Construct Validation of Direct Behavior Ratings: A Multitrait Multimethod Analysis

NASP Annual Convention 2014

**Presenters:**
Dr. Faith Miller, NCSP  
*Research Associate, University of Connecticut*
Daniel Cohen, M.P.H.  
*Research Assistant, University of Missouri*
Wesley Sims, CAGS, NCSP  
*Research Assistant, University of Missouri*

**Contributors:**
Dr. Megan Welsh  
*University of Connecticut*
Dr. Sandra Chafouleas  
*University of Connecticut*
Dr. Chris Riley-Tillman  
*University of Missouri*
Dr. Gregory Fabiano  
*University at Buffalo*
Purpose:

• To discuss the importance of understanding the psychometric properties of assessments
• To review the development of Direct Behavior Ratings – Single Item Scales
• To review results from a multitrait multimethod (MTMM) investigation of DBR
• To discuss implications for practice
NATIONAL ASSOCIATION OF SCHOOL PSYCHOLOGISTS

Model for Services by School Psychologists

PRACTICES THAT PERMEATE ALL ASPECTS OF SERVICE DELIVERY
- Data-Based Decision Making and Accountability
- Consultation and Collaboration

DIRECT AND INDIRECT SERVICES FOR CHILDREN, FAMILIES, AND SCHOOLS

<table>
<thead>
<tr>
<th>Student-Level Services</th>
<th>Systems-Level Services</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interventions and Instructional Support to Develop Academic Skills</td>
<td>School-Wide Practices to Promote Learning</td>
</tr>
<tr>
<td>Interventions and Mental Health Services to Develop Social and Life Skills</td>
<td>Preventive and Responsive Services</td>
</tr>
<tr>
<td>Family-School Collaboration Services</td>
<td></td>
</tr>
</tbody>
</table>

FOUNDATIONS OF SERVICE DELIVERY
- Diversity in Development and Learning
- Research and Program Evaluation
- Legal, Ethical, and Professional Practice

HELPING STUDENTS AND SCHOOLS ACHIEVE THEIR BEST
The importance of the assessment process:

• We need reliable and valid data in order to support students
• Nearly all of our decisions depend on it
• Understanding the strengths and limitations of our assessments is essential
• Different assessments provide us with different information...
Purpose of Assessment

- **Screening**
  - Who needs help?

- **Diagnosis**
  - Why is the problem occurring?

- **Progress Monitoring**
  - Is intervention working?

- **Evaluation**
  - How well are we doing overall?

Within each category, we can assess different traits using different methods: **what** are we measuring and **how** are we measuring it?

Emphasized within a Multi-Tiered Service Delivery Framework (RTI)
Behavioral Assessment

Rating Scales
- Teacher Report
- Parental Report
- Student Report

Observations
- Event recording
- Time sampling

Inattention

Social Skills

Disruptive Behavior

Interviews
- Unstructured
- Semi-structured
- Structured

Hyperactivity

Motivation

Internalizing Problems

Extant Data
- ODRs
- Attendance

Evaluation

Screening

Progress Monitoring

Diagnosis
School-based behavior assessment within RTI

- Current methods of behavior assessment were not built for multi-tiered assessment
- New options must possess four desirable characteristics...

Defensible  Efficient  Flexible  Repeatable

(Chafouleas, 2011; Chafouleas, Christ, & Riley-Tillman, 2009; Chafouleas, Volpe, Gresham, & Cook, 2010)
Direct Behavior Rating
What is DBR?

- An emerging alternative to systematic direct observation and behavior rating scales which involves brief ratings of target behaviors following a specified observation period.
A little 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:**

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)
**Example Scale Formats for DBR**

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

**Single Item Scale**

Interpretation: The student displayed academically engaged behavior during 80% of the observation period.

**Multi-Item Scale**

<table>
<thead>
<tr>
<th></th>
<th>Never</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did the student follow class rules?</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Did the student follow teacher directions?</td>
<td>0</td>
<td>1 2</td>
</tr>
<tr>
<td>Did the student do his/her best work?</td>
<td>0</td>
<td>1 2</td>
</tr>
</tbody>
</table>

Total number of points earned: **5**

Interpretation: The student earned 84% (5/6) of possible points during the observation period.
DBR-SIS

Directions: Place a mark along the line that best reflects the percentage of total time the student exhibited each target behavior. Note that the percentages do not need to total 100% across behaviors since some behaviors may co-occur.

* Remember that a lower score for “Disruptive” is more desirable.
DBR-SIS Target 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.
Development & Validation of DBR-SIS
RESEARCH: Project VIABLE (2006-2011) and Project VIABLE II (2011-current)

**Develop** instrumentation and procedures, then **evaluate** defensibility of DBR in decision-making

- **Rating Procedures**
- **Behavior Targets**
- **Scale Design**
- **Method Comparisons**
- **Rater Training**

**Evaluate** defensibility and usability of DBR in decision-making at larger scale

- **Triannual behavioral screening**
- **Multi-trait multi-method investigation**
- **Single-case design studies using DBR**
- **Teacher input regarding usability and perceptions**

Funding provided by the **Institute for Education Sciences**, U.S. Department of Education
## Development & Validation

### Development & Validation of DBR-SIS

<table>
<thead>
<tr>
<th>Scale development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior wording</td>
</tr>
<tr>
<td>Training</td>
</tr>
<tr>
<td>Influence of observation duration</td>
</tr>
<tr>
<td>How teachers assign ratings</td>
</tr>
<tr>
<td>Perceptions of usability</td>
</tr>
</tbody>
</table>

### Applications in Screening

- Developing cut scores to identify students at-risk
- Concurrent validity with established screeners: SRSS, BESS
- Examining bias

### Applications in Progress Monitoring

- Determining scale sensitivity to change
- Concurrent validity with SDO
Questions Remain...

- Foundational psychometric evidence of DBR-SIS
  - Reliability evidence
    - Accuracy or precision of scores
  - Validity evidence
    - The extent to which it is appropriate to use DBR-SIS for screening and progress monitoring
    - Many different types of validity evidence
    - Here, we focus on construct validity
Multitrait Multimethod Analysis
Rationale

• Test developers must accurately define, measure, and rigorously validate the construct(s) of interest
• Campbell and Fiske (1959) developed an approach to assessing construct validity
  ▫ MTMM analysis permits the examination of:
    • Convergent validity - evidence that scores are consistent with other measures of the same trait
    • Discriminant validity – evidence that scores diverge from measures of similar, but distinct traits
• Examining both convergent and discriminant evidence contributes to validity argument by determining not only whether a measure is consistent with criterion measures of the same construct, but also whether the measure is less strongly associated with measures of different, but related constructs
Purpose of MTMM Analysis

- Provides a way to systematically evaluate the correlations among a set of measures
  - Correlations tell us the degree of association between variables
- Evaluate construct validity
  - Convergent validity
  - Discriminant validity
- Evaluate variance attributed to traits vs. methods
Example MTMM Matrix

What are we looking for?

- High reliability coefficients
- Correlations between measures of the same trait obtained using different methods should be large
- Correlations between measures of the same trait obtained through different methods should be stronger than those observed between different traits using the same method
- The same pattern of trait correlations should hold for all methods and all combinations of methods

*K. Widaman (2010)*
Primary Research Questions

• How are scores obtained from DBR-SIS associated with other measures of school-based behavior?
  ▫ Evidence for convergent validity?
  ▫ Evidence for discriminant validity?

• Do there appear to be strong methods factors associated with various measures of behavior?
Methods

- Participants and Setting:
  - 993 students
  - 122 teachers
- Public school settings were located in 4 states: Connecticut, Rhode Island, New York, and Missouri
- Students were enrolled in a total of 19 different schools, including rural, suburban, and urban districts
- Participating students were in grades 3-8
<table>
<thead>
<tr>
<th>Student characteristic</th>
<th>$n$</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>452</td>
<td>45</td>
</tr>
<tr>
<td>Female</td>
<td>541</td>
<td>55</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>780</td>
<td>79</td>
</tr>
<tr>
<td>African American</td>
<td>154</td>
<td>16</td>
</tr>
<tr>
<td>Asian</td>
<td>35</td>
<td>4</td>
</tr>
<tr>
<td>Other</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>89</td>
<td>9</td>
</tr>
<tr>
<td>Non-Hispanic</td>
<td>904</td>
<td>91</td>
</tr>
<tr>
<td><strong>Grade</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Third</td>
<td>210</td>
<td>21</td>
</tr>
<tr>
<td>Fourth</td>
<td>204</td>
<td>21</td>
</tr>
<tr>
<td>Fifth</td>
<td>206</td>
<td>21</td>
</tr>
<tr>
<td>Sixth</td>
<td>166</td>
<td>17</td>
</tr>
<tr>
<td>Seventh</td>
<td>124</td>
<td>12</td>
</tr>
<tr>
<td>Eighth</td>
<td>83</td>
<td>8</td>
</tr>
</tbody>
</table>
Methods: Measures

- DBR-SIS teacher ratings: AE, DB, RS
- DBR-SIS student ratings: AE, DB, RS
- SDO observations: AE, DB, RS
  - Momentary time sampling, 10 second intervals
- Teacher rating scales
  - Attention Problems Subscale (BASC-2)
  - Hyperactivity Subscale (BASC-2)
  - Communication Subscale (SSIS Rating Scale)
- Student self-report rating scales
  - Attention Problems Subscale (BASC-2)
  - Hyperactivity Subscale (BASC-2)
  - Communication Subscale (SSIS)
Methods: Procedures

- Data collection occurred in a single assessment period in winter/spring of 2013
- Up to 10 students could participate per classroom
- Teachers and students were asked to complete:
  a) DBR-SIS scales over 10 occasions (one week)
  b) Behavior rating scales matched to the target constructs
- External observers completed SDO observations
  - Goal: 3+ 15 minute observations
  - IOA observations were also conducted
- Assessment order was counterbalanced in order to control for potential order effects
Results

- 3 (trait) x 5 (method) matrix
- Reliability coefficients were calculated as follows:
  - **DBR-SIS Teacher**: derived from intraclass correlation coefficient (ICC)
  - **DBR-SIS Student**: derived from intraclass correlation coefficient (ICC)
  - **SDO**: Pearson’s product moment correlations (inter-rater reliability)
  - Teacher rating scales: internal consistency ($\alpha$)
  - Student rating scales: internal consistency ($\alpha$)
<table>
<thead>
<tr>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
<th>Method 4</th>
<th>Method 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>1. DBR – Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Academic Engagement</td>
<td>.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Disruptive Behavior</td>
<td>-.87</td>
<td>.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. Respectful</td>
<td>.81</td>
<td>-.91</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td>2. DBR- Student</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Academic Engagement</td>
<td>.49</td>
<td>-.41</td>
<td>.41</td>
<td>.82</td>
</tr>
<tr>
<td>b. Disruptive Behavior</td>
<td>-.45</td>
<td>.48</td>
<td>-.44</td>
<td>-.75</td>
</tr>
<tr>
<td>c. Respectful</td>
<td>.45</td>
<td>-.47</td>
<td>.47</td>
<td>.96</td>
</tr>
<tr>
<td>3. SDO</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Academic Engagement</td>
<td>.37</td>
<td>-.39</td>
<td>.33</td>
<td>.27</td>
</tr>
<tr>
<td>b. Disruptive Behavior</td>
<td>-.29</td>
<td>.35</td>
<td>-.30</td>
<td>-.23</td>
</tr>
<tr>
<td>c. Respectful</td>
<td>.21</td>
<td>-.28</td>
<td>.30</td>
<td>.16</td>
</tr>
<tr>
<td>4. Rating Scale – Teacher</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Academic Engagement¹</td>
<td>-.75</td>
<td>.63</td>
<td>-.55</td>
<td>-.39</td>
</tr>
<tr>
<td>b. Disruptive Behavior²</td>
<td>-.58</td>
<td>.71</td>
<td>-.65</td>
<td>-.35</td>
</tr>
<tr>
<td>c. Respectful³</td>
<td>.55</td>
<td>-.50</td>
<td>.48</td>
<td>.33</td>
</tr>
<tr>
<td>5. Rating Scale - Student</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Academic Engagement¹</td>
<td>-.47</td>
<td>.41</td>
<td>-.34</td>
<td>-.53</td>
</tr>
<tr>
<td>b. Disruptive Behavior²</td>
<td>-.34</td>
<td>.39</td>
<td>-.32</td>
<td>-.38</td>
</tr>
<tr>
<td>c. Respectful³</td>
<td>.14</td>
<td>-.16</td>
<td>.11</td>
<td>.30</td>
</tr>
</tbody>
</table>

Rules of thumb for interpreting correlations:
-.20 = Weak
.20-.69 = Moderate
>.69 = Strong

Note. ¹ BASC-2 Attention Problems Subscale, ² BASC-2 Hyperactivity Subscale, ³ SSIS-RS Communication Subscale
Results

- Reliability coefficients were highest for the teacher rating scales, and lowest for the student rating scales
  - Reliability coefficients across methods were generally high
- Validity diagonals provide information on convergent validity
  - Coefficients were variable
  - Higher for AE & DB (Moderate to Strong)
  - Lower for RS (Weak to Moderate)
- Analysis of heterotrait-monomethod triangles suggests method effects
  - Same method, different traits, strong correlations
- Validity coefficients were often similar in magnitude to those in the heterotrait-heteromethod triangles
  - Are traits distinct? Does the method effect overpower the trait effect?
Primary Research Questions

• How are scores obtained from DBR-SIS associated with other measures of school-based behavior?
  ▫ Evidence for convergent validity?
    • Yes: Teacher DBR and Teacher Rating Scale
    • No: Student Rating Scale and SDO, Student DBR
  ▫ Evidence for discriminant validity?
    • Limited evidence

• Do there appear to be strong methods factors associated with various measures of behavior?
  ▫ Yes, method seems to matter
Next steps

• Structural Equation Modeling
  ▫ Account for nesting of students within teachers
  ▫ Estimate trait and method related variance
  ▫ Test the amount of trait-related and method-related variance statistically
Discussion

• **Implications for practice**
  ▫ **What are the implications of these findings on assessment selection?**
    • Our methods impact our results
  ▫ As school psychologists, should we be surprised when we find varied results using different assessment methods?
  ▫ Do you think these measurement challenges are unique to behavioral assessment?
Questions & Comments

Website: www.directbehaviorratings.org
Contact: faith.miller@uconn.edu