Direct Behavior Rating (DBR) sensitivity to change: Outcomes across consultation cases

Sandra M. Chafouleas, Ph.D.
Lisa M. Hagermoser Sanetti, Ph.D.
Stephen P. Kilgus, M.A.
Daniel M. Maggin, Ph.D.
University of Connecticut
Center for Behavioral Education and Research

Presentation at the 2010 Conference of the National Association of School Psychologists
Purpose:

- To review the logic and process of behavioral consultation
- To introduce Direct Behavior Rating (DBR) as an assessment method for progress monitoring of student behavior
- To review options for evaluating student behavioral response to intervention
- To demonstrate how DBR can be used to evaluate outcomes from consultation cases
REVIEW: Why do we need data?

**Purposes of Assessment**
- Screening
- Progress Monitoring
- Diagnosis
- Evaluation

Emphasized within a problem-solving framework
Foundations for Problem-Solving Models

**Problem Solving Model**
- Define the Problem
- Develop a Plan
- Implement Plan
- Evaluate

**“Train and Hope” Model**
- REACT to Problem Behavior
- WAIT for New Problem
- Select & ADD Practice
- Expect, But HOPE for Implementation
- Hire EXPERT to Train Practice
**What is “response to intervention”?”?**

**BASIC QUESTION:** How do we know if X is working?

- Foundations within *data-based decision making*
- Roots of data-based decision making come from the *problem-solving model*
- *Process* involved in “problem-solving” is ancient
  - model became clearly articulated within psychology and then education through applied behavior analysis — behavioral consultation

What is the problem?
What is it occurring?
What should we do about it?
Did it work?

(Bergan, 1977, Bergan & Kratochwill, 1990; Tilly, 2009; Reschly & Bergstrom, 2009)
Definitions: desirable features of assessment tools within PSM

- **Defensible**
  - established through psychometric research to provide evidence of reliability and validity for interpretation and use

- **Flexible**
  - established by methods useful in guiding a variety of assessment questions and situations

- **Efficient**
  - established by methods that require relatively few resources (feasible and reasonable)

- **Repeatable**
  - established by methods that yield necessary time series to evaluate intervention effectiveness

Adapted from Briesch & Volpe (2007)
BUT for behavior... it is complicated!

- Absence of a gold standard criterion
- One measure can’t do it all
  - Multiple measures are needed to evaluate different facets
- Co-morbidity of “problems”
  - What are the most relevant problem features?
- Multiple perspectives are valuable yet agreement may (will) be low!
- Moderators matter…

(Adapted from Kazdin, 2005)
What are the possibilities?

**Possible Methods?** Systematic direct observation, behavior rating scales, permanent products, Direct Behavior Rating

**Possible Metrics?** Visual analysis, reliable changes in behavior (RCI, percent change from baseline, PND, effect size), social validation, changes on social impact measures (e.g. dropout)
Methods

- Direct observations are costly
- Universally-accepted GOM for social behavior does not exist
- Traditional behavior rating scales not sensitive to change, not contextually relevant
- Permanent products lack defensibility

Metrics

- PND Does not index strength of response
- Decision rules for judging RTI not established
- Visual analysis does not allow “quantification”
- There are no social behavior “benchmarks”
- Effect sizes are often uninterpretable in SSD
DIRECT BEHAVIOR RATING: What is DBR?

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

Chafouleas, Riley-Tillman, & Christ (2009); Chafouleas, Riley-Tillman, & Sugai (2007); Chafouleas, Riley-Tillman, & McDougal (2002); Christ, Riley-Tillman, & Chafouleas (2009)
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:
- 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)
Example DBR scales

**Single Item Scale**

% of Total Time

0% 1% 2% 3% 4% 5% 6% 7% 8% 9% 10%

0% 50% 100%

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

**Multi-Item Scale**

<table>
<thead>
<tr>
<th>Did the student follow class rules?</th>
<th>Never</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Did the student follow teacher directions?</th>
<th>Never</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Did the student do his/her best work?</th>
<th>Never</th>
<th>Always</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>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.
Project VIABLE: 
Validation of Instruments for Assessing Behavior Longitudinally & Efficiently

GOAL: Develop and Evaluate DBR

Phases I & II: Develop instrumentation and procedures; evaluate defensibility of DBR in decision-making
• Large datasets; repeated observations of student behavior
• Understanding critical factors (e.g. scale, behavior targets)
• Pilot testing various aspects with classroom teachers

Phase III: evaluate feasibility and utility of DBR in school settings.
• Packaging what we have learned to then train teachers
• Establish groups of teachers willing to participate in DBR training
• Implement the training and provide feedback to researchers

Sandra M. Chafouleas
T. Chris Riley-Tillman
Theodore J. Christ
George Sugai

Funding provided by the Institute for Education Sciences, U.S. Department of Education (R324B060014).
Ratings should correspond to the **percentage of time** that the student was observed to display the target behavior.

Ex: When rating after 40-minute Independent Reading Block, if the student was engaged for 20 minutes, then the student receives a rating of 5 on the DBR.
DBR – Single Item Scales (DBR-SIS)

- Academically Engaged
- Respectful
- Non-Disruptive

**KEYS TO SUCCESS**

Academically Engaged
Respectful
Non-Disruptive
## Direct Behavior Rating (DBR) Form: 3 Standard Behaviors

<table>
<thead>
<tr>
<th>Date:</th>
<th>Student:</th>
<th>Activity Description:</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>T</td>
<td>W</td>
</tr>
<tr>
<td>T</td>
<td>W</td>
<td>Th</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Rater: 

#### Behavior Descriptions:

- **Academically engaged** is actively or passively participating in the classroom activity. For example: writing, raising hand, answering a question, talking about a lesson, listening to the teacher, reading silently, or looking at instructional materials.

- **Respectful** is compliant and polite behavior in response to classroom rules, adult directions, and/or peer interactions. For example: follows teacher direction, pro-social interaction with peers, positive response to adult request, conformity to classroom rules and norms.

- **Disruptive** is student action that interrupts regular school or classroom activity. For example: out of seat, fidgeting, playing with objects, acting aggressively, talking/yelling about things that are unrelated to classroom instruction.

### 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.

#### Academically Engaged

<table>
<thead>
<tr>
<th>% of Total Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
</tr>
<tr>
<td>Never</td>
</tr>
</tbody>
</table>

#### Respectful

<table>
<thead>
<tr>
<th>% of Total Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
</tr>
<tr>
<td>Never</td>
</tr>
</tbody>
</table>

#### Disruptive *

<table>
<thead>
<tr>
<th>% of Total Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
</tr>
<tr>
<td>Never</td>
</tr>
</tbody>
</table>

* Remember that a lower score for “Disruptive” is more desirable.
Summary: Characteristics of DBR-SIS

- Repeatable
- Efficient
- Flexible
- Defensible

Psychometric comparisons at single point

Evaluating sensitivity to change
Are DBR single item scales (SIS) sensitive to behavioral change?

Collaborative research project between

Dr. Lisa Sanetti & Dr. Sandy Chafouleas

with a school psych. consultant team involving Steve Kilgus, Katie Gritter, Rose Jaffery, Lindsay Beck, Lisa Dobey, Teri LeBel

& special guest appearances by Dr. Dan Maggin
- Participants included 20 teacher-student dyads
- Dyadic data was included if the teacher had completed DBR across 4 baseline and 10 intervention days.

<table>
<thead>
<tr>
<th>Activity</th>
<th>Number of Students</th>
<th>Number of Datapoints</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Baseline</td>
<td>Intervention</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>Range</td>
<td>M</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>6.25</td>
<td>4-12</td>
<td>17.40</td>
</tr>
<tr>
<td>2</td>
<td>19</td>
<td>6.32</td>
<td>4-11</td>
<td>17.63</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>6.17</td>
<td>4-11</td>
<td>16.78</td>
</tr>
</tbody>
</table>
Participants cont’d: Teachers

Teacher Characteristics

- **Teacher age**: Mean 34.8 years, Range 24 - 56 years
- **Years teaching**: Mean 9.1 years, Range 1 - 35 years

**Gender**
- Female 95%
- Male 5%

**Teacher Race/Ethnicity**
- Caucasian, 15
- African-American, 1
- Hispanic, 2
- Not Reported, 2
- African-American, 1

**Teacher Race/Ethnicity**

- Caucasian, 15
- African-American, 1
- Hispanic, 2
- Not Reported, 2
- African-American, 1
Participants cont’d: Students

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Mean</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>7.82 yrs</td>
<td>5-11 yrs</td>
</tr>
<tr>
<td>Gender</td>
<td>Male, 1</td>
<td>Female, 1</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td>Caucasian, 15</td>
<td>Hispanic, 2</td>
</tr>
<tr>
<td>Academic problems</td>
<td>Yes 10</td>
<td>No 9</td>
</tr>
<tr>
<td>Behavior problems</td>
<td>Yes 20</td>
<td>No 0</td>
</tr>
<tr>
<td>Special Education</td>
<td>Yes 4</td>
<td>No 16</td>
</tr>
</tbody>
</table>
Intervention Rating Profile-Adapted

Academic Subject:

DBR Form

1. This would be an acceptable intervention for the child’s problem behavior.
   - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

2. Most teachers would find this intervention appropriate for behavior problems in addition to the one described.
   - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

3. This intervention should prove effective in changing the child’s problem behavior.
   - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

4. I would suggest the use of this intervention to other teachers.
   - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

5. The child’s problem behavior is severe enough to warrant use of this intervention.
   - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

6. Most teachers would find this intervention suitable for the behavior problem described.
   - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

7. I would be willing to use this intervention in the classroom setting.
   - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

8. This intervention would not result in negative side-effects for the child.
   - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

9. This intervention would be appropriate for a variety of children.
   - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

10. This intervention is consistent with those I have used in classroom settings.
    - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

11. The intervention was a fair way to handle the child’s problem behavior.
    - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

12. This intervention is reasonable for the behavior problem described.
    - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

13. I like the procedures used in this intervention.
    - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

14. This intervention was a good way to handle this child’s behavior problem.
    - Strongly disagree: 1, Disagree: 2, Slightly disagree: 3, Agree: 4, Slightly agree: 5, Strongly agree: 6

Comments:
A series of consultative interviews were conducted to establish:

- Which teacher-nominated students may benefit from use of the DRC intervention
- The three activities within which the target student’s behavior was most problematic
- Which 3-5 behaviors should be targeted for intervention:
  - Did I follow class rules?
  - Did I follow teacher directions?
  - Did I do my best work?
  - 2 optional student-specific behaviors
- A menu of titrated rewards the student may earn and choose from each day if enough
Procedure cont.

- **Baseline**
  - Teachers completed DBR after each of the three specified activities each day for at least 5 days.
  - Consultants conducted the BOSS 2-3 times

- **Intervention**
  - At the end of each activity:
    - Teachers reviewed the DRC with student
    - Teachers were to complete the DBR immediately after DRC review
  - At end of school day/last activity:
    - Teachers reviewed the completed DRC with the student
    - Students could earn one of three levels of rewards depending on the number of “Yeses” they received.
  - During the 4th school week or in the last 5 days of intervention:
    - Consultants conducted the BOSS 2-3 times
Data Analysis

- Change Metrics (Gresham, 2005)
  - Absolute Change
  - Percent of nonoverlapping data
  - Percentage of change
  - Effect size
  - Reliable change index

- Spearman’s rho correlations
  - BOSS & DBR-SIS absolute change scores
  - DBR-SIS change metrics

- Pearson’s product-moment correlations
  - BOSS & DBR-SIS (collapsed across phases & activities)
  - DBR-SIS metrics & IRP-A
Change Metrics

- **Absolute change**
  - Intervention mean – Baseline mean

- **Percent of nonoverlapping data (PND)**
  - Number of intervention data points that exceeded the highest baseline data point (or fell below the lowest data point for DB), divided by the total number of intervention data points

- **Percentage of change**
  - (Intervention mean – Baseline mean)/Baseline mean

- **Effect size**
  - (Intervention mean – Baseline mean)/Baseline SD

- **Reliable change index (RCI)**
  - (Intervention mean – Baseline mean)/ \( SE_{diff} \)
### Results

#### Descriptive statistics across scales and phases

<table>
<thead>
<tr>
<th>Scale</th>
<th>Phase</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Disruptive Behavior</strong></td>
<td>Baseline</td>
<td>4.26</td>
<td>1.97</td>
<td>0.36 - 7.83</td>
</tr>
<tr>
<td></td>
<td>Intervention</td>
<td>2.58</td>
<td>1.41</td>
<td>0.00 - 5.55</td>
</tr>
<tr>
<td><strong>Academic Engagement</strong></td>
<td>Baseline</td>
<td>4.97</td>
<td>2.28</td>
<td>0.63 - 9.14</td>
</tr>
<tr>
<td></td>
<td>Intervention</td>
<td>6.82</td>
<td>1.50</td>
<td>1.90 - 9.65</td>
</tr>
<tr>
<td><strong>Compliance</strong></td>
<td>Baseline</td>
<td>5.74</td>
<td>1.93</td>
<td>2.25 - 9.25</td>
</tr>
<tr>
<td></td>
<td>Intervention</td>
<td>7.34</td>
<td>1.31</td>
<td>4.53 - 10.00</td>
</tr>
<tr>
<td><strong>On-task</strong></td>
<td>Baseline</td>
<td>69.98</td>
<td>19.76</td>
<td>14.00 - 98.00</td>
</tr>
<tr>
<td></td>
<td>Intervention</td>
<td>81.94</td>
<td>14.22</td>
<td>46.00 - 100.00</td>
</tr>
<tr>
<td><strong>Off-task</strong></td>
<td>Baseline</td>
<td>44.82</td>
<td>21.01</td>
<td>4.00 - 94.00</td>
</tr>
<tr>
<td></td>
<td>Intervention</td>
<td>28.69</td>
<td>18.54</td>
<td>2.00 - 77.00</td>
</tr>
</tbody>
</table>

1 – DBR-SIS ratings correspond to percentages (e.g., a DBR-SIS rating of 1 corresponds to 10%).

2 – Results are in the form of percentages.
### Results

**Descriptive statistics for change metrics across DBR-SIS**

<table>
<thead>
<tr>
<th>Change Metric</th>
<th>Disruptive Behavior</th>
<th>Academic Engagement</th>
<th>Compliance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>Range</td>
</tr>
<tr>
<td>Absolute Change</td>
<td>-1.68</td>
<td>1.80</td>
<td>-6.83 - 2.72</td>
</tr>
<tr>
<td>Percent Change</td>
<td>-0.32</td>
<td>0.49</td>
<td>-1.00 - 1.56</td>
</tr>
<tr>
<td>PND</td>
<td>0.30</td>
<td>0.29</td>
<td>0.00 - 0.95</td>
</tr>
<tr>
<td>Effect Size</td>
<td>-0.82</td>
<td>1.02</td>
<td>-3.56 - 1.32</td>
</tr>
<tr>
<td>RCI</td>
<td>-1.33</td>
<td>1.66</td>
<td>-5.77 - 2.15</td>
</tr>
</tbody>
</table>
Results

*Spearman’s rho correlations between DBR-SIS and BOSS absolute change metrics*

<table>
<thead>
<tr>
<th>BOSS Scale</th>
<th>DBR-SIS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Disruptive Behavior</td>
</tr>
<tr>
<td>On-task</td>
<td>-.458</td>
</tr>
<tr>
<td>Off-task</td>
<td>.487*</td>
</tr>
</tbody>
</table>

* - Statistically significant at the .05 level
Results cont.

- **Spearman’s rho correlations amongst DBR-SIS change metrics**
  - Analyses were kept within SIS. For example, disruptive behavior change metrics were only compared to other disruptive behavior change metrics.
  - Results revealed strong, statistically significant correspondences (at the $p = .001$ level) between each of DBR-SIS change metrics.
  - A sole exception was the correlation between the percent change and PND metrics for the disruptive behavior DBR-SIS, which did not reach statistical significance ($\rho = -.21, p = .118$)

- **Pearson’s product-moment correlations between SDO and DBR.**
  - SDO-AE & DBR-AE $\Rightarrow r = .344, p = .001$
  - SDO-OT & DBR-DB $\Rightarrow r = .292, p = .007$
## Results

### Correlations between DBR-SIS change metrics and average IRP-A score

<table>
<thead>
<tr>
<th>DBR-SIS</th>
<th>Absolute Change</th>
<th>Percentage of Change</th>
<th>PND</th>
<th>Effect Size</th>
<th>RCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disruptive Behavior</td>
<td>-.05</td>
<td>-.03</td>
<td>.04</td>
<td>-.06</td>
<td>-.06</td>
</tr>
<tr>
<td>Academic Engagement</td>
<td>.13</td>
<td>.03</td>
<td>.08</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td>Compliance</td>
<td>*.29</td>
<td>.21</td>
<td>.21</td>
<td>*.31</td>
<td>*.31</td>
</tr>
</tbody>
</table>

* - Statistically significant at the .05 level
Discussion

- DBR-SIS and BOSS descriptive data indicate change in student behavior across phases, in the expected direction.

- High correspondence between DBR-SIS and BOSS absolute change metrics suggests that students were ranked similarly across the two measures with regard to responsiveness to the DRC intervention.
  - Provides preliminary support for further research into the use of DBR-SIS to differentiate between those who have or have not responded to intervention.

- High correlations amongst DBR-SIS change metrics indicates that each affords similar information.
  - Yet, conceptual limitations of some metrics may hinder their usefulness.
Discussion

- Low (yet statistically significant) association b/w DBR & SDO
  - Suggests similarity across the methods with regard to summative evaluations?

- Small/non-existent association between teacher perceptions (acceptability/effectiveness) and student RTI
  - Consistent with literature indicating teachers don’t have to like an intervention for it to work (e.g. Axelrod, 1996)?
Discussion

- **Absolute change** may be good, but level of change indicative of “adequate response” is not consistent across DBR scale.

- Due to floor and ceiling effects, **PND** is not a good indicator.

- **Percentage of change** was not as interpretable as others have found (e.g., Cheney et al., 2008). However, should DBR cut scores be established, may become more useful.

- **Effect size** may be a good indicator given the ability to interpret magnitude of effect. However, challenges with interpretation are apparent, but may be good for low stakes decisions.

- In accordance with past recommendations and findings (Cheney et al., 2008; Gresham, 2005), **RCI** seems a bit too stringent of a criteria. However, may be suitable for higher stakes decisions.
Questions/Comments...

Contact:
Dr. Sandra Chafouleas – sandra.chafouleas@uconn.edu
Dr. Lisa Sanetti – lisa.sanetti@uconn.edu