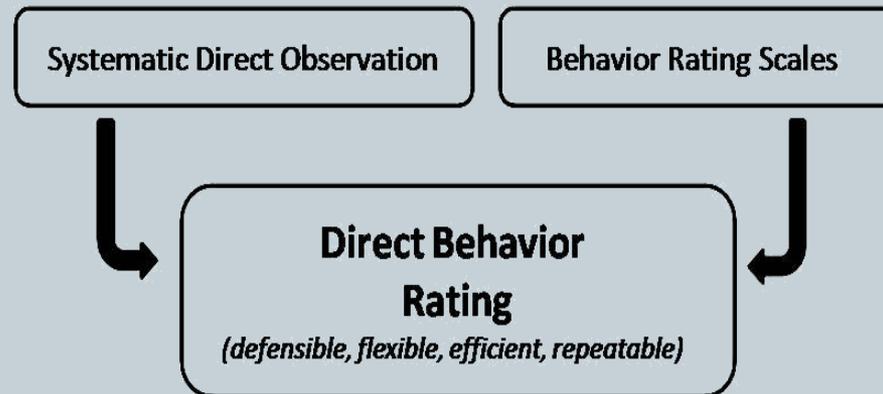# Direct Behavior Ratings

BACKGROUND

Overview of Direct Behavior Ratings

- ## What is DBR?

  - An <u>emerging alternative</u> to systematic direct observation and behavior rating scales which involves *brief rating* of target behavior following a specified observation period

    - For example, a teacher might rate how well students followed directions during a math lesson

Chafouleas, Riley-Tillman, & Christ (2009); Chafouleas, Riley-Tillman, & Sugai (2007); Chafouleas, Riley-Tillman, & McDougal (2002); Christ, Riley-Tillman, & Chafouleas (2009)

DBR Background

Other Names for DBR-like Tools:

- Home-School Note
- Behavior Report Card
- Daily Progress Report
- Good Behavior Note
- Check-In Check-Out Card
- Performance-based behavioral recording

Contemporary Defining Features:

SDO

BRS
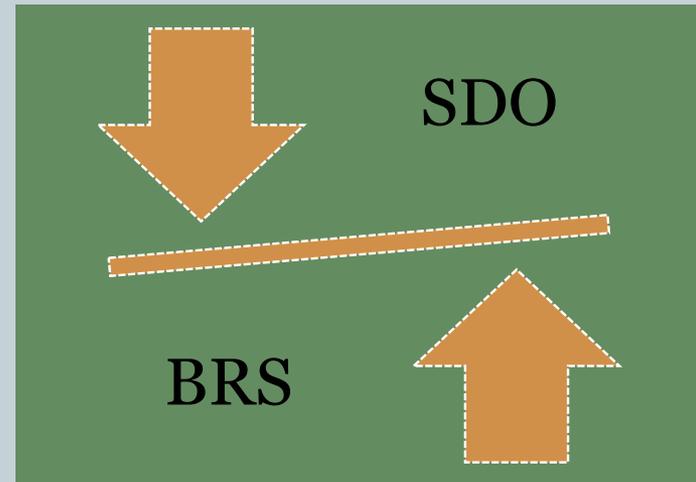
Used repeatedly to represent behavior that occurs over a specified period of time (e.g., 4 weeks) and under specific and similar conditions (e.g., 45 min. morning seat work)

MU UB
UCONN

DBR: Defining Features

- The defining features of DBR are that they are
  - **Direct**
    - Ratings are recorded immediately at the end of an observation
  - **Behavior**
    - Specific behaviors are rated such as Academic Engagement and Disruptive Behavior
  - **Rating**
    - Ratings are conducted repeatedly in a psychometrically sound manner similar to behavior rating scales

DBR Core Behaviors

- Although ANY target can be selected, the following core **behaviors** have been conceptualized as relevant to student success in the classroom:

  - *Academically Engaged*

  - *Respectful*

  - *Non-Disruptive*

Academically Engaged

KEYS TO SUCCESS

Respectful
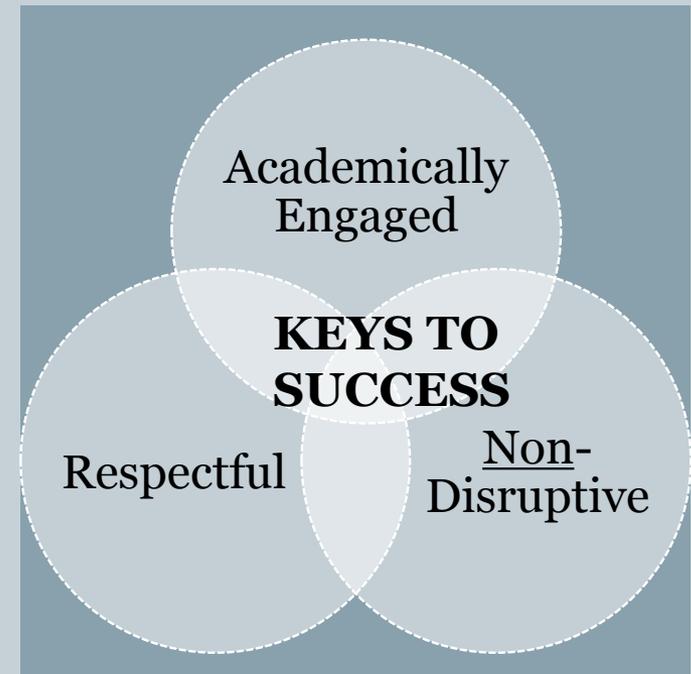
Non-Disruptive

DBR Core Behaviors

Academic Engagement:
Actively or passively participating in the classroom activity.

Respectful:
Compliant and polite behavior in response to adult direction and/or interactions with peers and adults.

Disruptive Behavior:
A student action that interrupts regular school or classroom activity.

DBR Single Item Scale (SIS): What Does it Look Like?

- DBR-SIS is a scale format that has only one target rated per scale. Typically, a single broad behavior (e.g., disruptive) is used to represent a class of behaviors in general (e.g., out of seat, playing with objects)

Academically Engaged

Never | | | | | | | | | | | | Always

0   1   2   3   4   5   6   7   8   9   10
0%   50%   100%
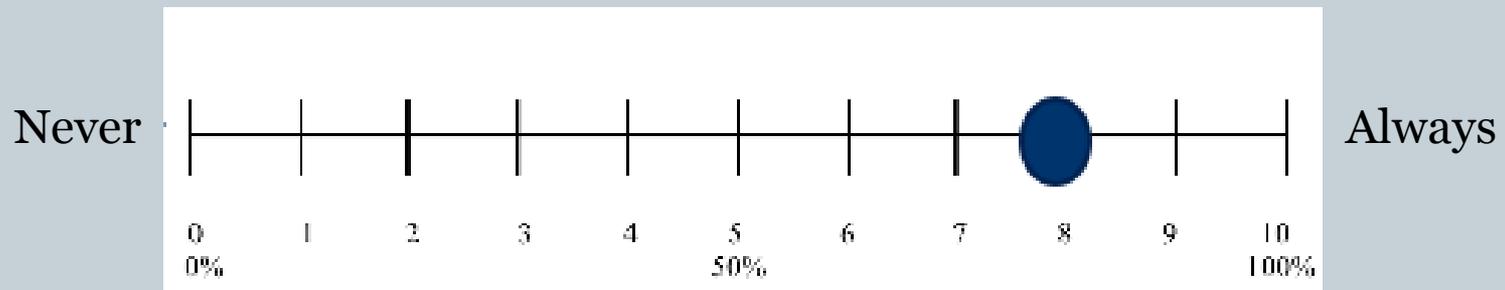
Interpretation: The student displayed academically engaged behavior 80% of large group math instruction today

UCONN

Direct Behavior Ratings are….

- *efficient* as ratings are simple and quick to complete

- *repeatable* for use in progress monitoring assessment

- *defensible* given increasing evidence of technical adequacy for some DBR formats

- *flexible* as it can be used across a range of <u>assessment</u>, <u>intervention</u>, and <u>communication</u> purposes

(Chafouleas, Riley-Tillman, & Christ, 2009)

DBR in Assessment

- In assessment, DBR can be used for the following purposes:

  - Screening
    - Correct identification of students in need

  - Progress monitoring
    - Monitoring of a student's response to an intervention

- When making these decisions, it is critical to have reliable (consistent, repeatable) results

Literature & Background

DIRECT BEHAVIOR RATINGS

An investigation into reliability

- DBR is similar to direct observation in that data are collected in close proximity to the occurrence to the time of behavior (Chafouleas et al., 2010)

- Direct observation has been shown to result in reliable data in middle school (Cushing, Horner, Barrier, 2003)

- Direct observation data have demonstrated high inter-rater reliability in a middle school population (Simonsen, Myers, & Briere, 2010)

An investigation of reliability

- DBR-SIS ratings have demonstrated high levels of reliability in assessing middle school students (Chafouleas, S.M., Kilgus, S.P, Jaffery, R., Riley-Tillman, T.C., Welsh, M.E., & Christ, T.J., 2013).

- DBR-SIS has produced reliable data in evaluating the effects of an intervention for middle school students (Chafouleas, Sanetti, Jaffery, & Fallon, 2011).

An investigation into reliability

- The reliability of data generated by DBR increases as the number of data points increases (Chafouleas, Christ, Riley-Tillman, Briesch, & Chanese, 2007; Christ, Riley-Tillman, Chafouleas, & Boice, 2009)

- Reliability of the data increases when ratings are conducted within the same teacher for each student (Chafouleas et al. 2010)

The Present Study

METHODS & PURPOSE

Purpose

- To determine whether multiple ratings conducted within the same lesson are more reliable than single ratings in a middle school population
  - How reliable are scores based on one rating per day?

- To determine whether there is a difference in the reliability of morning and afternoon ratings in middle school students
  - How reliable are scores in the morning vs. scores in the afternoon?

Methods & Materials

- In the present study, teachers rated 7th & 8th grade students using the DBR method
- Ratings were completed twice per class period
  - Typically less than an hour
- Teachers rated students for five days
- Ratings were completed 3 times during the year
  - Fall, Winter, Spring

Methods & Materials

- Students were rated on the 3 core DBR behaviors
  - Academic Engagement
  - Respectful Behavior
  - Disruptive Behavior

- A composite score was generated by summing across the three core behaviors

- The composite score ranges from 0 – 30 with higher scores reflecting more desirable behavior

Intra-Class Correlation Background

- An intra-class correlation (ICC) can be used to assess the variability:
  - Between participants
  - Within participants

- In our case, the ICC is used for the following:
  - To determine the proportion of variance attributable to differences between students for different times of day (AM/PM ratings) and different ratings per period (1st/2nd ratings)

- ICC was then used to calculate the reliability coefficient for each group using a formula proposed by Shrout & Fleiss (1979)

Analysis

1ST/ 2ND RATINGS

1ˢᵗ/2ⁿᵈ Rating Analysis

- Research Question
- *In a middle school population, is a DBR screening schedule of two ratings per period necessary in terms of identifying risk and procuring accurate scores?*

- Why does this matter?
  - If 1ˢᵗ and 2ⁿᵈ ratings are equally reliable in capturing student behavior, both time and resources can be saved by rating only once per period in a middle school population
  - Middle school teachers only see students for one instructional period per day. This means that the amount of instructional time captured by two middle school ratings may be similar to or less than the amount of time captured by just one rating in elementary grades.

1ˢᵗ/2ⁿᵈ Ratings Analysis

- N = 590 7ᵗʰ and 8ᵗʰ grade students
- Descriptive statistics for DBR composite scores

	Mean	**Standard Deviation**
1ˢᵗ Rating	27.36	4.05
2ⁿᵈ Rating	27.14	4.12

UCONN

1ˢᵗ/2ⁿᵈ Rating Analysis Approach

- Intraclass Correlation & Reliability Coefficients
  - How reliable are the measure's data when estimated using both 1ˢᵗ and 2ⁿᵈ ratings?
  - How reliable are the measure's data when estimated using only 1ˢᵗ ratings?
  - How reliable are the measure's date when estimated using only 2ⁿᵈ ratings?
  - To what extent do the reliability coefficients of the first and second DBR ratings resemble each other?
    - Are they similar enough to only require 1 DBR rating per period?

- Student nested within teacher for all analyses
- Restricted analyses to students who were rated at both first and second rating times
  - Omitted students who were only rated during either first OR second
- Analyses run for combined 1ˢᵗ AND 2ⁿᵈ ratings across seasons
- Analyses run for separate 1ˢᵗ VERSUS 2ⁿᵈ ratings across seasons
- Analyses run to determine variance attributions

1ˢᵗ/2ⁿᵈ Rating Analysis: Combined Approach

- ## Reliability of First AND Second rating combined

  - ### Reliability coefficients separated by season

Season	Reliability Coefficient
Fall	.91
Winter	.93
Spring	.92

  - ### What did we find?

    - Reliability VERY HIGH when looking at the composite score by season
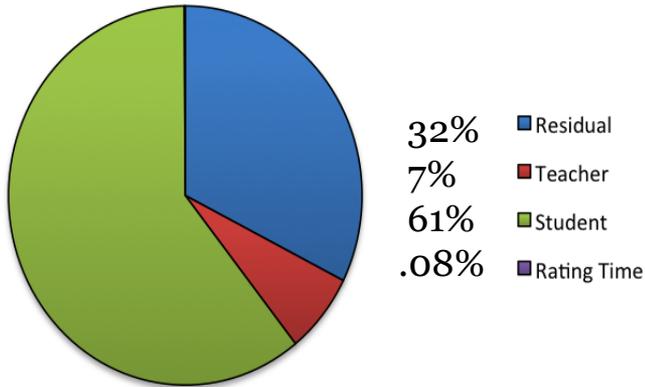
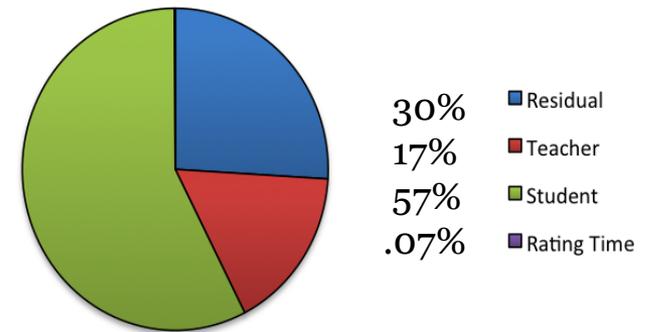# 1ˢᵗ/2ⁿᵈ Rating Analysis: Variance

- Proportion of Variance
  - In a separate analysis, we looked at the proportion of variance attributable to:
    - Student differences
    - Teacher differences
    - Rating time differences

  - Rating time was found to make up a very small proportion of total variance
  - A large proportion of unexplained variance remains, meaning there are still other things affecting ratings that we have not captured

Proportion of DBR Score Attributable to Student, Teacher, and to First or Second Rating

DBR Composite Fall 1st & 2nd Ratings

32% ■ Residual
7% ■ Teacher
61% ■ Student
.08% ■ Rating Time

DBR Composite Winter 1st & 2nd Ratings

30% ■ Residual
17% ■ Teacher
57% ■ Student
.07% ■ Rating Time

DBR Composite Spring 1st & 2nd Rating

29% ■ Residual
8% ■ Teacher
63% ■ Student
2% ■ Rating Time
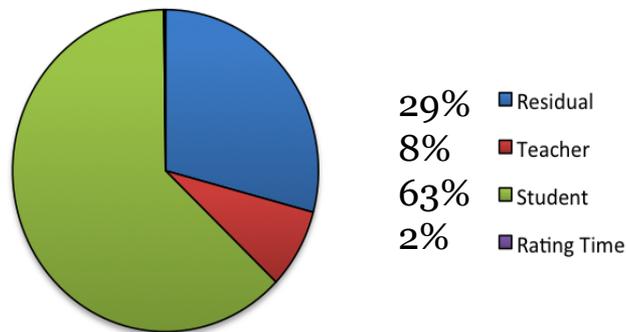
- Since there is so little variability due to rating time (first or second), we can expect first and second rating to be equally reliable...but are they?

  - Brings us to our next analysis
    - 1st VERSUS 2nd ratings

1st/2nd Rating Analysis: Separate Approach

- Reliability of 1st VERSUS 2nd rating when ratings separated into Fall/Winter/Spring time points
  - Things to watch out for
    - When looking at only 1st ratings or only 2nd ratings, we will be capturing different behaviors on different days, so reliability may not be as high
- Reliability coefficients when separated by rating and season:

Season	1st Rating	2nd Rating	Combined Ratings
Fall	.92	.90	.91
Winter	.93	.93	.93
Spring	.92	.91	.92

- 1st and 2nd ratings are equally reliable!
  - Suggests that we should not need to rate middle school students twice per period

Analysis

AM/PM RATINGS

Time of Day Analysis

- Research Question
- *In a middle school population, is there a time of day that yields more reliable DBR ratings in terms of identifying risk and procuring accurate scores?*

- Why does this matter?
  - If AM and PM ratings are equally reliable in capturing student behavior, DBR ratings may be completed at the teacher's convenience, regardless of time of day

Time of Day Analysis Approach

- Intraclass Correlation & Reliability Coefficients
  - How reliable are the measure's data when using only AM ratings?
  - How reliable are the measure's data when using only PM ratings?
  - Are reliability coefficients for AM and PM similar enough to allow ratings to be done at any time of day?

- Student nested within teacher for all analyses
- Restricted analyses to students who had 6 or more DBR ratings
- Analyses run for separate AM versus PM ratings across seasons

Time of Day Analysis Approach

- Reliability of AM *versus* PM rating when ratings separated into Fall/Winter/Spring time points
  - N= 301 middle school students
  - AM defined as classes starting before 11:00 AM
    - 47% of students
  - PM defined as classes starting after 11:00 AM
    - 53% of students
- Descriptive statistics for DBR composite scores

	Mean	Standard Deviation
AM Ratings	27.16	4.25
PM Ratings	27.17	3.69

Reliability of AM/PM Ratings

- Reliability coefficients when separated by time of day and season:

Season	AM Rating	PM Rating
Fall	.92	.90
Winter	.96	.94
Spring	.96	.92

- Equally reliable ratings in the morning AND the afternoon!
  - Suggests that students can be rated at any time of day

Implications

Implications

- We can compare reliability between our two analyses
- 1st vs. 2nd Ratings

Season	1st Rating	2nd Rating
Fall	.92	.90
Winter	.93	.93
Spring	.92	.91

- AM vs. PM Ratings

Season	AM Rating	PM Rating
Fall	.92	.90
Winter	.96	.94
Spring	.96	.92

MU UB
UCONN

Implications

- Based on both analyses presented here, we can conclude:
  - Reliable DBR data in a middle school population can be generated using *one rating at any time of day*

- Requiring only one rating per day at any time allows school personnel to save resources and complete the DBR at their convenience

- Since DBR is already a flexible assessment tool that can be completed quickly and within the natural context of the classroom, these results only improve the efficiency of DBR data collection

Future Directions

- How many data points are necessary when using 1 DBR rating per day in a middle school population?

- What other sources might be affecting the residual variance?
  - Class subject
  - Class type
  - Student demographics

Questions?