Development and Validation of Progress Monitoring Tools for Social Behavior: Lessons from Project VIABLE

Sandra M. Chafouleas, Project Director
Presented at the 2010 IES Conference

Project VIABLE:

Validation of ***I***nstruments for ***A***ssessing ***B***ehavior ***L***ongitudinally &***E***fficiently

GOAL: Develop and Evaluate Direct Behavior Rating (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 format, behavior targets, training requirements)
- Pilot testing various aspects with classroom teachers

Phase III: Evaluate feasibility and utility of DBR in school settings at small scale.

- Packaging what we have learned to train users
- Establish groups of teachers/schools willing to participate in DBR training and use
- Evaluate data/feedback

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).

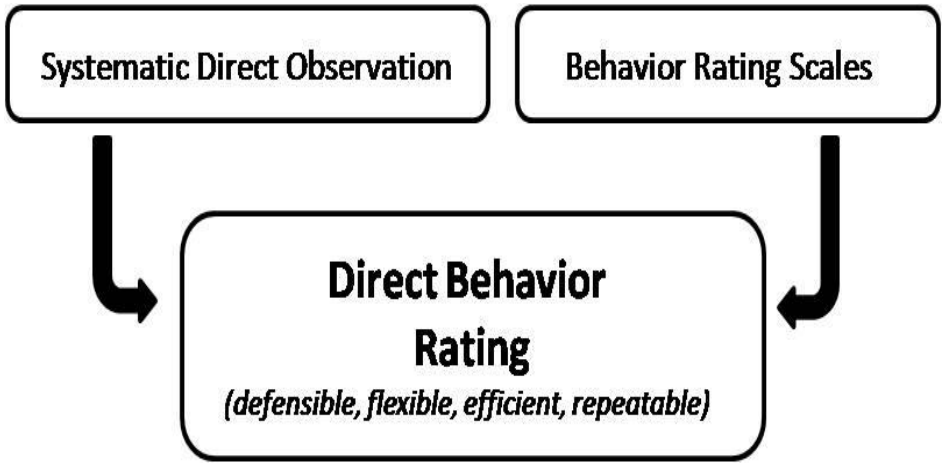
University of Connecticut
East Carolina University
University of Minnesota

CBER

Overview of DBR in Assessment: History & Defining Features

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)

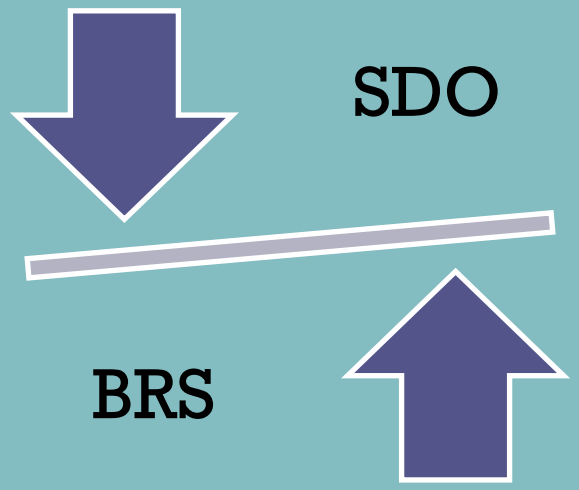
CBER

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

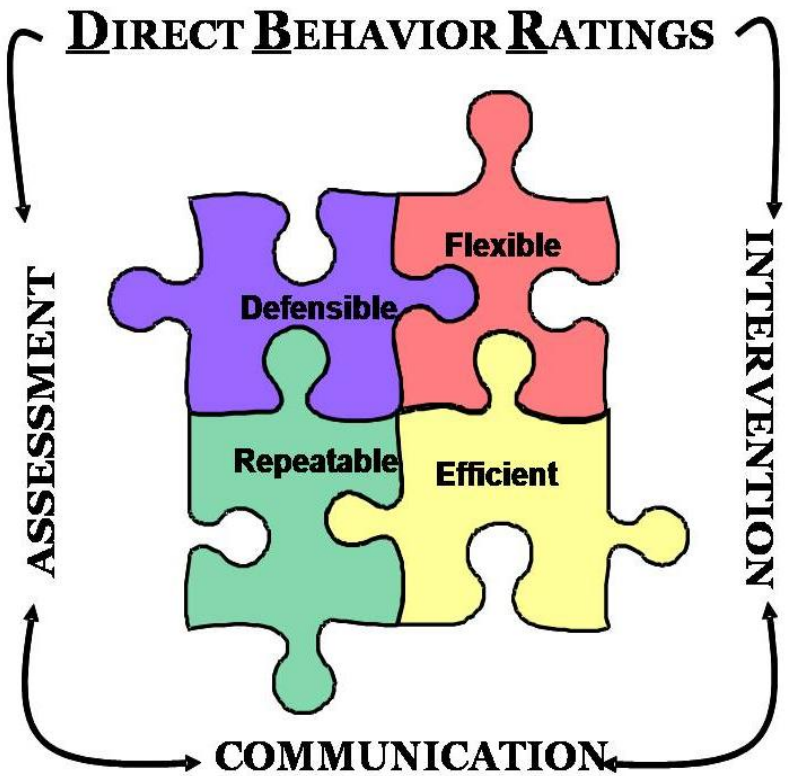
Contemporary Defining Features:



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)

CBER

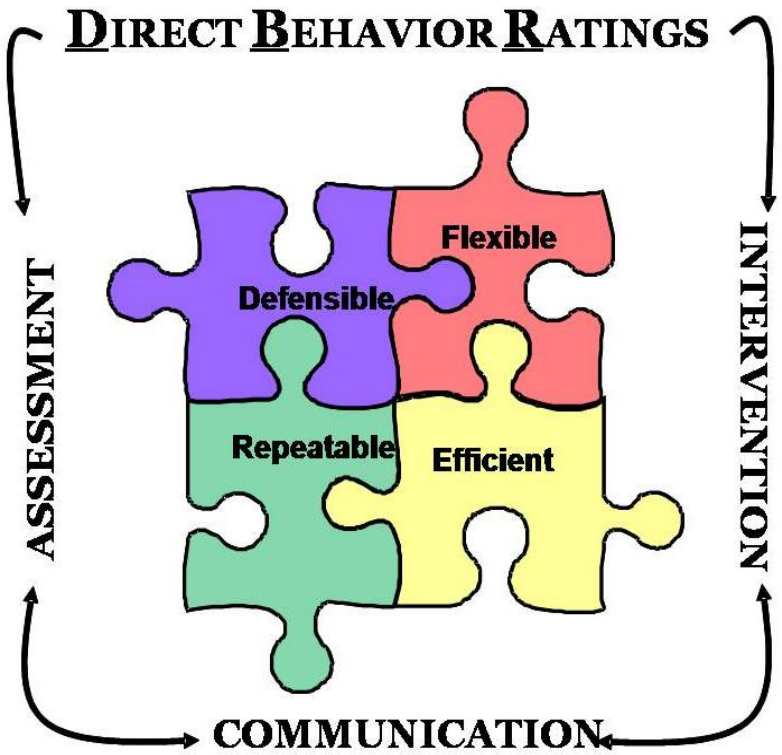
Direct Behavior Rating



Direct

- establishes that the observation and rating **occur at the time and place** that behavior occurs.
- This minimizes
  - inference &
  - retrospective judgments

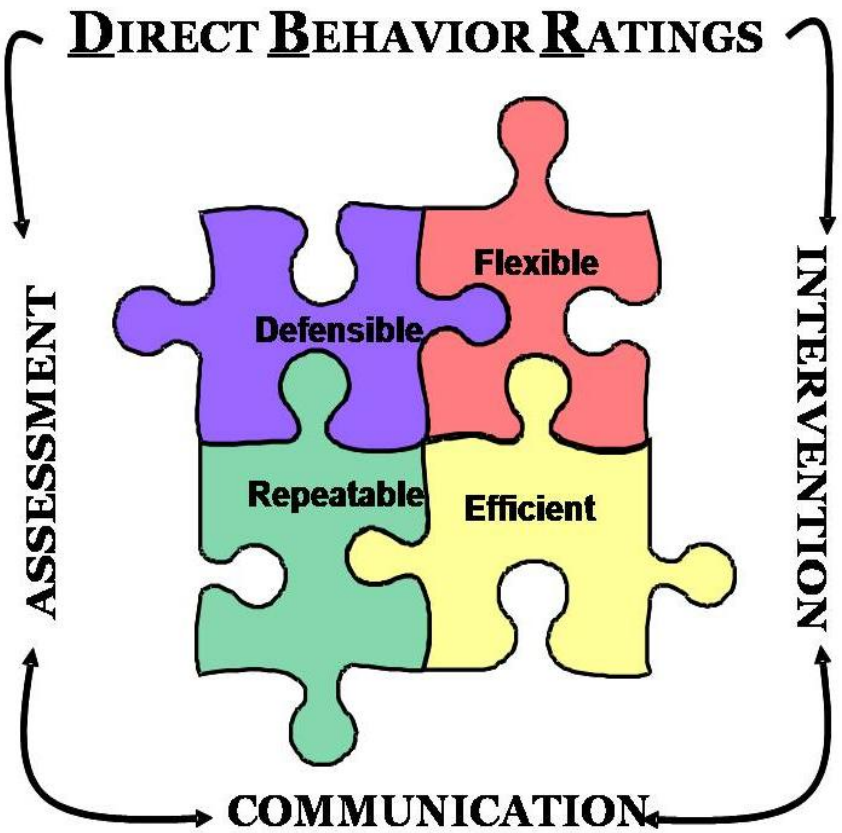
Direct Behavior Rating



Behavior

- the target of assessment must be **accessible for observation** and evaluation by the intended rater.
- the preference is to observe behavior within the **naturalistic setting**.
- contents/modalities for behavioral assessment are **motor**, **physiological**, and **cognitive** (Cone, 1978).

Direct Behavior Rating



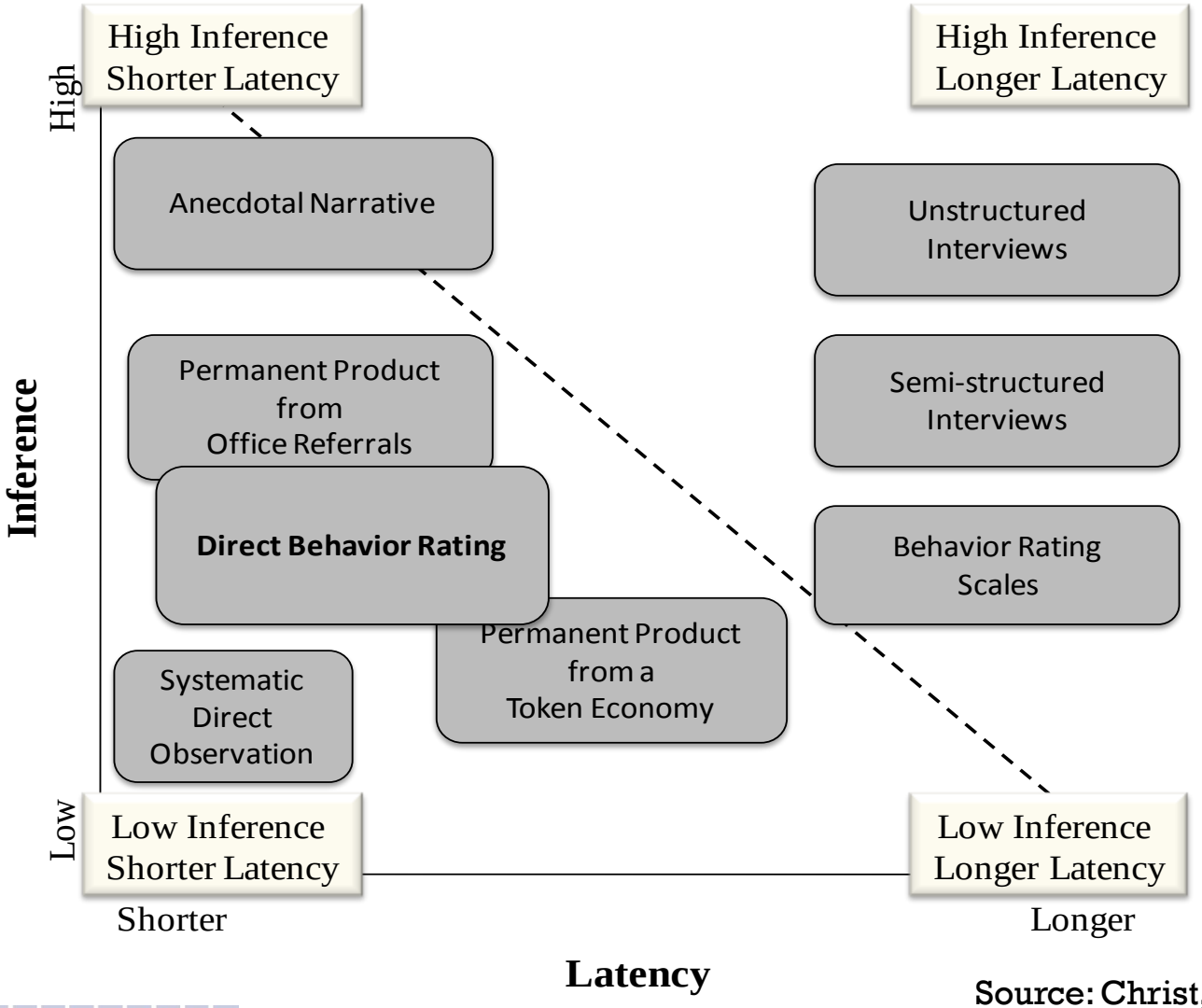
Rating

- quantify a person's **perception** or attitude toward something.

- DBR can be compared to any of a variety of other problem solving and behavioral methods
  - SDO
  - Interviews
  - behavioral rating scales

CBER

Direct Behavior Rating & Other Methods

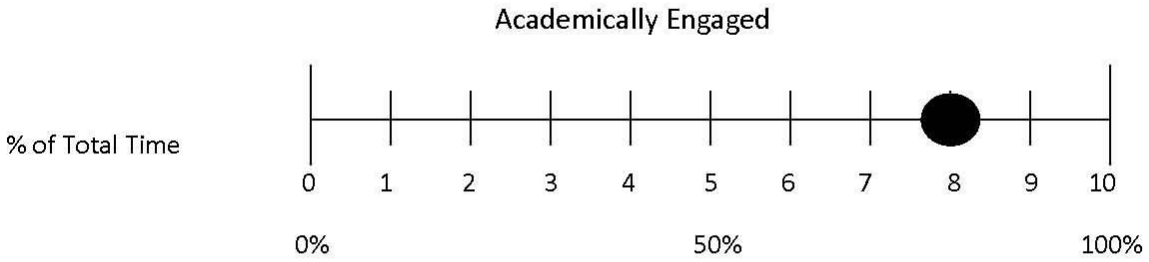


Source: Christ, Riley-Tillman, & Chafouleas (2009)

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

	Never		Always
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)

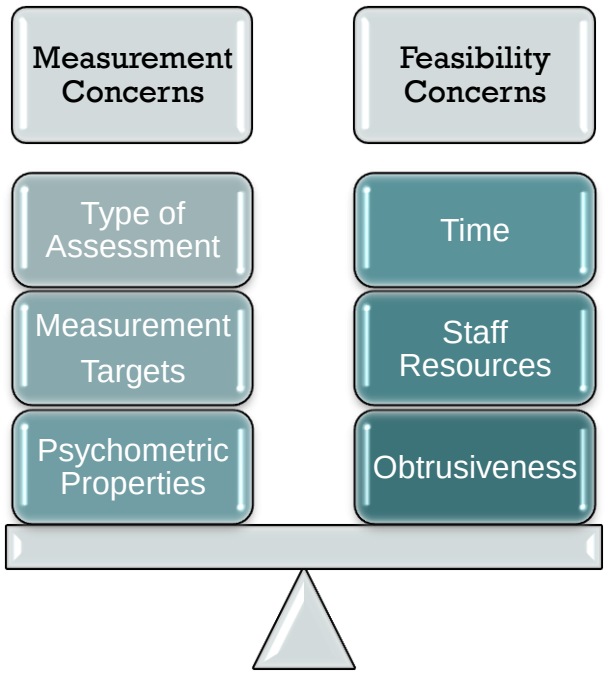
Total number of points earned: 5

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

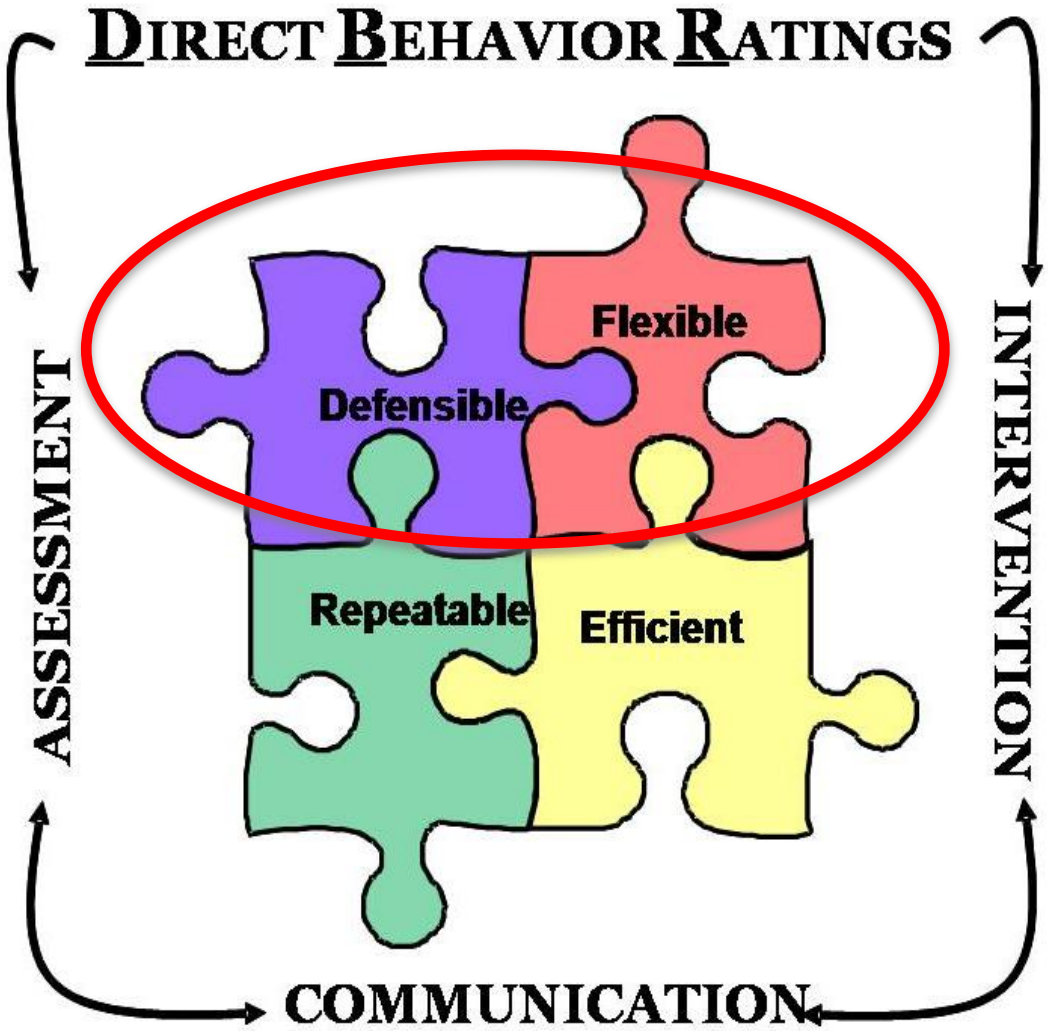
What are desirable features of progress monitoring tools for behavior?

- **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

Source: Chafouleas, Riley-Tillman, & Christ, 2009; Chafouleas, Riley-Tillman, & Sugai, 2007; Christ, Riley-Tillman, & Chafouleas, 2009)

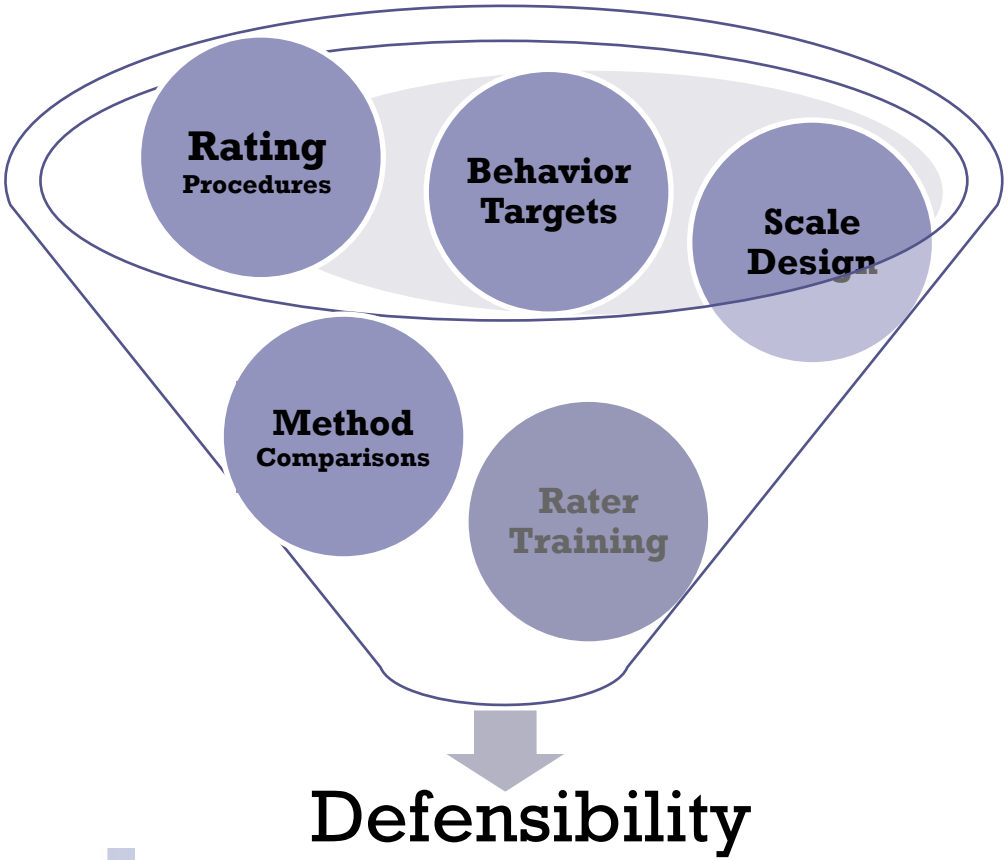


Adapted from Briesch & Volpe (2007)


DIRECT BEHAVIOR RATINGS
ASSESSMENT
INTERVENTION
COMMUNICATION
Flexible
Defensible
Repeatable
Efficient

Project VIABLE:

<u>Phases I & II</u> Develop instrumentation and procedures; evaluate defensibility of DBR in decision-making



Behavior Targets

So many options…

Considerations

- Molar v. molecular wording?
  - E.g. Disruptive – Out of Seat
- Negative v. positive wording?
  - E.g. Disrespectful – Respectful
- General Outcome v. Individualized Targets
  - Applicable to all – Relevant to Some/Few

Studies using Large Samples of Undergraduate Students

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

- The first attempt… DBR vs. SDO ("true score") comparisons of 3 target constructs and 2 wording.
- Wording and specificity of target construct can impact rater accuracy. Molar wording resulted in stronger correspondence and positive phrasing was stronger for academic engagement yet unclear for disruptive. "Compliance" definition needed revision.

Christ, Riley-Tillman, Chafouleas, & Jaffery (in review)

- Adding on… analyses to separate rater bias and error, and influence of base rates.
- High correspondence between DBR and SDO for Academic Engagement and Disruptive Behavior, but results for molecular behaviors were weak. Substantial rater bias was present (underestimate desirable and vice versa).

Chafouleas, Riley-Tillman, Jaffery, Sen, Music, & Christ (2010)

- And adding further… only molar behaviors of academic engagement, disruptive, and respectful. Comparisons with SDO and DBR-Expert and controlled the clips (base rates).
- DBR-Expert resulted in closer correspondence than SDO, Stronger evidence for Academic Engagement and Disruptive than Respectful, Medium levels of behavior harder to rate than low and high

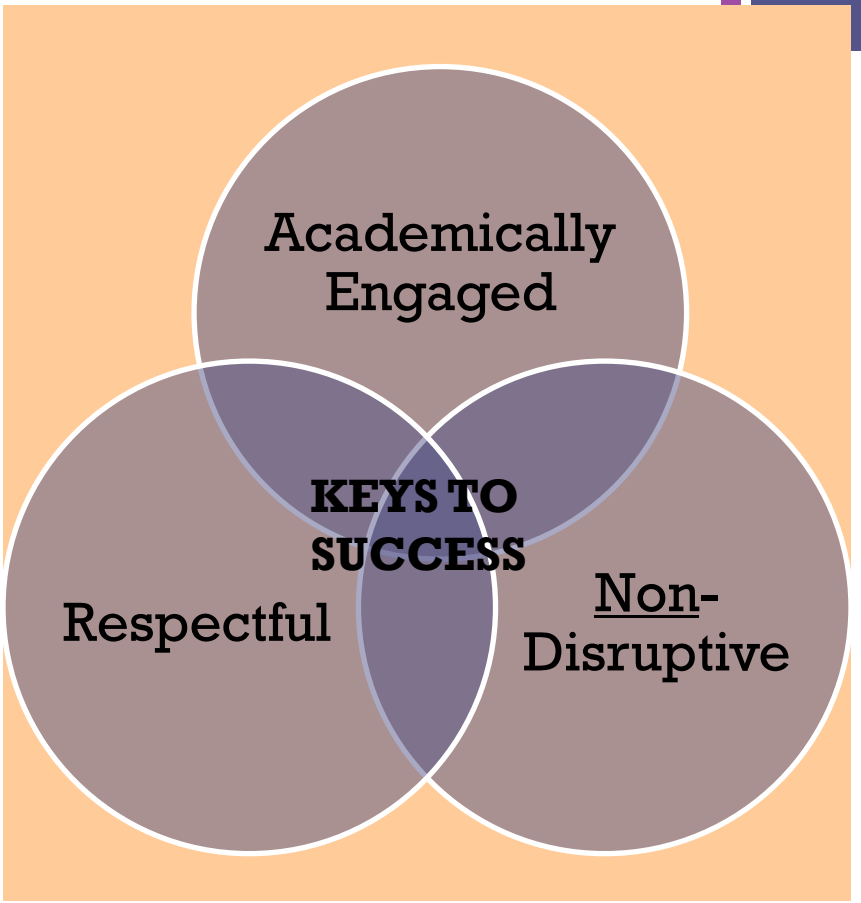
CBER

DBR-SIS Targets: "The Big 3" General Outcomes

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.



CBER

Example: Idiographic vs. General Outcome Target Selection

Vujnovic, Fabiano, Chafouleas, & Sen (under review)

- Sample: 13 boys with diagnosis of attention-deficit hyperactivity disorder
- Intervention: DRC-based intervention
- Design: Point, level, slope comparisons over 20 data collection days with both measures
- Measures: teacher-completed DBR-SIS (once at end of day) and DBR-MIS (completed multiple times each day)
- Conclusion: DBR instrumentation and procedures can be flexibly determined to match assessment situation

Point, Level, and Slope Estimates for DBR

		Mean (SD)
DBR-MIS		
	point	71.67*(31.68)*
	level	79.18*(18.52)*
	slope	-0.19 (0.61)
DBR-SIS: Academic Engagement		
	point	7.13*(2.19)*
	level	7.57*(1.36)*
	slope	-0.04 (0.05)
DBR-SIS: Non-Disruptive		
	point	8.05*(2.54)*
	level	7.66*(2.30)*
	slope	-0.06(0.08)

		DBR-SIS	
		AE	Non-DB
	Point	.854**	.830**
DBR-MIS	Level	.715**	.741**
	Slope	.415	.758**

Scale Design

So many options…

Considerations

- ***Number of Gradients?***
- ***Anchors?***
- ***Qualitative Descriptors?***
- ***Visual Cue?***

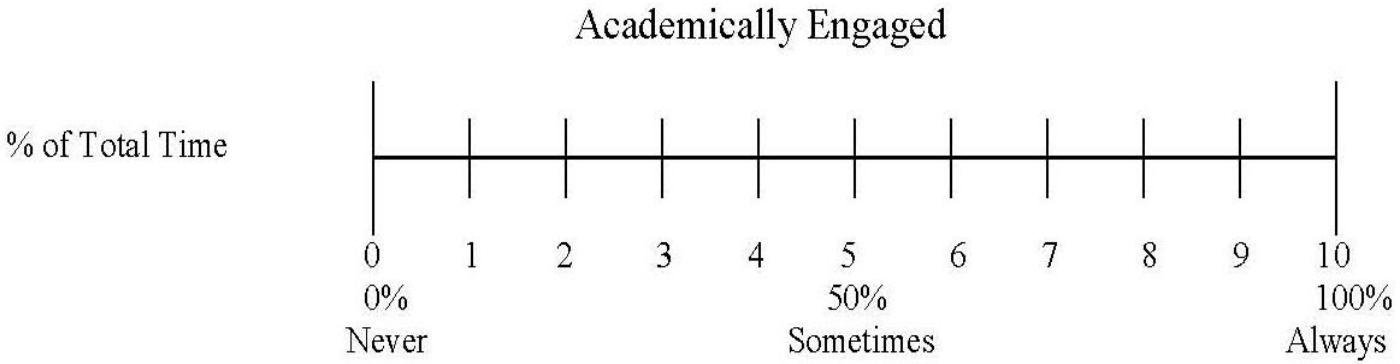
Guiding Principles: Built from Review, Large N Rater Samples, and Teacher Preference Assessments

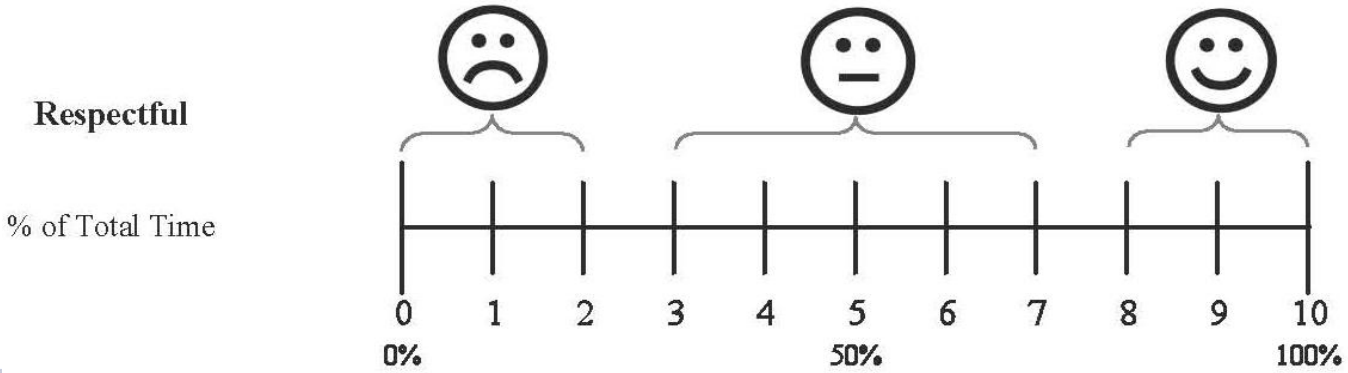
- Christ & Boice (2009); Christ, Riley-Tillman, & Chafouleas (2009)
  - Scales should be comprised of at least 6 gradients yet 10 appears optimal to facilitate ease of data interpretation and utility within visual analysis of formative data.
- Briesch, Kilgus, Chafouleas Riley-Tillman, & Christ (2010); Christ & Boice (2009)
  - Scales can use a variety of physical options. A line can be used to provide a visual cue toward rating, although the total length of the line does not impact reliability or accuracy.
- Riley-Tillman, Christ, Chafouleas, Boice, & Briesch (2009); Riley-Tillman, Chafouleas, & Music (2009)
  - Scales may vary with regard to WHAT is rated (duration, proportion), and no strong preferred design has emerged among teachers

CBER

Our DBR-SIS Scale

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.





Comparisons Across Methods and Raters

$E\hat{\rho}^2$

Method Comparison: Reliability of Data Obtained from DBR and SDO

G Study Full Model Results: Comparison of Variance Components by Rater Type

	Teachers % Var	Observers % Var
Person	47	48
Day	0	1
Occasion:Day, Occ x Day	2.5	5
Rater	7.5	0
Person x Rater	20	0
Person x Day	0	2
Rater x Day	2	0
Person x Rater x Day	3	1
Residual	17	44
(Person x Occasion:Day)	(0)	(30)
(Rater x Occasion:Day)	(4)	(0)
(3-way interaction plus error)	(13)	(14)
$E\rho^2$	.82	.98
Φ	.77	.97

Results of Decision Studies with Conditions Varied by Day and Rater Type

		1 day	5 days	10 days	15 days	20 days	100 days
1 observation/day[1]							
Researcher-conducted SDO	$E\rho^2$	.50	.83	.91	.93	.98	.99
	Φ	.48	.82	.90	.93	.97	.99
Teacher-completed DBR	$E\rho^2$	.54	.66	.68	.69	.70	.70
	Φ	.47	.58	.61	.62	.63	.63
3 observations/day[2]							
Researcher-conducted SDO	$E\rho^2$	.73	.93	.96	.97	.98	.99
	Φ	.70	.92	.96	.97	.97	.99
Teacher-completed DBR	$E\rho^2$	.62	.68	.69	.69	.70	.70
	Φ	.55	.60	.62	.62	.63	.63

Effects of Rater and

Chafouleas, Briesch, Riley-Tillman, Christ, Black, & Kilgus (2010)

- Sample: 2 teachers and 2 research assistants – 7 middle school students in the same Language Arts classroom
- Measures: researcher-completed and teacher-completed DBR-SIS for Academic Engagement and Disruptive Behavior over 6 days (3x/period)
- Analyses : Multiple imputation to handle substantial missing data, Generalizability theory
- Conclusion: Degree of reliability-like estimates can differ substantially depending on individual rater. In the absence of estimates of rater reliability and firm recommendations regarding rater training, ratings obtained from DBR-SIS, and subsequent analyses, be conducted within rater.

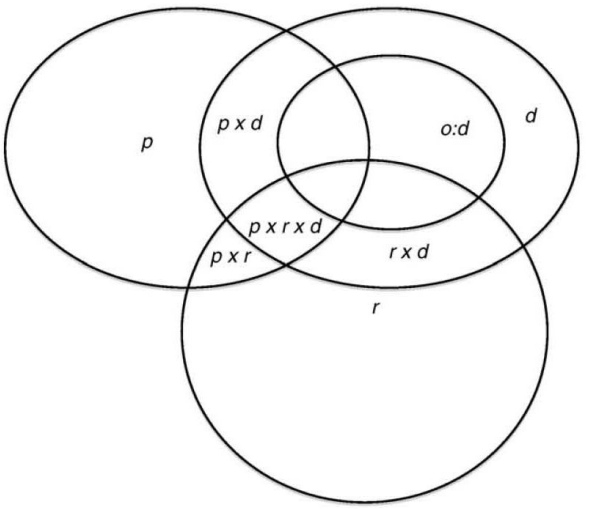


Fig. 1. Estimable variance components for design involving facets of person (*p*), rater (*r*), and occasion within day (*o*:*d*).

Table 1
Interpretation of estimated variance components and interactions.

Variance component	Interpretation
Person (*p*)	Absolute differences in observed behavior among students
Day (*d*)	Changes in overall student behavior across time, given common instructional block
Rater (*r*)	Differences in overall rating behavior among raters
Occasion:day (*o*:*d*) [a]	Differences in mean ratings between occasions within a particular instructional block (i.e., day)
Person × day (*p* × *d*)	Changes in the relative standing of students across time
Person × rater (*p* × *r*)	Differences in the relative standing of particular students among raters
Day × rater (*d* × *r*)	Changes in rating behavior across time
Person × day × rater (*p* × *d* × *r*)	Changes in how individual raters judge the relative standing of students across time
Error [a]	Residual variance, including variance contributed by interactions involving occasion:day

[a] Interactions associated with *o*:*d* are not separately presented for interpretation due to nesting of occasion within day. Attributable variance for po:*d*, ro:*d*, and pro:*d* is accounted for within the residual term.

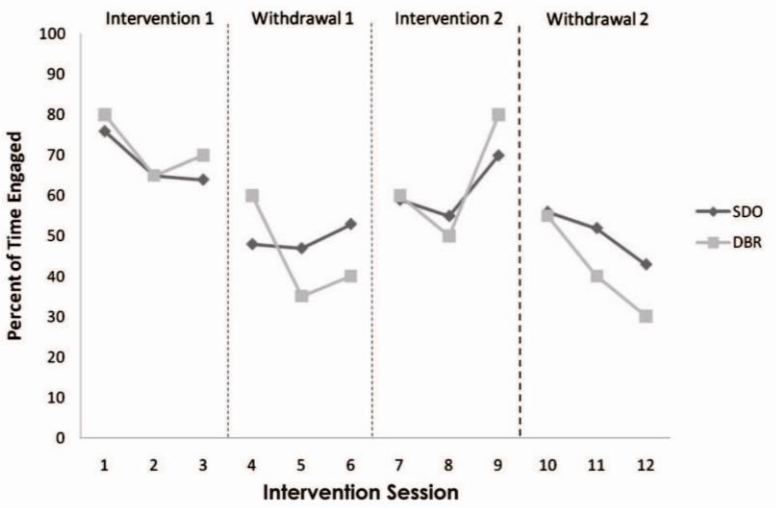
CBER

Case Study: Method Comparison in Classwide Assessment

Riley-Tillman, Methe, & Weegar (2009)

- Sample: First grade classroom with 14 students
- Design: B-A-B-A
- Intervention: modeling and prompting of silent reading
- Measures: researcher-completed SDO, teacher-completed DBR-SIS
- Conclusion: DBR data can be sensitive to classroom-level intervention effects, maps closely to resource-intensive SDO

Systematic Direct Observation and Direct Behavior Rating Data of Engagement



	Phase Mean			
	B1	A1	B2	A2
DBR	72	45	63	42
SDO	68	49	61	50

Evaluating DBR-SIS Sensitivity to Change

Chafouleas, Sanetti, Kilgus, & Maggin (in prep)

- Sample: 20 teacher-student dyads in elementary grades
- Design and Intervention: A-B intervention involving behavioral consultation and DRC-based intervention. Five options for "change metrics" were calculated.
- Measures: researcher-completed SDO, teacher-completed DBR-SIS
- Conclusion: Change (in expected directions) in student behavior across phases and sources. High correspondence between DBR-SIS and BOSS absolute change metrics suggests that students were ranked similarly across the two measures with regard to intervention responsiveness. Provides preliminary support for the use of DBR-SIS to differentiate between those who have or have not responded to intervention.

Descriptive statistics across scales and phases

			Mean	SD
DBR-SIS	Disruptive Behavior	Baseline	4.26	1.97
		Intervention	2.58	1.41
	Academic Engagement	Baseline	4.97	2.28
		Intervention	6.82	1.50
	Compliance	Baseline	5.74	1.93
		Intervention	7.34	1.31
BOSS	On-task	Baseline	69.98	19.76
		Intervention	81.94	14.22
	Off-task	Baseline	44.82	21.01
		Intervention	28.69	18.54

Correlations between DBR-SIS and BOSS absolute change metrics

BOSS Scale	DBR-SIS		
	Disruptive Behavior	Academic Engagement	Compliance
On-task	-.458	.441	.299
Off-task	.487*	-.582*	-.554*

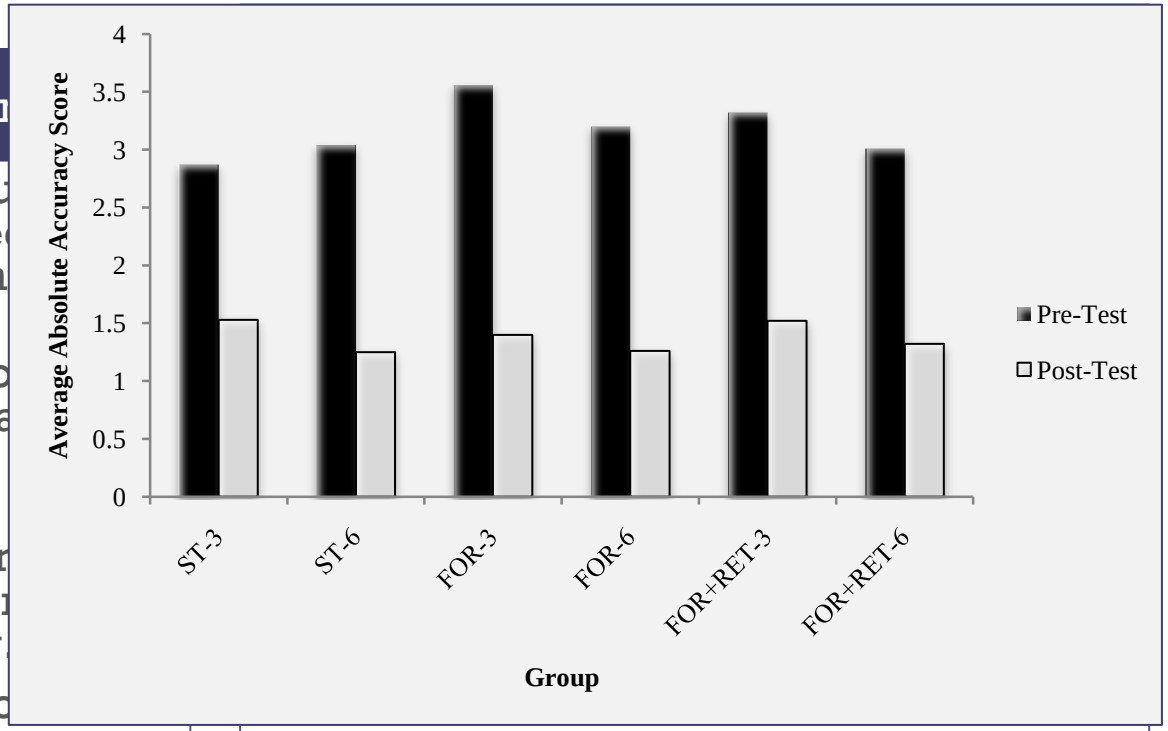
CBER

Rater Training

+ Options for RATER TRAINING

Considerations

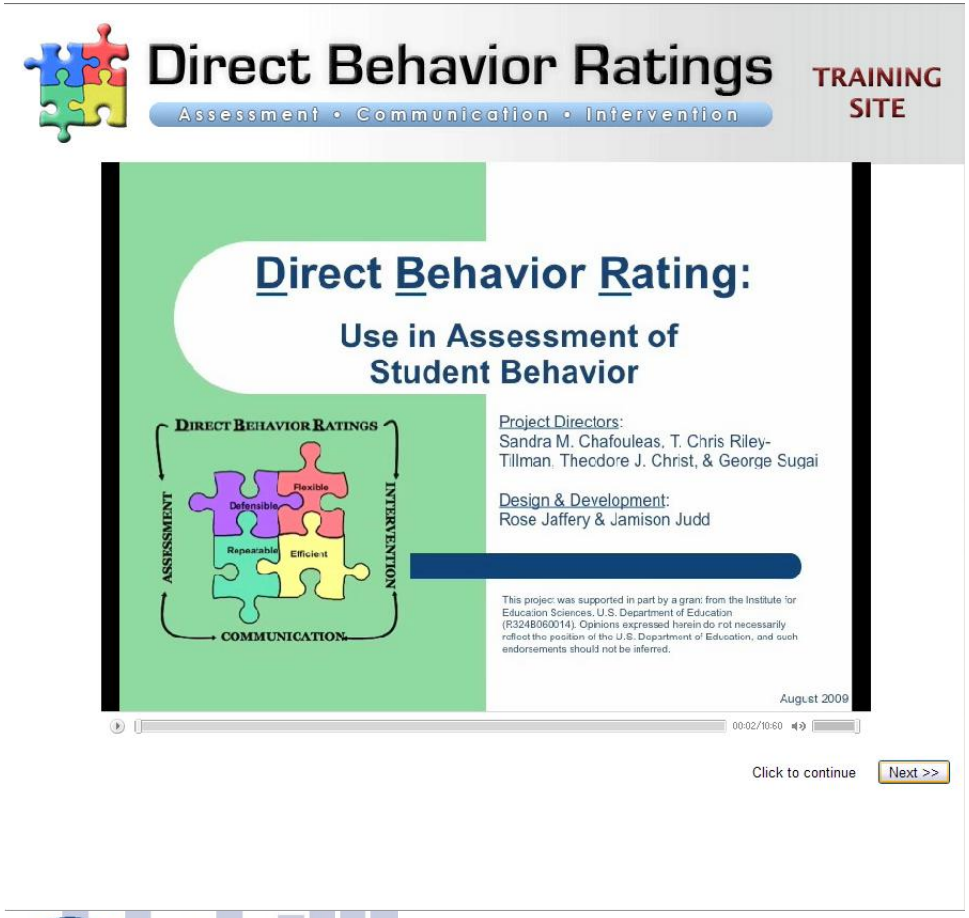
- What level of accu might be expecte absence of trainin
- Are some behavio more difficult to ra accurately?
- What improvemer be expected give training involving
  - Information abo
  - Information about Rater Bias
  - Modeling
  - Demonstration
  - Performance feedback?

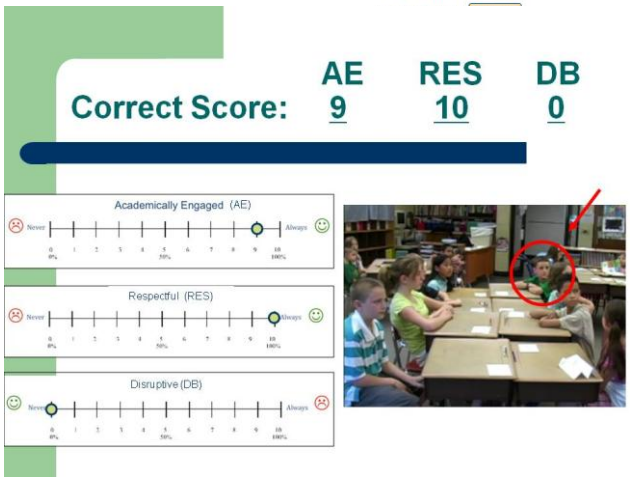


- *Chafouleas, Kilgus, Riley-Tillman, & Jaffery (2010)*
  - Adding on… impact of Frame of Reference and Rater Error Training added … control of base rates of behavior and varied "exposure" to performance feedback
  - "**Exposure**" mattered for some clips… thus, "Standard Training" should suffice as long as sufficient opportunities for practice and feedback are provided.

CBER

DBR-SIS:
3-Part On-Line Training Module





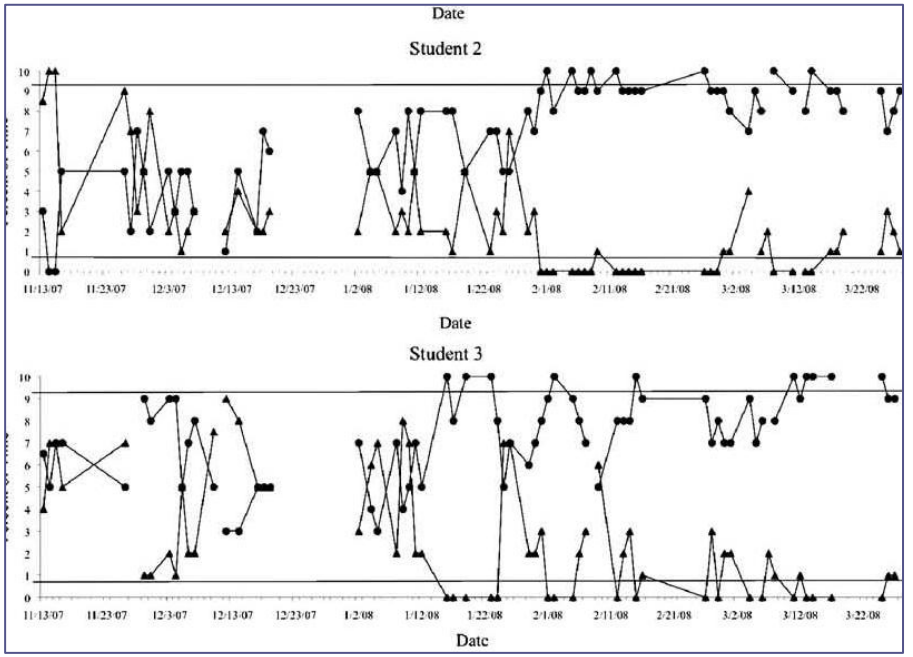
DBR Evaluation: Next Steps

Variability across Time and Grade

Chafouleas, Kilgus, & Hernandez (2009)

- Sample: full day K inclusive classroom, 2 teachers and 22 students
- Measures: teacher-completed DBR-SIS following am and pm over Nov-March for ALL students
- Conclusion: "Local" cut-score comparisons can be useful in examining individual student performance. Periodic re-assessment of all may be needed to re-confirm appropriate comparison

Target Behavior	Rating Time	FALL M (SD)	SPRING M (SD)
Academic	AM	8.72 (1.31)	9.40 (0.63)
Engagement	PM	8.25 (2.03)	9.37 (0.88)
Disruptive	AM	1.30 (1.47)	0.60 (0.62)
Behavior	PM	1.61 (2.08)	0.42 (0.52)



Understanding "Cut-Points"

Kilgus, Chafouleas, Riley-Tillman, & Welsh (in prep)

- Purpose: To evaluate the diagnostic accuracy of DBR-SIS (Disruptive Behavior, Academic Engagement, Compliance)
- Sample: Second grade teachers and randomly selected students in their classrooms
- Measures: teacher-completed DBR-SIS following am and pm over 1 week, BESS and SSiS Perf.Screener
- Analyses: Diagnostic accuracy statistics
- Conclusion: DBR may provide efficient initial identification of potential risk, but may need to be confirmed through additional gates. Interpretation of DBR-SIS "cut-score" highly dependent on what is considered to be a "criterion" indicator of behavioral risk.

Example: DBR-SIS with BESS Criterion

Target Behavior	Cut Score	SN	SP	PPP	NPP
Disruptive	>=0	100.00	0.00	19.5	
Behavior	>0	95.65	38.95	27.5	97.4
	>1	91.30	67.37	40.4	97.0
	>2	78.26	85.26	56.2	94.2
	>3	56.52	89.47	56.5	89.5
	>4	34.78	97.89	80.0	86.1
	>5	21.74	97.89	71.4	83.8
	>6	13.04	100.00	100.0	82.6
	>7	8.70	100.00	100.0	81.9
	>9	0.00	100.00		80.5
Academic	< 3	0.00	100.00		80.5
Engagement	<=3	8.70	100.00	100.0	81.9
	<=4	17.39	96.84	57.1	82.9
	<=5	30.43	93.68	53.8	84.8
	<=6	47.83	91.58	57.9	87.9
	<=7	86.96	81.05	52.6	96.2
	<=8	100.00	66.32	41.8	100.0
	<=9	100.00	37.89	28.0	100.0
	<=10	100.00	0.00	19.5	

CBER

Rater Flexibility

Chafouleas, Sanetti, Jaffery & Fallon (in prep)

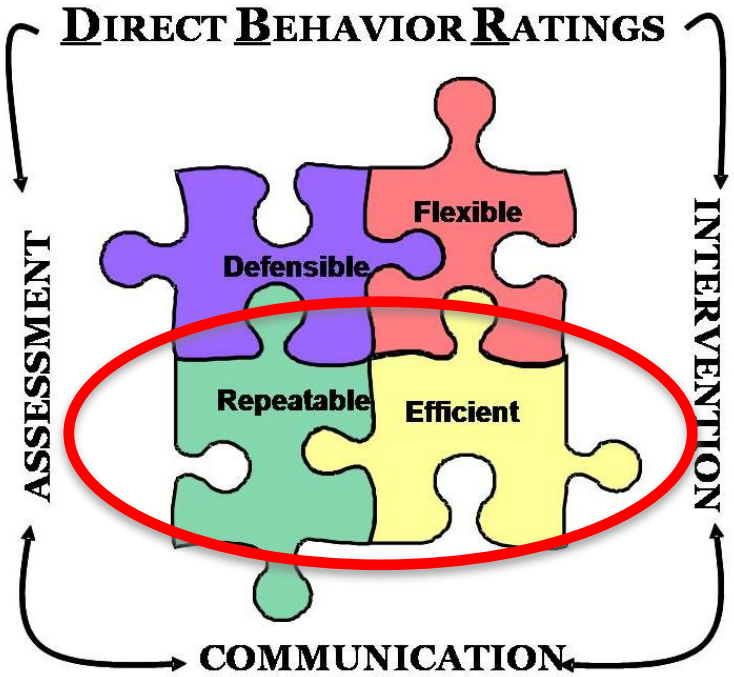
- Sample: 8th grade, 2 teachers and 3 classrooms (17-24 students)
- Design: Multiple baseline across classrooms
- Intervention: Self-monitoring and a group contingency package, implemented over about 2 months
- Measures: student-completed DBR (teacher-checked), researcher-completed SDO
- Conclusion: Classwide intervention overall effective, think about target identification and need for supports based on baseline

DBR-SM and SDO Data Across Classes

		Baseline	Intervention		
		M (SD)	Phase 1 M (SD)	Phase 2 M (SD)	
Ms. S – Period 5					
DBR-SM	Prepared.	7.9 (2.03)	7.6 (1.95)	8.8 (1.33)	↑
	Engagement	6.4 (2.80)	6.8 (2.31)	8.0 (1.71)	
SDO	Engagement	36.2 (12.51)	79.0 (5.08)	83.1 (.34)	↑↓
	Off-Task	70.4 (7.60)	30.7 (6.30)	21.7 (8.16)	
Ms. B – Period 3					
DBR-SM	Prepared.	9.6 (1.05)	9.9 (0.48)	9.9 (0.24)	↔
	Engagement	8.6 (1.36)	9.3 (0.99)	9.6 (0.76)	
SDO	Engagement	75.9 (5.68)	86.7 (2.36)	86.7 (5.87)	↑↓
	Off-Task	34.7 (4.58)	19.2 (5.53)	16.7 (6.41)	
Ms. S – Period 1					
DBR-SM	Prepared.	8.1 (1.90)	8.3 (1.35)	8.9 (0.92)	↑
	Engagement	7.4 (2.02)	7.8 (1.59)	8.1 (1.35)	
SDO	Engagement	57.9 (7.75)	71.0 (13.86)	80.6 (14.94)	↑↓
	Off-Task	47.5 (5.00)	34.6 (20.78)	28.9 (14.18)	

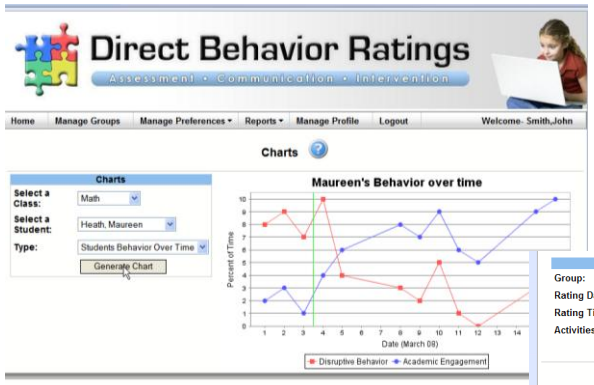
CBER

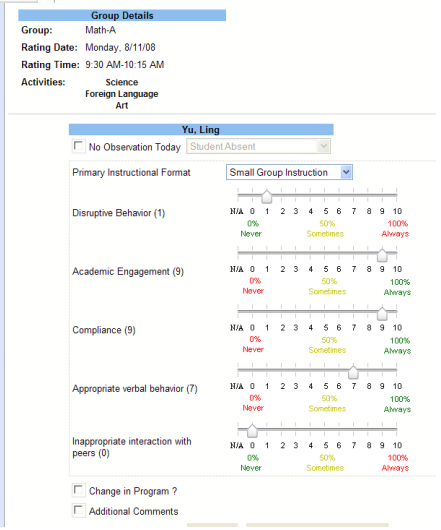
Efficiency of Repeated Measurement



DBR – BASIS

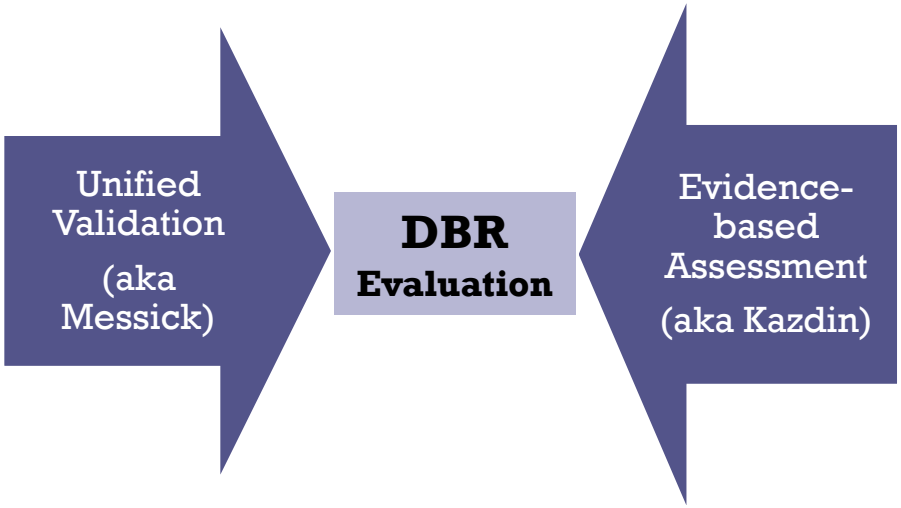
A web-based application will serve to increase utility of the DBR in behavioral assessment given ease of data entry, analysis, and presentation.





CBER

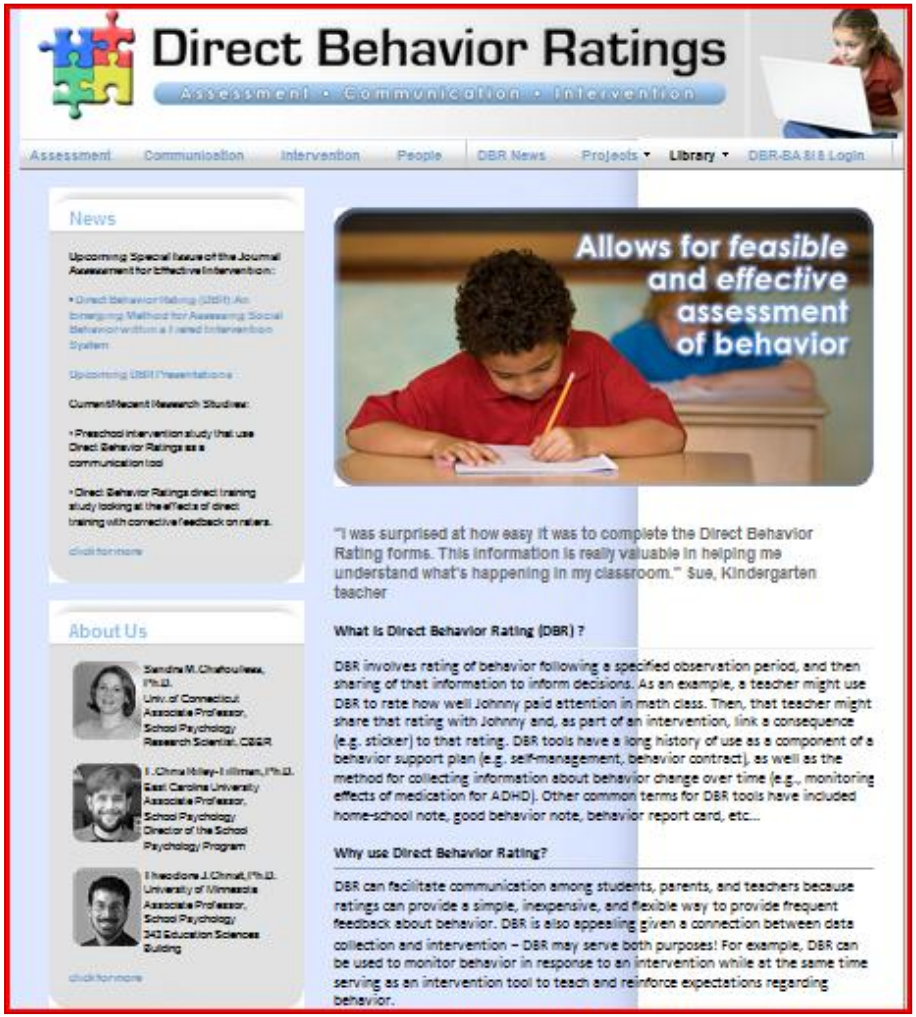
Data Use and Interpretation:
Schools, Teachers, Students at Scale



- Do teachers interpret and apply DBR data as intended?
- How do teachers perceive utility of the DBR method for different purposes?
- How does the use of DBR impact teacher problem-solving behavior about students?

CBER

Questions & Comments…



Contact: Dr. Sandra Chafouleas
sandra.chafouleas@uconn.edu
www.directbehaviorrating.org