

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

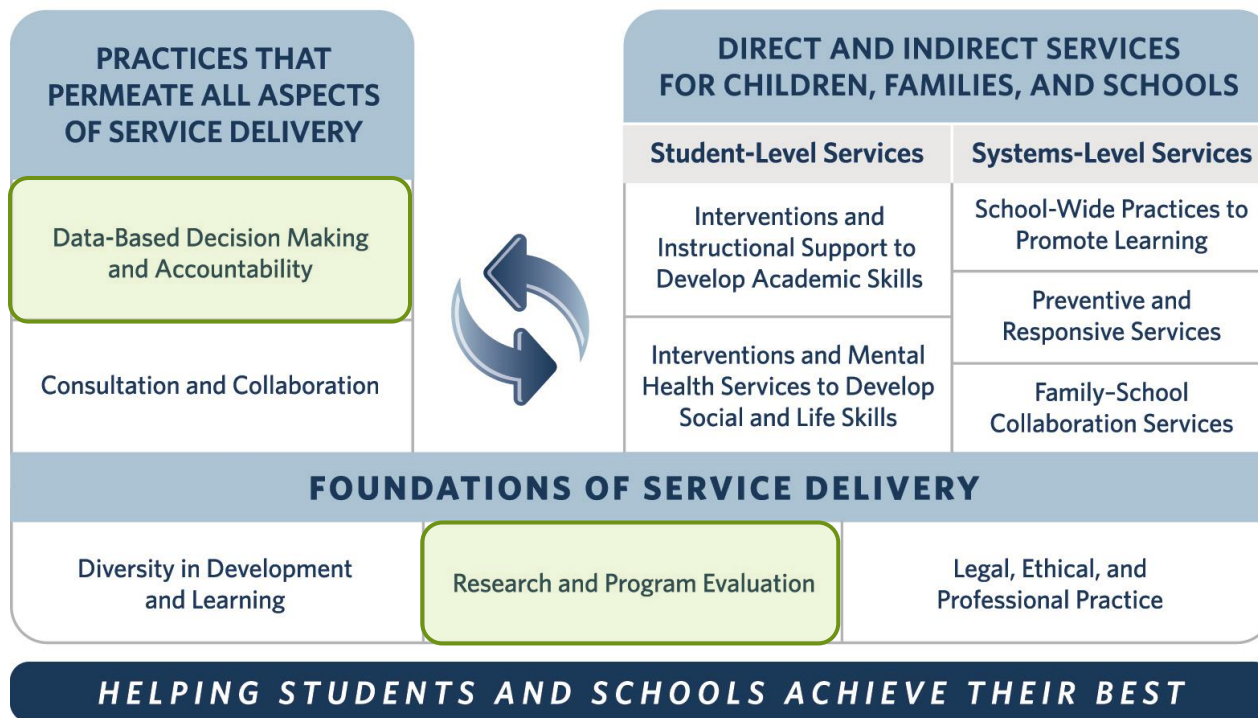
University at Buffalo

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

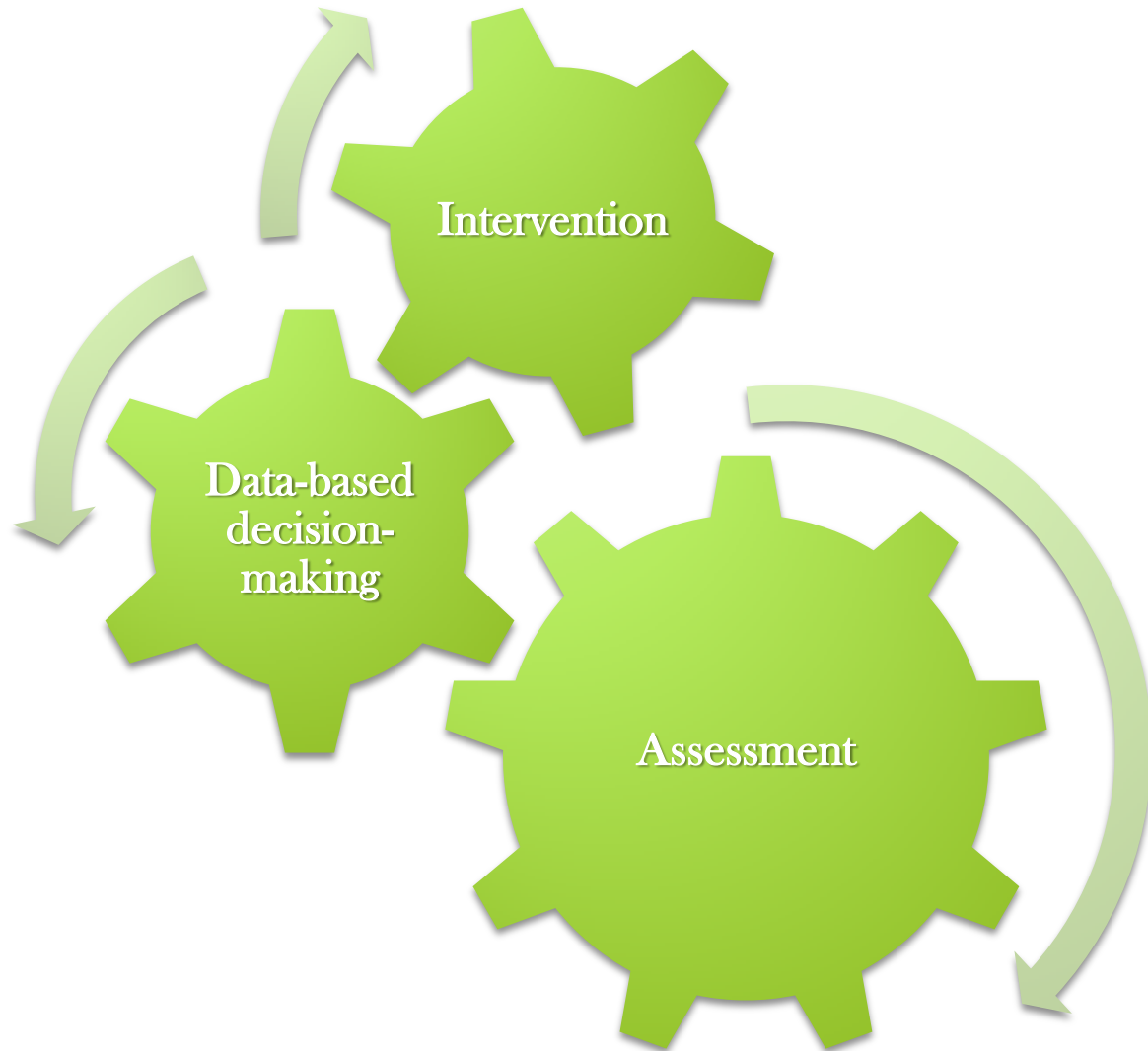
NATIONAL ASSOCIATION OF SCHOOL PSYCHOLOGISTS

Model for Services by School Psychologists



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

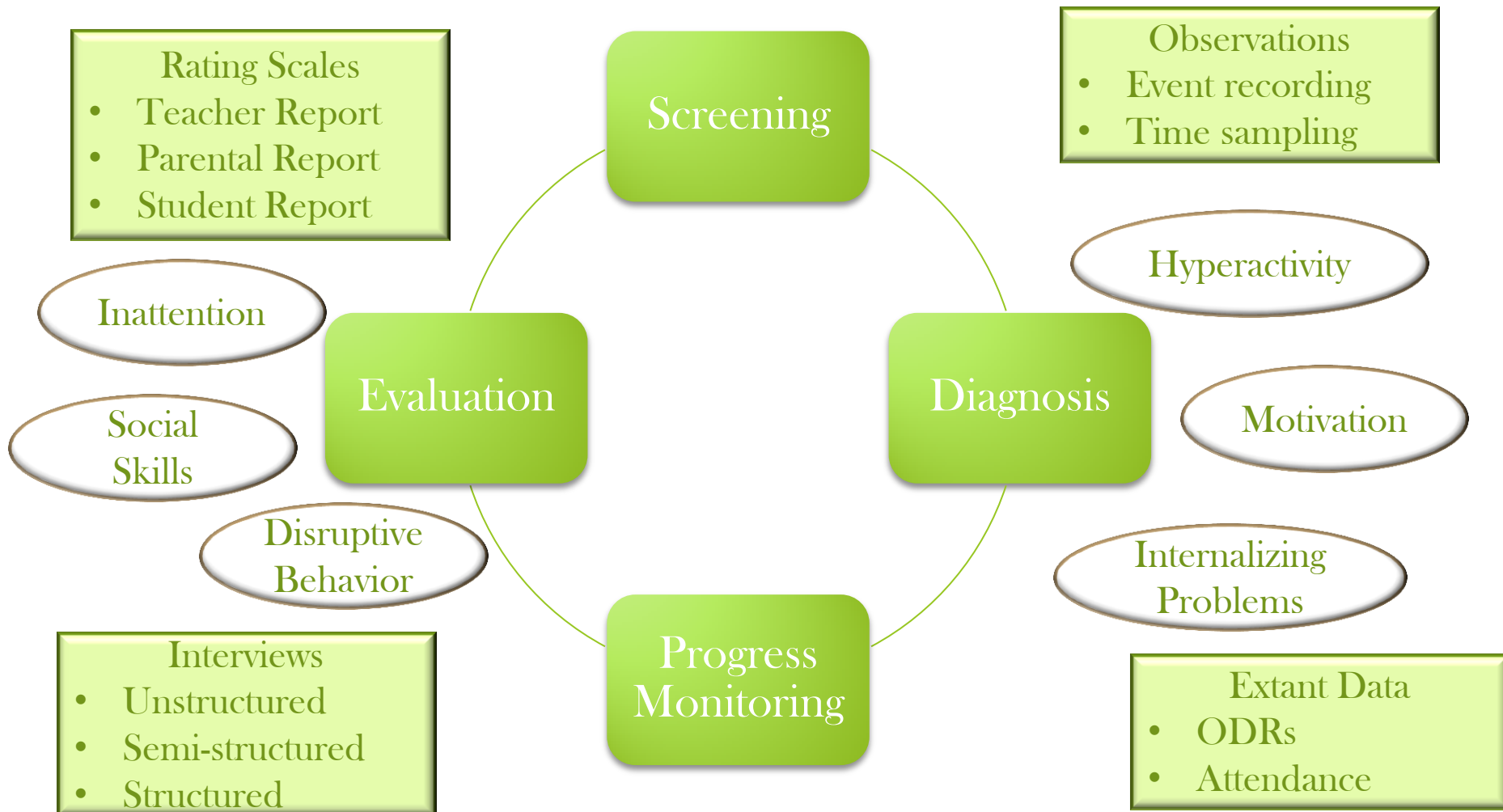
- Screening
 - Who needs help?
- Diagnosis
 - Why is the problem occurring?
- Progress Monitoring
 - Is intervention working?
- Evaluation
 - How well are we doing overall?

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



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

Behavioral Assessment



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...



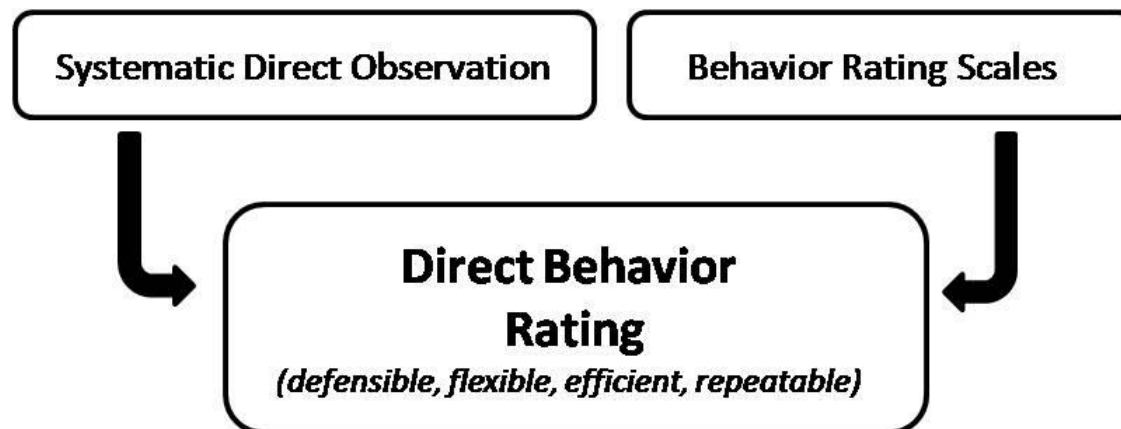
(Chafouleas, 2011; Chafouleas, Christ, & Riley-Tillman, 2009; Chafouleas, Volpe, Gresham, & Cook, 2010)



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

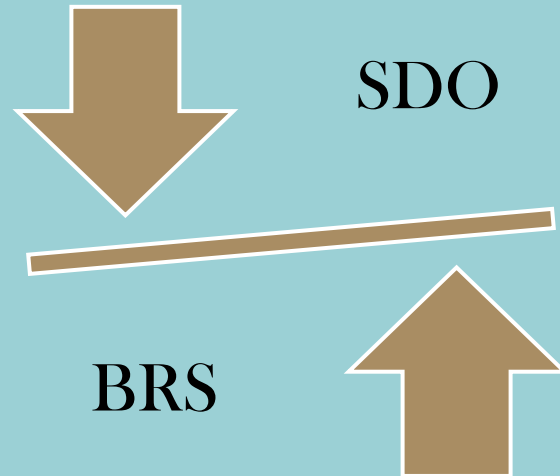


A little background...

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:

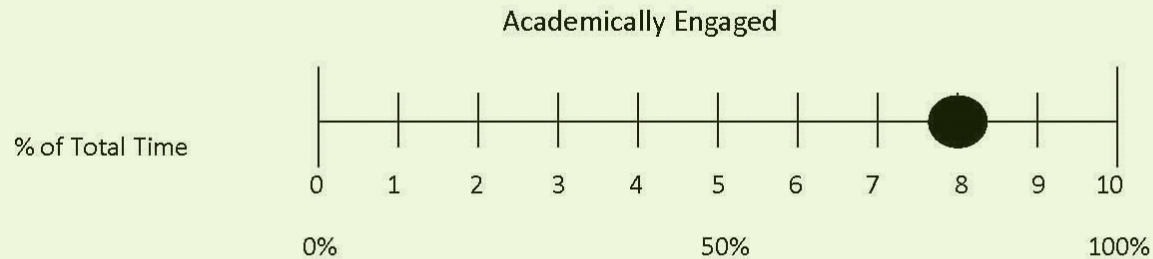


Used repeatedly to represent behavior that occurs over a specified period of time (e.g., 4 weeks) and under specific and similar conditions (e.g., 45 min. morning seat work)

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.

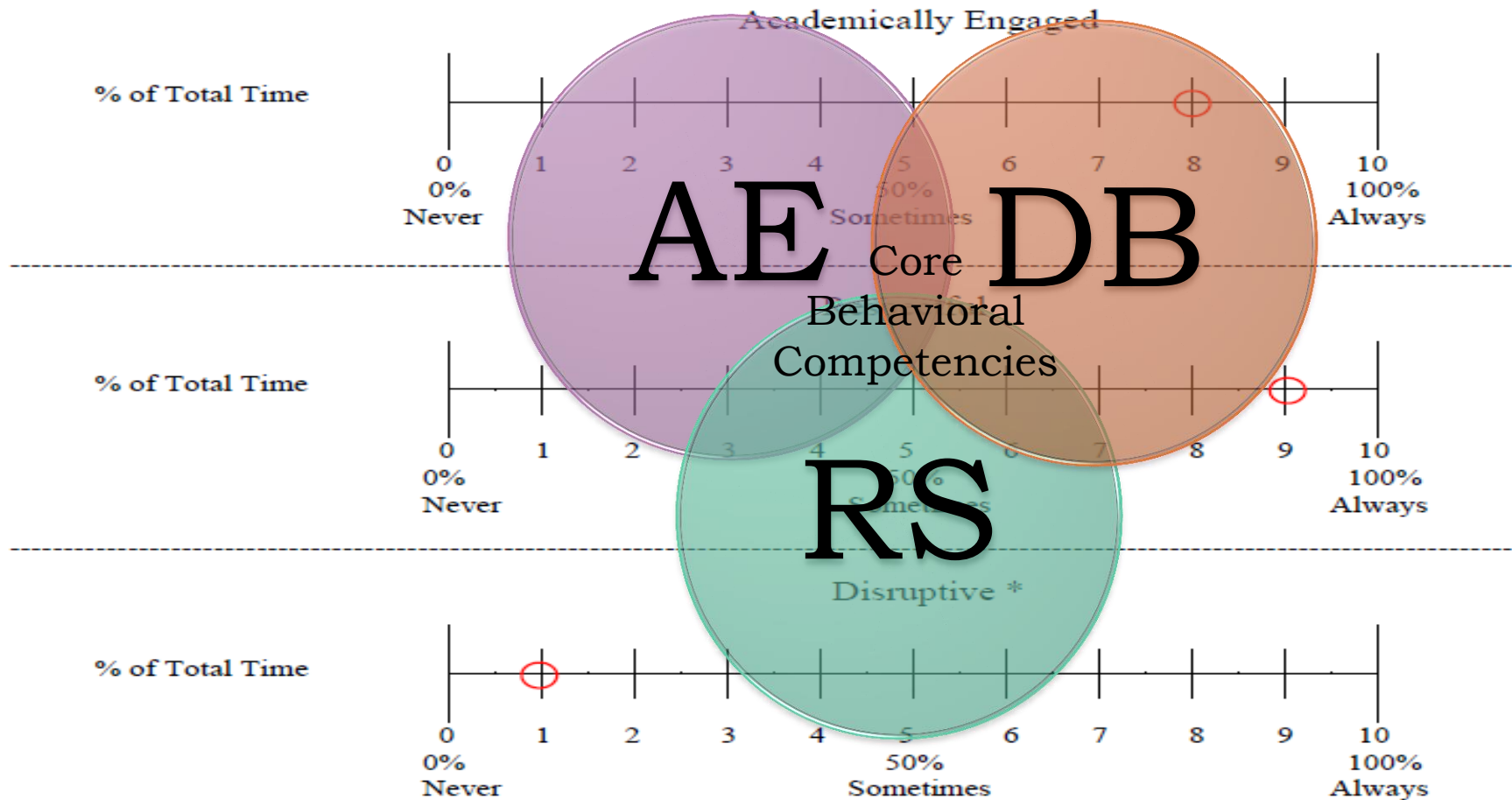
Multi-Item Scale

	<u>Never</u>		<u>Always</u>
Did the student follow class rules?	0	(1)	2
Did the student follow teacher directions?	0	1	(2)
Did the student do his/her best work?	0	1	(2)
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:

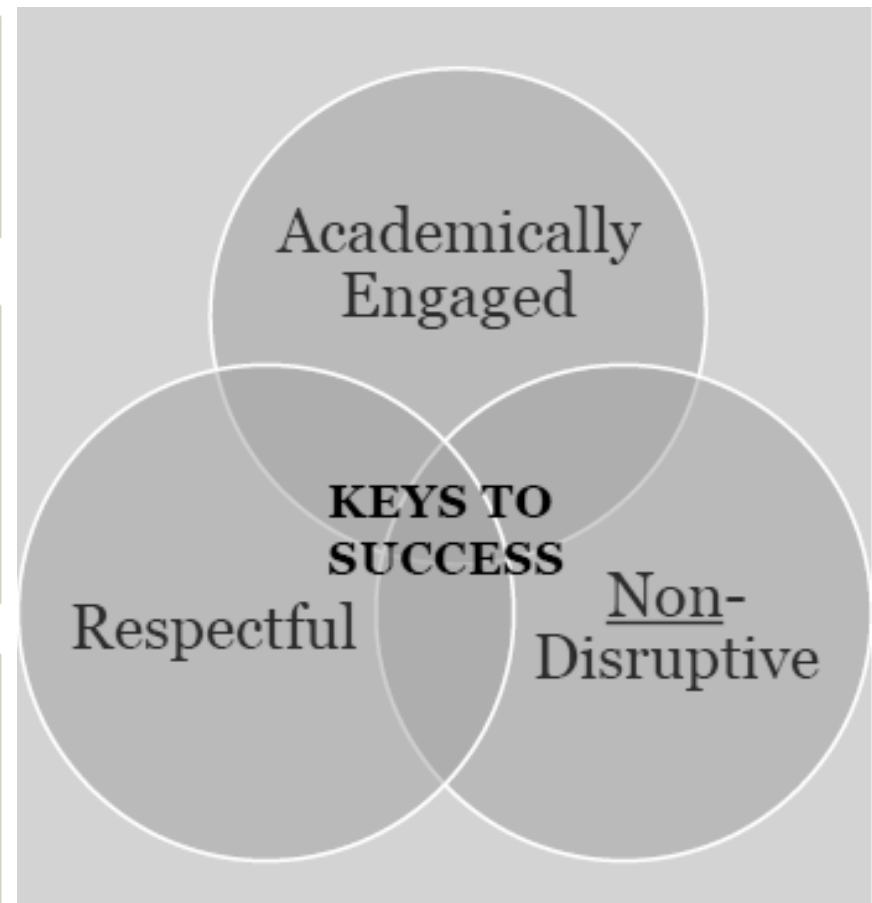
Actively or passively participating in the classroom activity.

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.

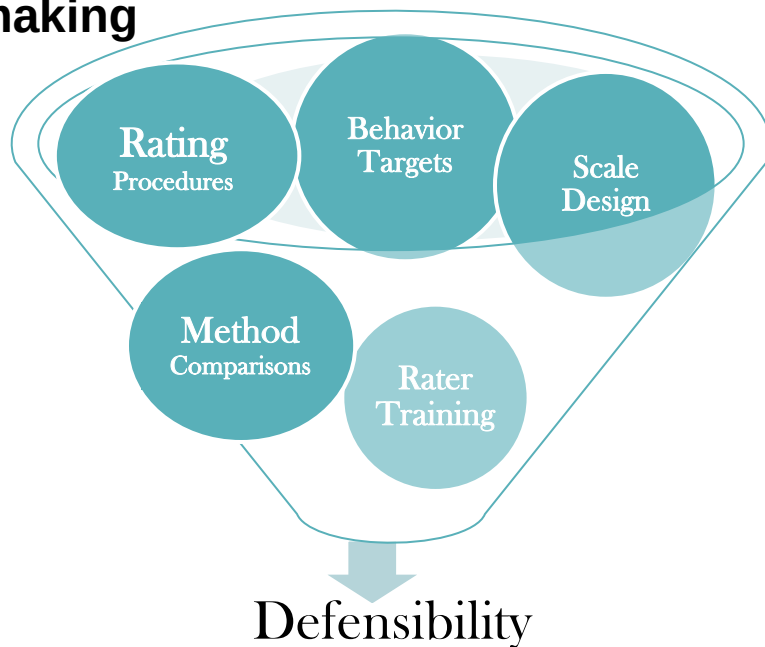




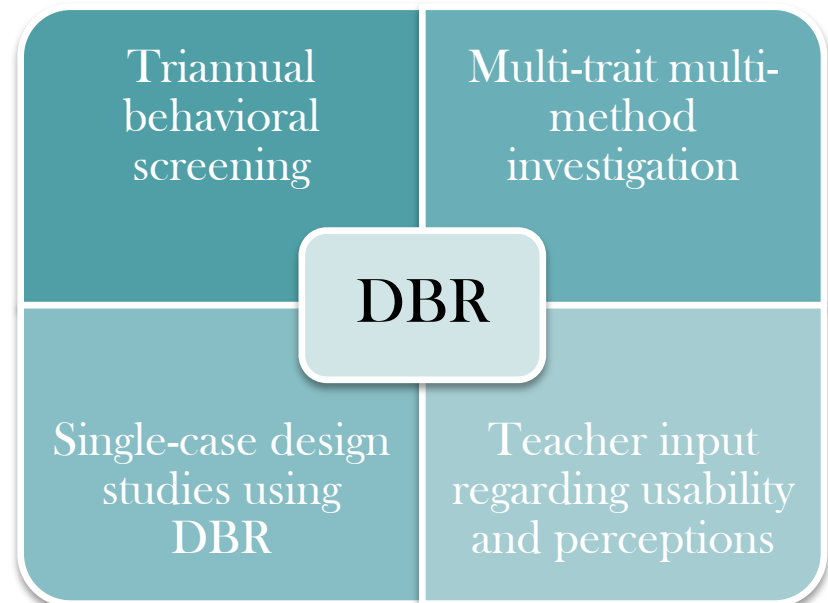
Development & Validation of DBR-SIS

RESEARCH: Project VIABLE (2006-2011) and Project VIABLE II (2011-current)

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



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



Funding provided by the **Institute for Education Sciences**, U.S.
Department of Education

Development & Validation

Development & Validation of DBR-SIS

Scale development

Behavior wording

Training

Influence of observation duration

How teachers assign ratings

Perceptions of usability

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



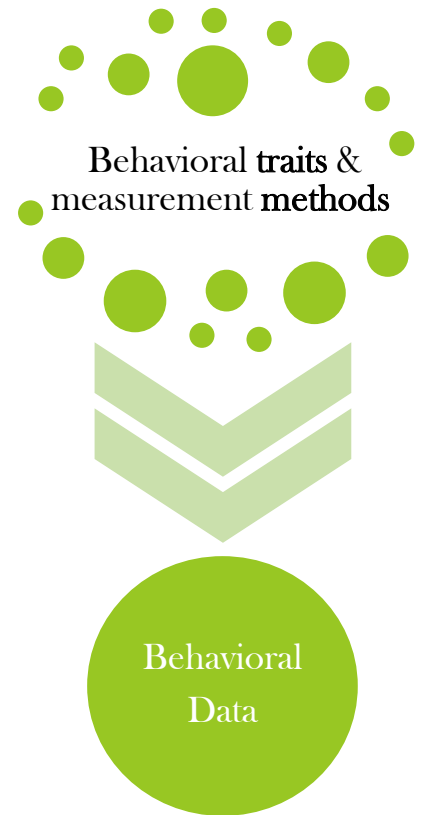
Multitrait Multimethod Analysis

Rationale

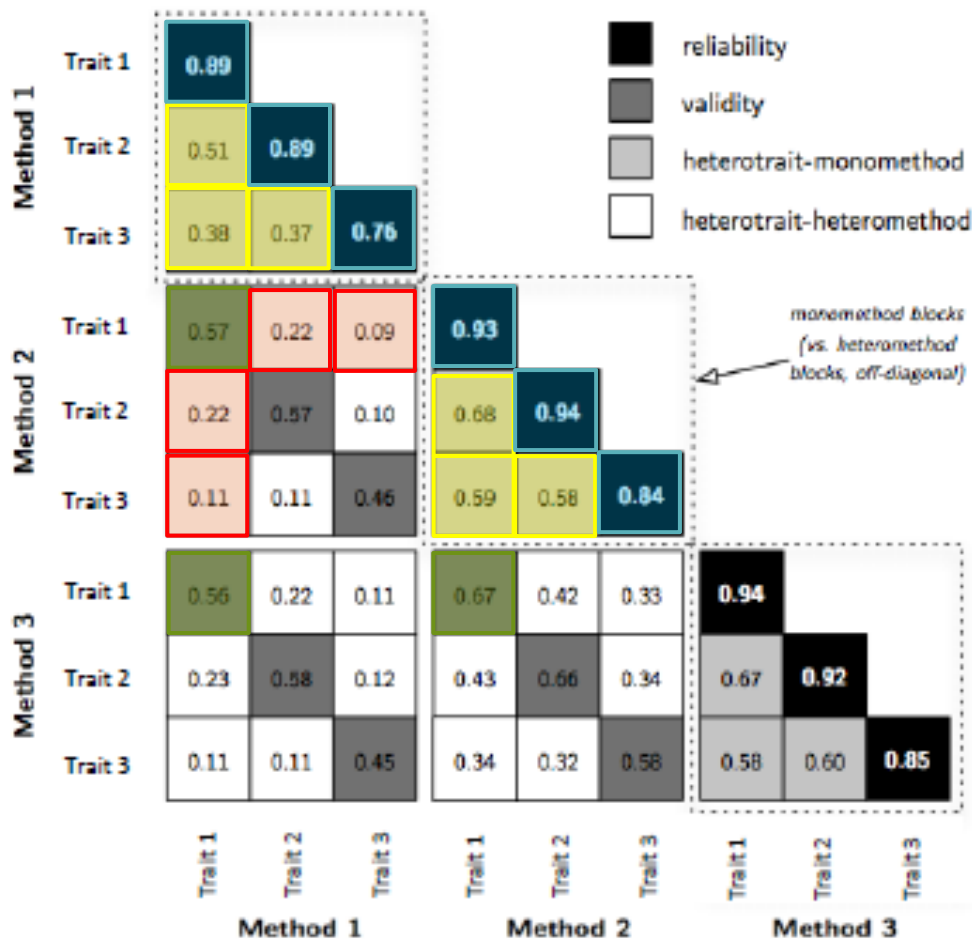
- Test developers must accurately define, measure, and rigorously validate the construct(s) of interest
- Campbell and Fiske (1959) developed an approach to assessing construct validity
 - MTMM analysis permits the examination of:
 - Convergent validity - evidence that scores are consistent with other 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

Student characteristic	<i>n</i>	%
Gender		
Male	452	45
Female	541	55
Race		
Caucasian	780	79
African American	154	16
Asian	35	4
Other	12	1
Ethnicity		
Hispanic	89	9
Non-Hispanic	904	91
Grade		
Third	210	21
Fourth	204	21
Fifth	206	21
Sixth	166	17
Seventh	124	12
Eighth	83	8

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 (α)
 - Student rating scales: internal consistency (α)

	Method 1			Method 2			Method 3			Method 4			Method 5		
	A	B	C	A	B	C	A	B	C	A	B	C	A	B	C
1. DBR – Teacher															
a. Academic Engagement	.90									Rules of thumb for interpreting correlations: <.20 = Weak .20-.69 = Moderate >.69 = Strong					
b. Disruptive Behavior	-.87	.88													
c. Respectful	.81	-.91	.88												
2. DBR- Student															
a. Academic Engagement	.49	-.41	.41	.82											
b. Disruptive Behavior	-.45	.48	-.44	-.75	.80										
c. Respectful	.45	-.47	.47	.96	-.84	.81									
3. SDO															
a. Academic Engagement	.37	-.39	.33	.27	-.29	.30	.93								
b. Disruptive Behavior	-.29	.35	-.30	-.23	.23	-.24	-.80	.96							
c. Respectful	.21	-.28	.30	.16	-.19	.23	.48	-.61	.78						
4. Rating Scale – Teacher															
a. Academic Engagement ¹	-.75	.63	-.55	-.39	.41	-.35	-.24	.23	-.20	.95					
b. Disruptive Behavior ²	-.58	.71	-.65	-.35	.41	-.39	-.28	.28	-.27	.76	.95				
c. Respectful ³	.55	-.50	.48	.33	-.31	.31	.23	-.18	.20	-.67	-.55	.93			
5. Rating Scale - Student															
a. Academic Engagement ¹	-.47	.41	-.34	-.53	.50	-.53	-.25	.23	-.25	.48	.39	-.40	.77		
b. Disruptive Behavior ²	-.34	.39	-.32	-.38	.45	-.42	-.23	.24	-.21	.36	.47	-.24	.80	.75	
c. Respectful ³	.14	-.16	.11	.30	-.29	.33	.10	-.08	.03	-.15	-.15	.16	-.42	-.36	.71

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

