



Direct Behavior Ratings as Screeners of Social Behavior Risk

Presentation at NASP 2012

Dr. Stephen P. Kilgus, *East Carolina University*

Dr. Sandra M. Chafouleas, *University of Connecticut*

Ms. Rose Jaffery, M.A., *University of Connecticut*

Dr. Chris Riley-Tillman, *University of Missouri-Columbia*

Dr. Megan E. Welsh, *University of Connecticut*

+ Purpose of the Presentation

- To discuss desirable characteristics of assessment methods in multi-tiered frameworks.
- To review directions in behavior assessment within multi-tiered frameworks.
- To present results from a line of studies examining the utility of Direct Behavior Rating Single Item Scales (DBR-SIS) as behavior screeners.
- To review implications for use of DBR-SIS in early identification efforts.





Multi-Tiered Frameworks (MTF)



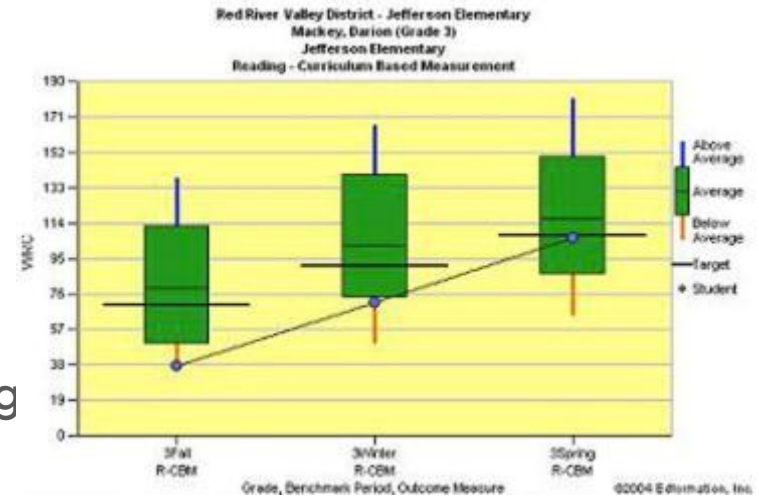
- Adoption of multi-tiered frameworks (e.g., RtI, PBIS) as the basis of academic and behavior service delivery models is on the rise.
- The National Center on Response to Intervention (rti4success.org) emphasizes two main assessment purposes:
 - Progress monitoring
 - Universal screening
- Importance of both is echoed by both IDEA (2004) and NCLB (2001), and numerous professional organizations (e.g., NASP).





Academic MTF – Measures

- Meta-analyses support the relationship between **curriculum-based measurement** (CBM) and key state and national outcome measures (Reschly, Busch, Betts, Deno, & Long, 2009; Yeo, 2010).
- Supported for use across both assessment purposes (Deno, 2005).
- Procedures associated with CBM result in efficient screening and progress monitoring
- Use of a common measure across both purposes yields an overall efficient problem-solving process (Shapiro, 2004; Shinn, 2010).





Behavior MTF – Measures



Progress monitoring (Gresham, 2010)

- Systematic direct observation (SDO)
- Brief behavior rating scales (BBRS)
- Direct Behavior Ratings (DBR)

Universal screening (Severson, Walker, Hope-Doolittle, Kratochwill, & Gresham, 2007)

- *Systematic Screening for Behavior Disorders* (Walker & Severson, 1990)
- *Behavioral and Emotional Screening System* (Kamphaus & Reynolds, 2004)
- *Strengths and Difficulties Questionnaire* (Goodman, 1997)
- *Student Risk Screening Scale* (Drummond, 1994)



Multi-Purpose Behavior Measures



- There is no single behavior measure that may be used across both purposes.
 - That is, there is no CBM analogue for behavior (Chafouleas, Riley-Tillman, & Christ, 2009; Gresham et al., 2010).
- May be related to limited adoption of universal screening for non-academic risk in school settings (Romer & McIntosh, 2005).
- A review of existing approaches supports consideration of defensible progress monitoring tools as potential behavioral *general outcomes measures* (GOMs).

+ Glover & Albers, 2007



Evaluating fit...

- Usability
- Contextual relevance
- Technical adequacy

+ Usability

DBR

Feasibility of administration

Balance of costs and benefits*

Infrastructure

Acceptability

Utility of outcomes

BBRS

Feasibility of administration

Balance of costs and benefits

Infrastructure

Acceptability*

SDO

Utility of outcomes*



Contextual Relevance & Technical Adequacy



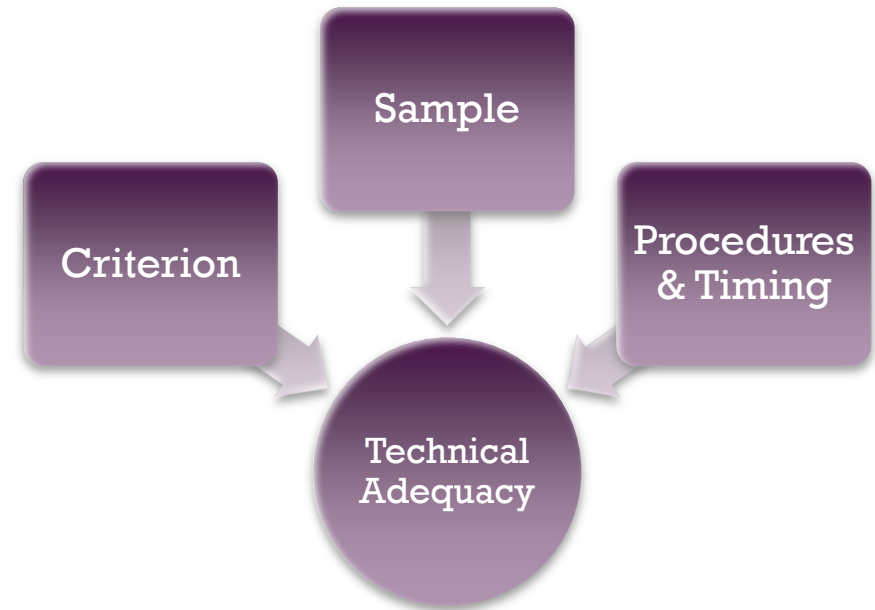
- Contextual Relevance
 - Compatibility with service delivery needs; Alignment with constructs of interest; Theoretical support; Population fit
- Technical Adequacy (or construct validity)
 - Reliability; Criterion-related validity; Content validity; Adequacy of norms
 - Criterion-related validity as demonstrated by:
 - Correlations between predictor and criterion measures
 - Evidence of diagnostic accuracy (*sine qua non*; Jenkins, Hudson, & Johnson, 2007)



Contextual Relevance & Technical Adequacy



- SDO, BBRS, and DBR each possess:
 - Alignment with constructs of interest
 - Theoretical support
- Evidence needed across...
 - Compatibility with local service delivery needs (e.g., risk stratification; Shinn, 2010).
 - Population fit
- No screening technical adequacy evidence available to date
 - Should consider moderators (Cook, 2007).
 - Address the remaining areas of contextual relevance.



+ Statement of the Problem

- Few schools are conducting screening.
- Adoption may be enhanced through the development and validation of more efficient measures and procedures.
- One way to enhance efficiency may be through use of multi-purpose measures.
- Research is necessary to examine the tenability of using existing progress monitoring tools as screeners...

+ DBR as Screeners



- Initial correlational research justified further examination of Direct Behavior Rating – Single Item Scales (DBR-SIS) in screening applications (Chafouleas, Kilgus, & Hernandez, 2009).
- A series of studies have since examined their use in screening through the collection of concurrent validity evidence, including correlations and diagnostic accuracy.



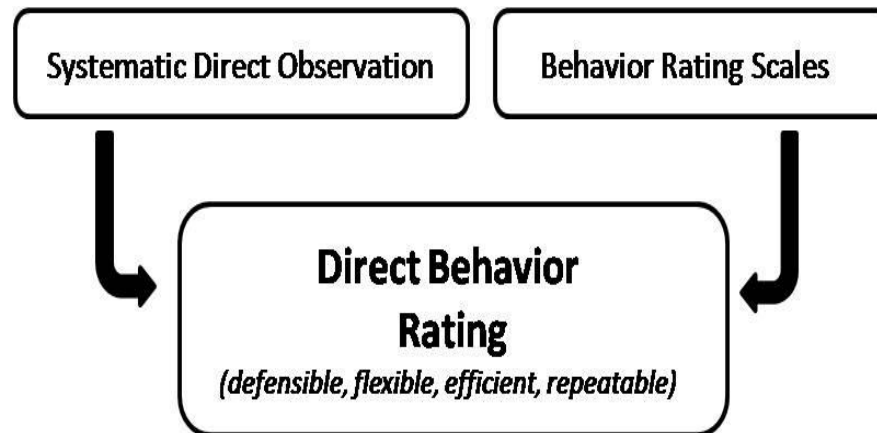
What is
Direct
Behavior
Rating?



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

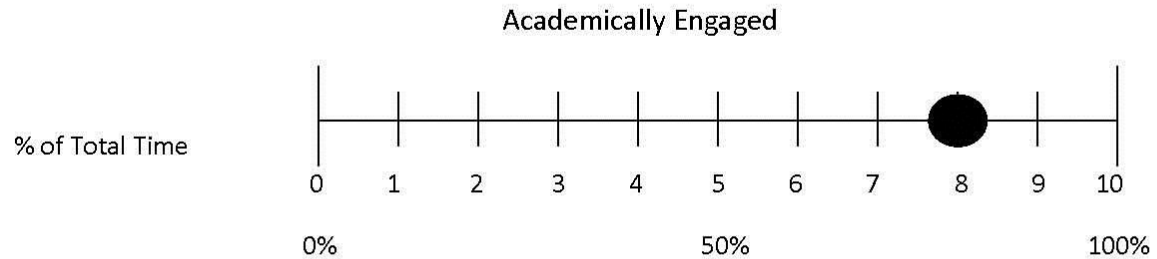


Chafouleas, Riley-Tillman, & Christ (2009); Chafouleas, Riley-Tillman, & Sugai (2007); Chafouleas, Riley-Tillman, & McDougal (2002); Christ, Riley-Tillman, & Chafouleas (2009)

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		

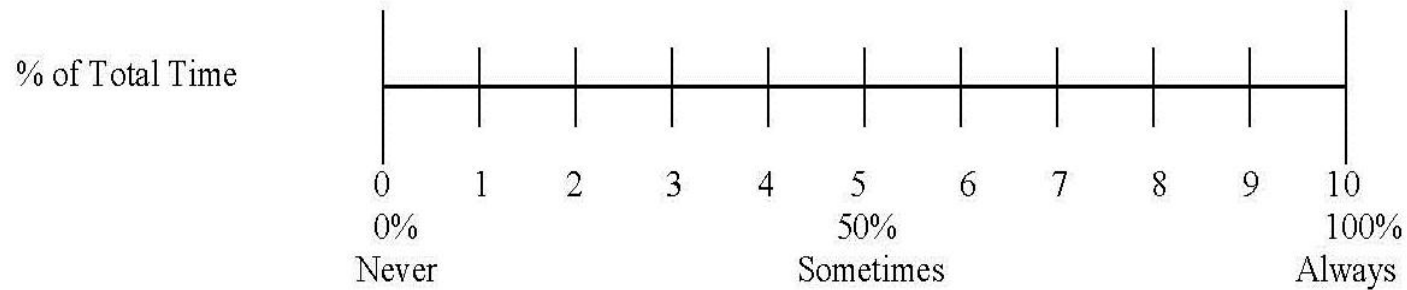
Interpretation: The student earned 84% (5/6) of possible points during the observation period.



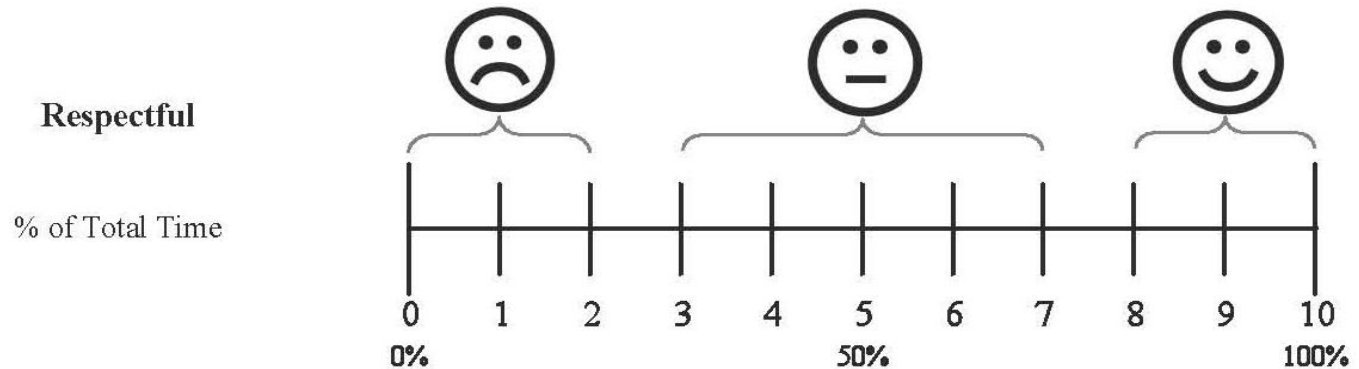
Our DBR-SIS Scale

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.

Academically Engaged



Respectful

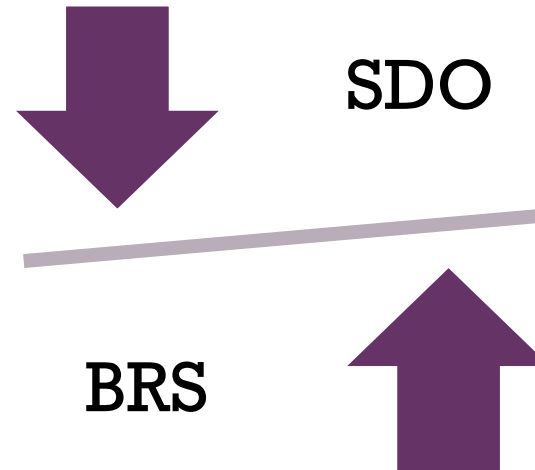


+ A little background...

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

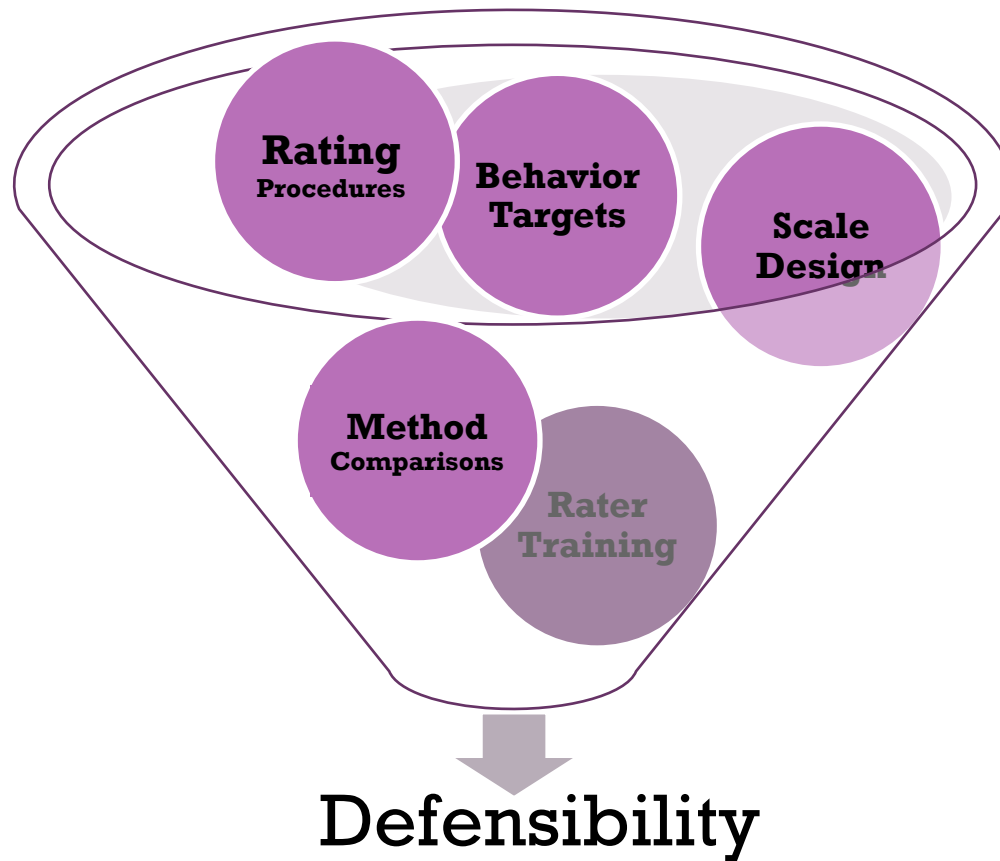
Contemporary Defining Features:



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)

+ Project VIABLE (2006-2011)

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



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



† DBR-SIS Targets: “The Big 3” General Outcomes



Academic Engagement:

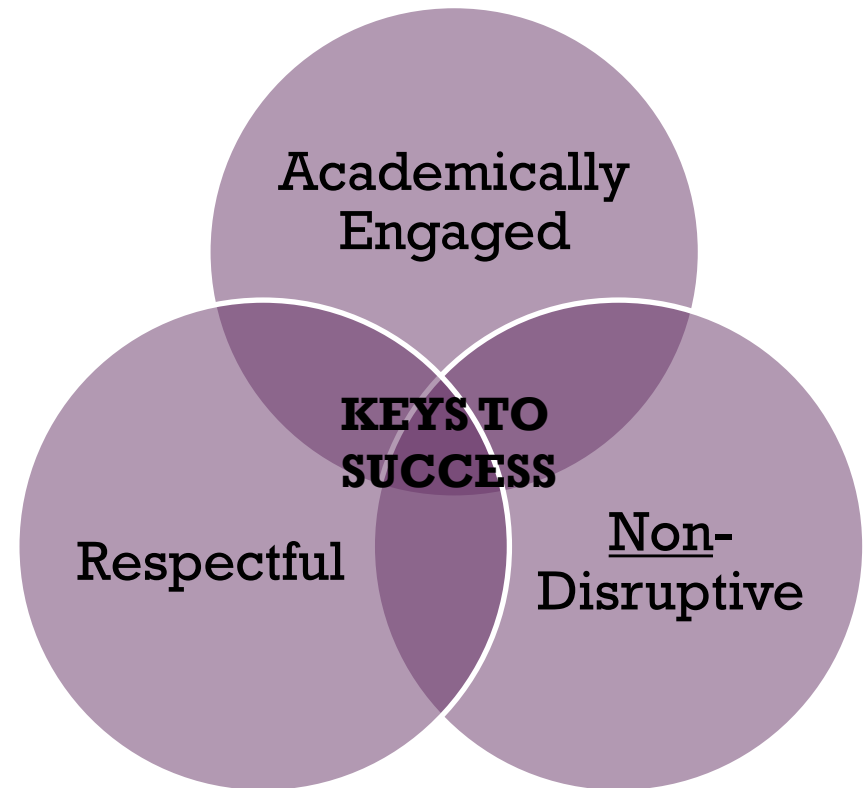
Actively or passively participating in the classroom activity.

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.





Purposes for THIS Presentation on DBR-SIS in Screening



Broad Question:

Does screening research support use of DBR-SIS, a defensible progress monitoring tool, as a **GOM** within multi-tier frameworks?

Specific Purposes:

Receiver operating characteristic (ROC) curve analysis

1. Examine overall diagnostic accuracy associated with each DBR-SIS target
2. Identify cut scores for determining risk for behavior problems
3. Examine suitability of single targets relative to multiple gating procedures

Bivariate correlations

4. Evaluate concurrent validity of DBR-SIS score-based inferences as they pertain to behavioral difficulty



1. Conditional Probabilities

	BESS	
DBR-SIS	At-Risk	Not At-Risk
Abnormal Score	A	B
Normal Score	C	D

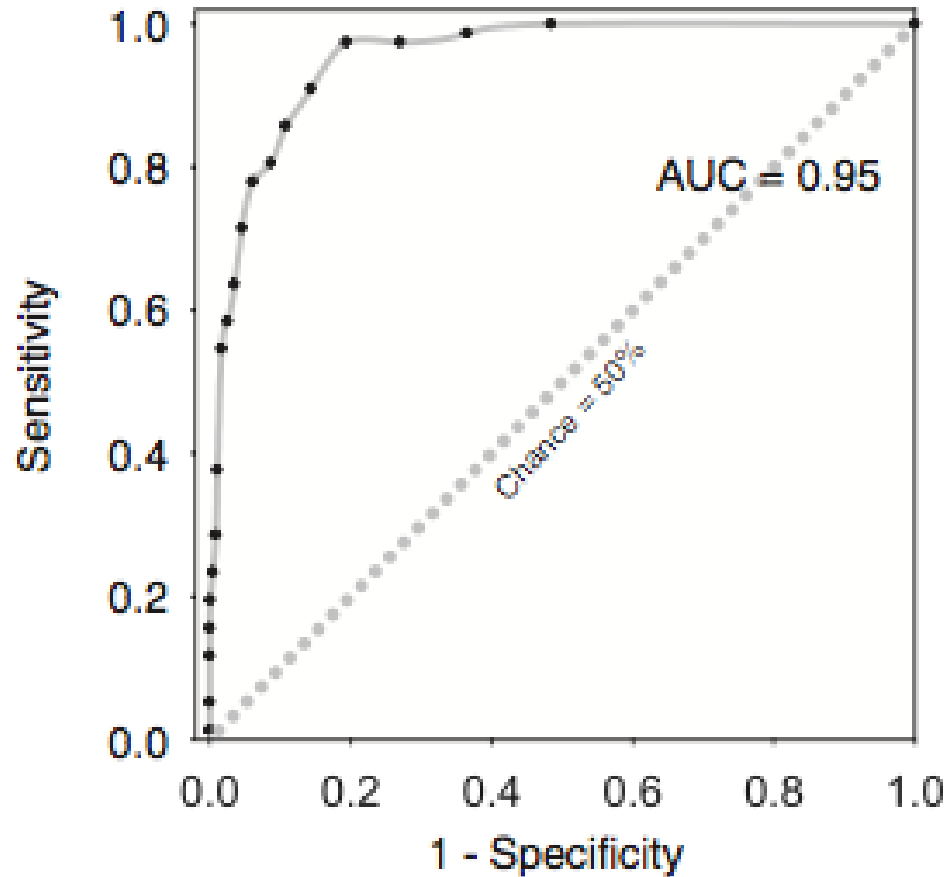
$$\text{SN} = A/(A+C)$$

$$\text{SP} = D/(B+D)$$

$$\text{PPP} = A/(A+B)$$

$$\text{NPP} = D/(D+C)$$

+ 2. Area Under the Curve (AUC)

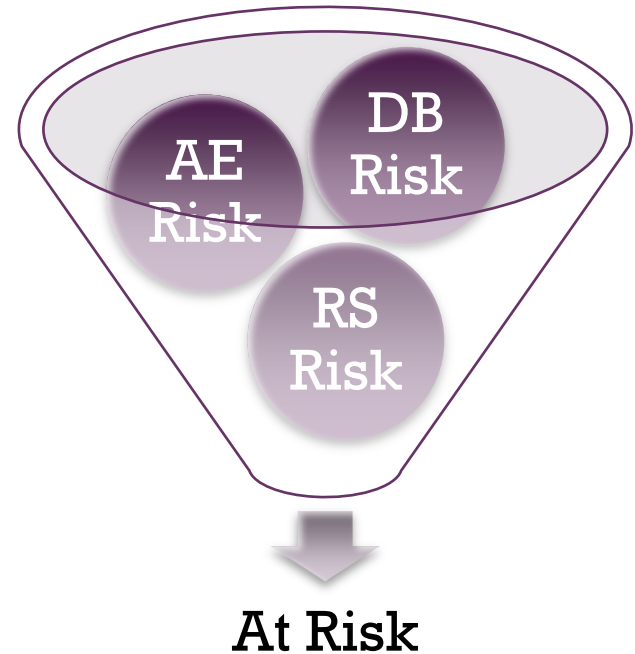




3. Multiple Gating Procedure



- A student must be at risk on all targets of interest if he/she is to be considered at risk overall.
- Means by which to increase the specificity of a screening process.
- Must be careful, as increasing the stringency of the decision criteria may also decrease sensitivity.



+ Study 1



Kilgus, S. P., Chafouleas, S. M., Riley-Tillman, T. C., & Welsh, M. (2012). Direct Behavior Rating scales as screeners: A preliminary investigation of diagnostic accuracy in elementary school. *School Psychology Quarterly*, doi:10.1037/a0027150

- Second-grade sample - 118 students and 12 teachers
- Base rate = 20%

+ Study 1



■ Predictors

- DBR-SIS (DB, AE, & CO)
 - Mean of 5 days * 2 ratings (AM & PM)
- Social Skills Improvement System (SSiS) Performance Screening Guide (Elliott & Gresham, 2007)
 - Prosocial Behavior & Motivation to Learn
 - Once following DBR-SIS

■ Criterion

- BASC-2 Behavioral and Emotional Screening System (BESS; Kamphaus & Reynolds, 2007)
 - Once following DBR-SIS

+ Conditional Probabilities

Score	SN	SP	PPP	NPP	BR	CC
DB (2)	.91	.67	.40	.97	44.07	.72
DB (3)	.78	.85	.56	.94	27.12	.84
AE (7)	.87	.81	.53	.96	32.20	.82
AE (8)	1.00	.66	.42	1.00	46.61	.73
CO (8)	.78	.82	.51	.94	29.66	.88
CO (9)	.96	.60	.37	.98	50.85	.67
PB (3)	.83	.76	.47	.95	35.59	.79
ML (3)	.91	.79	.51	.97	34.75	.82
DB+AE	.91	.80	.53	.97	33.90	.82
DB+CO	.91	.75	.47	.97	38.14	.78
AE+CO	.96	.78	.51	.99	36.44	.81
PB+ ML	.78	.85	.56	.94	27.12	.84
DB+AE+CO	.91	.83	.57	.98	31.36	.85

+ Area Under the Curve



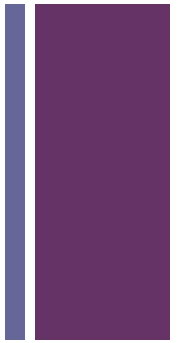
	AUC	<i>p</i>	CI-95
DB	0.87	<.001	0.79-0.92
AE	0.89	<.001	0.82-0.94
CO	0.87	<.001	0.78-0.93
SSiS-ML	0.91	<.001	0.85-0.96
SSiS-PB	0.85	<.001	0.78-0.91

- Statistically significant differences between each scale and random decision making.
- No statistically significant AUC differences between scales

+ Study 2

Chafouleas, S. M., Kilgus, S. P., Jaffery, R., Riley-Tillman, T. C., & Welsh, M. E. (2012). Diagnostic accuracy of direct behavior rating as a behavior screener for elementary and middle school students. *Manuscript under review.*

- Investigation 1
 - Elementary; 617 students & 44 teachers
 - Base rate = 15%
- Investigation 2
 - Middle; 214 students & 14 teachers
 - Base rate = 9%



+ Study 2

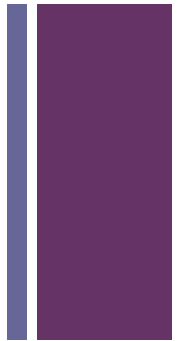


■ Predictors

- DBR-SIS (DB, AE, & RS)
 - Mean of 5 days * 2 ratings
 - Investigation 1 – AM & PM
 - Investigation 2 – First and second half of one period
- Student Risk Screening Scale (SRSS; Drummond, 1994)
 - Once following DBR-SIS

■ Criterion

- BESS
 - Once following DBR-SIS



Group	Scale	SN	SP	PPP	NPP	Base Rate
Early Elementary	DB (2)	.80	.79	.42	.95	30.77
	DB (3)	.58	.90	.54	.92	17.31
K-2	AE (7)	.64	.85	.44	.93	23.08
	AE (8)	.76	.67	.30	.94	40.06
	RS (9)	.64	.82	.40	.92	25.64
	SRSS (4)	.98	.66	.36	.99	44.23
	DB + AE	.66	.82	.42	.93	25.32
	DB + RS	.60	.87	.46	.92	20.83
	AE + RS	.58	.84	.41	.91	22.76
	DB + AE+ RS	.56	.87	.46	.91	19.55
	Late Elementary	DB (1)	.90	.61	.26	.98
DB (2)		.63	.91	.53	.94	16.07
AE (7)		.68	.93	.60	.95	15.41
AE (8)		.88	.73	.34	.97	35.08
RS (9)		.51	.91	.47	.92	14.75
SRSS (4)		.83	.82	.41	.97	26.89
DB + AE		.83	.81	.40	.97	27.54
DB + RS		.51	.94	.57	.93	12.13
AE + RS		.51	.95	.64	.93	10.82
DB + AE+ RS	.51	.95	.64	.93	10.82	
Middle	DB (1)	.58	.83	.25	.95	20.56
	AE (8)	.63	.85	.29	.96	19.16
	AE (9)	.89	.47	.14	.98	56.07
	RS (9)	.47	.90	.32	.95	13.08
	SRSS (4)	.79	.84	.33	.98	21.50
	DB + AE	.58	.83	.25	.95	20.56
	DB + RS	.32	.95	.40	.93	7.01
	AE + RS	.42	.91	.32	.94	11.68
	DB + AE+ RS	.32	.95	.40	.93	7.01



Scale	Early Elementary			Late Elementary			Middle		
	AUC ¹	<i>p</i>	95% CI	AUC ²	<i>p</i>	95% CI	AUC ³	<i>p</i>	95% CI
DB	.823	<.001	.78-.86	.857	<.001	.81-.89	.707	.001	.64-.77
AE	.810	<.001	.76-.85	.893	<.001	.85-.93	.780	<.001	.72-.83
RS	.740	<.001	.69-.79	.720	<.001	.67-.77	.692	.002	.63-.75
SRSS	.928	<.001	.89-.95	.896	<.001	.86-.93	.848	<.001	.79-.89

- Holm-Bonferroni multiple comparisons
- Statistically significant AUC differences between:
 - RS scale and all other measures in Late Elementary
 - SRSS and all other scales in Early Elementary

+ Study 3

Chafouleas, S. M., Riley-Tillman, T. C., Christ, T. J., & Kilgus, S. P. (2012). *Examining Direct Behavior Rating Single Item Scale diagnostic accuracy and concurrent validity in elementary and middle school settings: Replication across sites*. Manuscript in preparation.

	Northeast	Southeast	Midwest
1 st	116 (8)	181 (14)	114 (9)
4 th	122 (9)	137 (10)	96 (6)
7 th	155 (11)	70 (5)	119 (7)

+ Study 3



■ Predictors

- DBR-SIS (DB, AE, & RS)
 - Mean of 5 days * 2 ratings
 - Investigation 1 - AM & PM
 - Investigation 2 – First and second half of one period
- SRSS
 - Once following DBR-SIS

■ Criterion

- BESS
 - Once following DBR-SIS

Grade	Scale (cut score)	SN	SP	PPP	NPP
1 st	DB (1)	.90	.52	.25	.97
	DB (2)	.79	.82	.43	.96
	AE (7)	.71	.81	.39	.94
	AE (8)	.90	.58	.28	.97
	RS (9)	.69	.79	.37	.94
	SRSS (5)	.92	.73	.38	.98
	DB+AE	.84	.70	.33	.96
	DB+RS	.68	.81	.39	.93
	AE+RS	.66	.83	.40	.93
	DB+AE+RS	.66	.83	.41	.93
4 th	DB (1)	.74	.61	.33	.90
	AE (8)	.84	.71	.43	.94
	RS (9)	.53	.80	.41	.87
	SRSS (4)	.93	.70	.45	.98
	DB+AE	.66	.80	.47	.90
	DB+RS	.53	.83	.44	.87
	AE+RS	.51	.88	.53	.87
	DB+AE+RS	.51	.88	.54	.87
7 th	DB (1)	.87	.65	.37	.95
	DB (2)	.66	.86	.54	.91
	AE (8)	.87	.73	.44	.96
	RS (9)	.69	.84	.52	.92
	SRSS (5)	.90	.80	.52	.97
	DB+AE	.82	.79	.49	.95
	DB+RS	.66	.88	.56	.91
	AE+RS	.63	.90	.60	.91
DB+AE+RS	.63	.90	.61	.91	



Scale	1 st Grade			4 th Grade			7 th Grade		
	AUC ¹	CI-95	<i>p</i>	AUC ²	CI-95	<i>p</i>	AUC ³	CI-95	<i>p</i>
DB	.85	.81-.88	<.001	.74	.69-.78	<.001	.83	.79-.87	<.001
AE	.84	.80-.88	<.001	.84	.80-.88	<.001	.87	.83-.91	<.001
RS	.77	.73-.81	<.001	.68	.63-.73	<.001	.79	.74-.83	<.001
SRSS	.90	.87-.93	<.001	.92	.88-.94	<.001	.92	.88-.94	<.001

- Differences in diagnostic accuracy between DBR-SIS targets were predominantly non-statistically significant.
- Across all grades, the SRSS consistently outperformed both DB and RS, but not AE, to a statistically significant degree

+ Summary Across Studies

- To date, accumulated evidence provides initial support for the use of DBR-SIS data to inform screening decisions.
- Overall DBR-SIS diagnostic accuracy was consistently in the moderate range.
 - AE performed consistently well, particularly in advanced grades
 - DB performed well in lower grades. Performance in advanced grades varied.
 - CO/RS consistently underperformed
- Concurrent validity supported through moderate to high correlations between each DBR scale and the SSiS, SRSS, and BESS.
- Comparison measures:
 - SSiS Performance Screening Guide performed similarly to all targets
 - SRSS performed as well if not better than DB and RS. Performed similarly to AE.

+ Summary



- Each DBR-SIS target's optimal cut score varied by grade.
- Two classes of cut score were commonly identified
 - High SN/NPP, moderate SP/PPP
 - Higher SP/PPP, moderate SN/NPP (higher correct classification)
- DBR-SIS multiple gating procedure occasionally yielded a more balanced approach with greater correct classification.



Implications & Future Directions

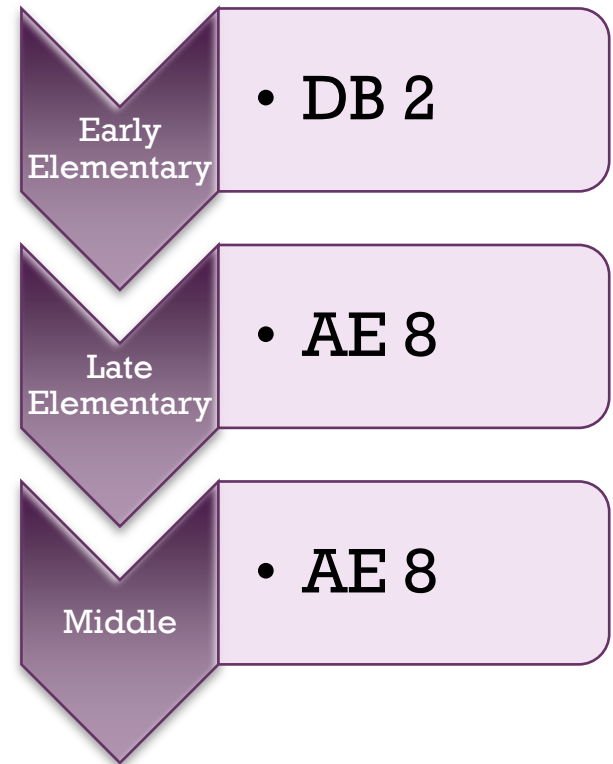


Research:

- Current promising findings support need for further research

Practice:

- Through additional research, there is promise for use of DBR-SIS in formative risk evaluation.
- There is need for time-specific cut scores.
- The best DBR-SIS target is likely to change over time.
- Multiple gating procedure is inconsistent
- Intensity of DBR-SIS procedures may not support its use at the universal level (that is, ALL students)... may need to combine with gating procedure to reduce total number to evaluate





Questions, Comments, Contacts...

Dr. Stephen Kilgus, *kilguss@ecu.edu*

Dr. Sandra Chafouleas, *sandra.chafouleas@uconn.edu*

Dr. Chris Riley-Tillman, *rileytillmant@missouri.edu*