



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

REVIEW OF METHODS FOR FORMATIVE ASSESSMENT OF CHILD SOCIAL BEHAVIOR

Chair: Sandra M. Chafouleas (U of Connecticut)

Amy M. Briesch (Northeastern U)

T. Chris Riley-Tillman (East Carolina U)

Theodore J. Christ (U of Minnesota)

Robert J. Volpe (Northeastern U)

Discussant: Frank Gresham (Louisiana State U)

Date: August 9, 2009

American Psychological Association



Purpose

- To provide a critical review of social behavior assessment methods that include
 - systematic direct observation, behavior rating scales, direct behavior rating, and permanent products
- To introduce current knowledge about each method with regard to **formative assessment**
 - content framed around the characteristics of *defensibility*, *flexibility*, *efficiency*, and *repeatability*





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Statement of the Problem

Evolving trend away from pathology and toward *promotion* of mental health means increased psychologist role in assessment focused on early identification and monitoring of key behavioral indices

What constitutes evidence-based assessment?





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Statement of the Problem

Commentary by Kazdin (2005):

Developing evidence-based assessment (EBA) begins through *a priori* delineation of

- a) the purposes of assessment, and then
- b) identification of the special requirements for each purpose (and associated criteria for stating when requirement is met)



Statement of the Problem

A Few Caveats to Establishing EBA...

- 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)



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Statement of the Problem

What is ~~THE~~ measure I should use?





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Statement of the Problem

Evolving trend away from pathology and toward *promotion* of mental health means increased psychologist role in assessment focused on early identification and monitoring of key behavioral indices



**Formative
Assessment**



Statement of the Problem

Why is formative assessment important?

- data streams are needed to make on-going decisions about treatment effectiveness
 - Even with implementation of EBI, you need ideographic analysis of effects for a particular individual

What are the critical features (requirements) of formative assessment tools?

- *defensibility, flexibility, efficiency, and repeatability*



Definitions

- **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 (while maintaining above characteristics).



What is ~~THE~~ measure I should use?

Methods

- 1) Systematic Direct Observation
(Briesch & Riley-Tillman)
- 2) Behavior Rating Scales (Volpe)
- 3) Direct Behavior Rating (Christ)
- 4) Permanent Product (Chafouleas)

Discussion & Questions (Gresham)





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

REVIEW OF SYSTEMATIC DIRECT OBSERVATION

Amy M. Briesch, Ph.D.

Department of Counseling and Applied Psychology, Northeastern University

T. Chris Riley-Tillman, Ph.D.

Department of Psychology, East Carolina University





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Systematic direct observation

- Often synonymous with behavioral assessment (Reynolds & Kamphaus, 2003)
- Revered due to direct nature and opportunity to assess behavior-environment interactions
- Limited research evidence with regard to questions of defensibility, flexibility, efficiency, and repeatability





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

SDO – Practice Issues

- Practitioners need to actually use SDO
 - Acceptability
 - Use
 - Flexibility
 - Feasibility



SDO –Acceptability

- Studies show high levels of acceptability (Riley-Tillman et. al., 2008)
 - Specifically, on a 6-point scale (1 = strongly disagree, 6 = strongly agree)
 - 5.1 - This technique is an acceptable strategy to assess intervention effectiveness for this child's problem
 - 4.9 - Overall, using this technique would be beneficial for the child
 - 4.8 - This technique provides a feasible method of assessing the effectiveness of an intervention



SDO Use

- Studies suggest moderate to high levels of reported use
 - 67% of school psychologist report using direct observation in 4 of their last 10 case Shapiro & Heick (2004), 63% to 73% of School Psychologist report moderate to frequent use (Riley-Tillman et. al, 2008).
 - Are practitioners actually using a formal SDO system with integrity – or are they doing naturalistic observation and calling it SDO?





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

SDO Use

- No research which measures actual use of formal SDO
- No research which measures the integrity of usage by practitioners
 - Considering that IOA data is demanded for publication, it is interesting that such a practice is essentially unheard of in applied settings





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

SDO – Needed Future Research

- Use
 - What is the actual use in terms of rate and integrity in applied settings?
 - What are the impacts on the natural environment of SDO use?
 - For example, what are the reactivity impacts, and at what rate do they typically dissipate?





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

SDO - Flexibility

- One of the significant strengths of SDO is the flexibility of instrumentation
 - Behavioral definitions can be altered
 - The setting can be selected to maximize the likelihood of observing the target behavior
- There is a lack of research on the impact (e.g reduced accuracy) of altering SDO instrumentation



The logo features a stylized illustration of the CN Tower on the left side of the slide. The tower is depicted with a blue and white color scheme, showing its distinctive structure including the observation deck and the spire. It is positioned vertically, extending from the bottom left towards the top left of the slide.

TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

SDO – Needed Future Research

- Flexibility
 - What are the implications of changing core features of SDO within a method?
 - For example, using Momentary Time Sampling, for 15 minutes, are the psychometric qualities of SDO consistent across a range of target behaviors





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

SDO - Feasibility

- A single SDO is rather feasible – 10-15min.
- Feasibility though decreases as observation numbers increase
 - Assuming a min number of observations (5), this balloons to 50-75 minutes of observation with additional entry/exit time.
 - Over 100 cases (a rather typical school psychologist yearly load), this is 5,000 – 7,500 minutes, or 83 – 125 hours.





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

SDO – Needed Future Research

- Feasibility
 - What are the minimum number of observations needed for defensible decisions?





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Defensibility

- The BIG question...
- Interobserver agreement = “the bedrock upon which sound behavioral measurement rests”
(Watkins & Pacheco, 2000, p. 206)
- Reports of interobserver reliability generally high for published systems

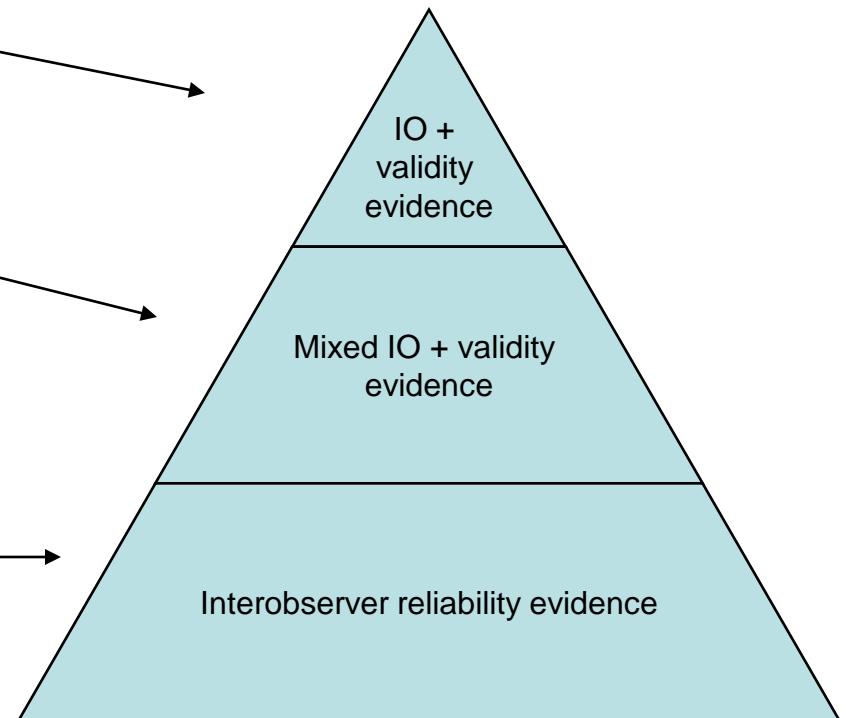


Volpe et al. (2005)

- *Academic Engaged Time Code of the SSBD (2)*
- *Direct Observation Form (3-6)*

- *ADHD School Observation Code*

- *Classroom Observation Code*
- *Behavioral Observation of Students in Schools*
- *State-Event Classroom Observation System*





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Issues

- Few systems have established reliability/validity evidence in multiple samples
- Uncertainty re: use of formal coding systems
- Reliability \neq Validity





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

What type of accuracy is most important?

- Johnston and Pennypacker's (1980)
observational accuracy
 - Collect sufficient behavioral samples to be “sufficiently representative” of times and behaviors targeted





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Marcus (1980)

- *Cooperative play* of 31 preschool students
- Momentary time-sampling with 10-second interval
- At least 10 observations needed to achieve adequate ($r = .81$) reliability



McWilliam & Ware (1994)

- 47 1-4 year-old children in day care setting
- 4 free play behaviors (*engagement with adults, engagement with peers, engagement with materials, nonengaged*)
- Momentary time-sampling with 10-second interval
- Between 12 (*engagement with peers*) and 40 (*engagement with materials*) 15-minute observation sessions necessary



Doll & Elliott (2004)

- 24 preschool children
- 20-minute observation using 15-second partial-interval recording
- 13 mutually exclusive social behaviors (e.g., *share request, complimentary statement*)
- Need at least 5 10-minute observations to adequately represent behavior



Hintze & Matthews (2004)

- *Active/passive engagement* of 14 5th grade students
- Momentary time-sampling w/ 15-second intervals
- Typical practice (3 15-min observations) = low reliability (.25)
- Four observations per day / 40 days = adequate reliability





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Briesch et al. (in preparation)

- *Academic engagement* of 12 kindergarten students
- Momentary time-sampling w/ 15-second interval
- High (.9) levels of dependability reached given 10 days of data collection, sufficient (.8) levels after 5



Coming to conclusions

- Number of observations has varied significantly
 - Minimum ranging from 5 (Briesch et al., in prep; Doll & Elliott, 2004) to upwards of 40 (Hintze & Matthews, 2004)
 - Issue of different behaviors assessed with varying populations



Considerations for future research

- Focus has been on generalizing across time
 - How representative is 15-minute observation of larger school day?
- Have not studied targeted populations
 - Do reliability/generalizability estimates look different?
- Convergent/discriminant validity evidence needed for systems & less formal codes



A vertical illustration of the CN Tower, showing its distinctive needle-like top and the circular observation deck. The tower is colored in shades of blue and green, with a yellow light at the base.

TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

In sum...

- SDO direct, flexible, and familiar
- Need better understanding of what it looks like in applied practice
- Need to dig deeper with regard to defensibility





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

REVIEW OF ADAPTIVE BEHAVIOR RATING SCALES

Robert J. Volpe, Ph.D.

Department of Counseling and Applied
Psychology, Northeastern University



Northeastern





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Overview

- Types of Rating Scales
- Strengths and Limitations
- The Adaptive Rating Scale Approach
- Preliminary Data
- Future Directions and a preliminary model



Traditional Rating Scales

- **Strengths**

- Wide selection for use across informants
- Relatively broad assessment of key constructs
- Well examined psychometric properties
- Useful for low frequency behaviors

- **Disadvantages**

- Majority not designed for progress monitoring
 - Long, time consuming
 - Instructions typically request ratings over a long time period



Brief Rating Scales

- **Shortened measures***
 - **Conners Short Forms**
 - **BASC Monitor**
 - **ADHD-SC4**
 - **CDI Short Form**
- **Typically derived via factor analysis**
- **Targets of assessment usually psychiatric symptoms as opposed to areas of impairment**

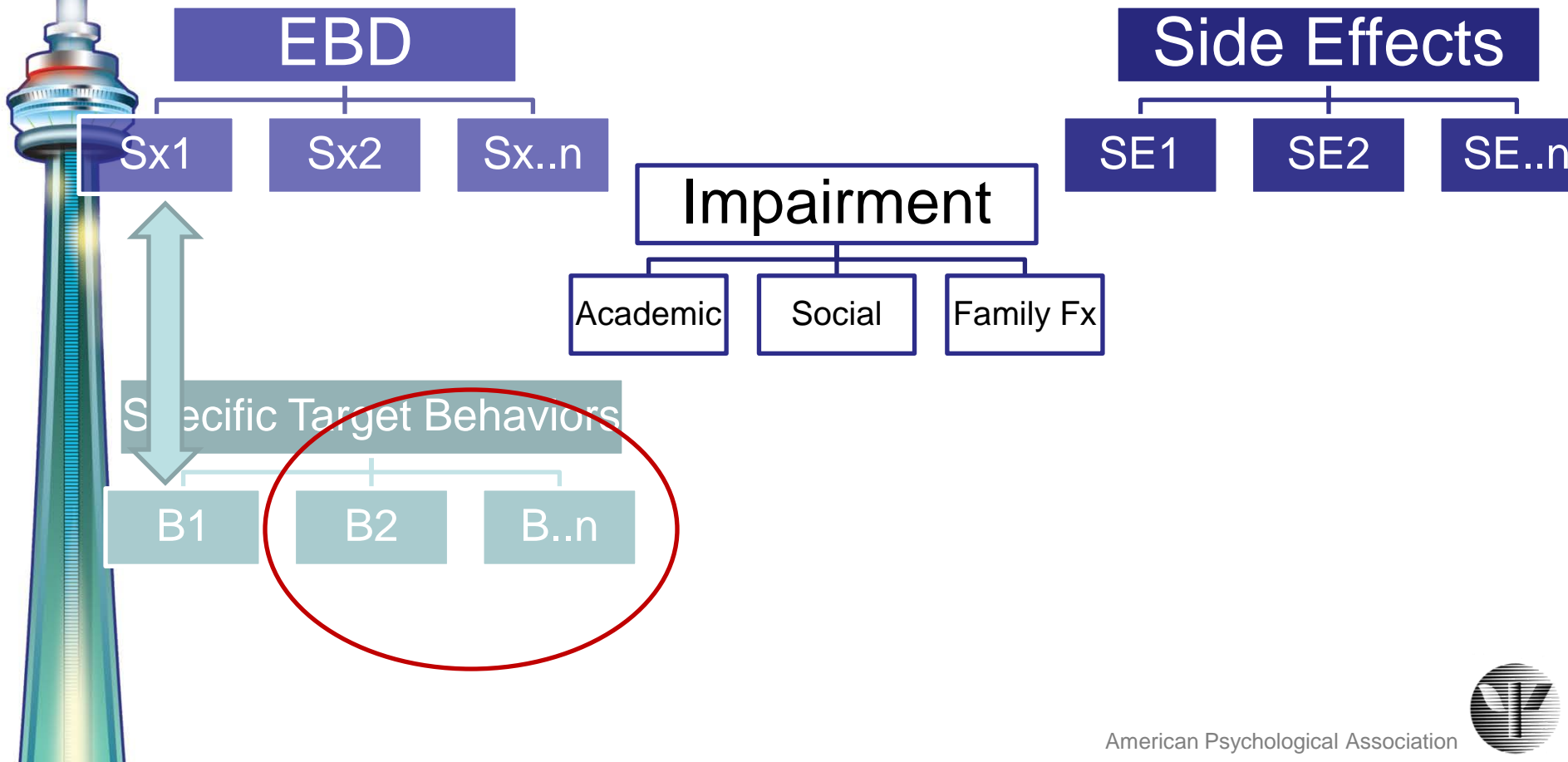


TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Measurement Targets

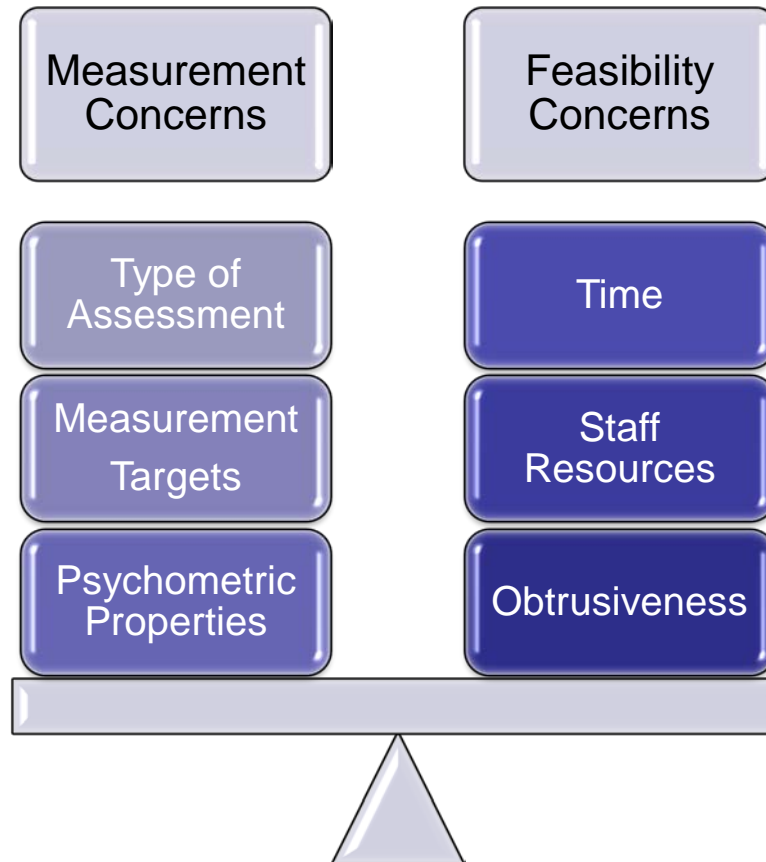


TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Selecting Appropriate Progress Monitoring Measures



Adapted from Briesch & Volpe (2007)



Adaptive Behavior Rating Scales

- Adaptive rating scales are small groups of items drawn from existing scales measuring particular constructs.
- Separate short scales can be generated for each construct of interest.
- The assessment can be tailored to each individual student



Unique Features of ABRS

- **Differ from direct behavior ratings:**
 - Breadth and depth of assessment
 - Confidence in psychometric properties
 - Longer
- **Differ from brief rating scales:**
 - Flexibility
 - Potentially broader assessment





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

AN INITIAL EXAMINATION OF TWO METHODS FOR CREATING ABRS

Factor Analytic Approach
Individualized Approach

Volpe, R. J., Gadow, K. D., Blom-Hoffman, J., & Feinberg, A. B. (2009). Factor analytic and individualized approaches to constructing brief measures of ADHD behaviors. *Journal of Emotional and Behavioral Disorders*, 17, 118-128

American Psychological Association



Participants

- N=29 (24 males, 5 females)
- Between 4 and 17 years of age ($M=10.3$; $SD=3.5$).
- Diagnosed with ADHD by a child psychiatrist:
 - Rating scales completed by parents and teachers
 - CBCL, TRF, CSI-4, IOWA Conners'
 - Laboratory and school observations
 - Clinical interviews
 - Review of school records





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

The ADHD-Symptom Checklist-4

(ADHDSC-4; Gadow & Sprafkin)

The 50-item ADHD-SC4 contains five scales:

- ADHD:Inattentive (IA)
- ADHD:Hyperactive-Impulsive (HI)
- Oppositional Defiant Disorder
- Peer Conflict Scale
- Symptom Side-effects Checklist

Response format:

- never=0, sometimes=1, often=2, very often=3

Reliability and Validity

- Internal consistency for the IA and HI scales is high (.95)
- Good test-retest reliability coefficients (2-week interval) for both IA (.84) and HI (.85).
- Evidence convergent and discriminant validity with Teacher Report Form (TRF; Achenbach, 1991b) and IOWA Conners' Teacher Rating Scale (Loney & Milich, 1982).



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Design

Week 1	Week 2	Week 3	Week 4
Baseline	Low Dose	Med. Dose	High Dose
<div>SC 4</div> <div>SC 4</div>	<div>SC 4</div> <div>SC 4</div>	<div>SC 4</div> <div>SC 4</div>	<div>SC 4</div> <div>SC 4</div>



Two Approaches

- **Factor-derived Approach (FAC)**
 - Traditional/generic approach
 - Select subgroup of items with highest factor loadings
 - Commonly used to create short forms of larger rating scales
- **Individualized Approach (IND)**
 - Behavioral approach
 - Each case has a specific scale based on items that were most problematic during an initial administration of a complete scale or scales.



Hypotheses

- Scores from the FAC, IND, and FULL version of scales will demonstrate similar profiles across dose conditions reflecting the utility of all three approaches.
- Each method will demonstrate adequate psychometric properties, but the full version of scales will prove most favorable in this regard.



Comparing Methods

Mean test-retest coefficients

- FAC = .69
 - IND = .63
 - Full = .67
-
- Differences in mean test-retest coefficients across conditions were larger than differences across methods but were not statistically significant.
 - The largest difference was between average teacher ratings in the baseline (.50) and high-dose (.73) conditions ($z = 1.29, p < .10$).





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Internal Consistency

Scale		Alpha
FAC		
	IA	.72
	HI	.93
IND		
	IA	.93
	HI	.96
Full		
	IA	.87
	HI	.96



Treatment Sensitivity

- All methods demonstrated treatment sensitivity
- Treatment sensitivity overall was comparable across methods
- The only significant difference was between the individualized version and the two other versions (Factor and Full), but only for the dose x method interaction for baseline-low dose.

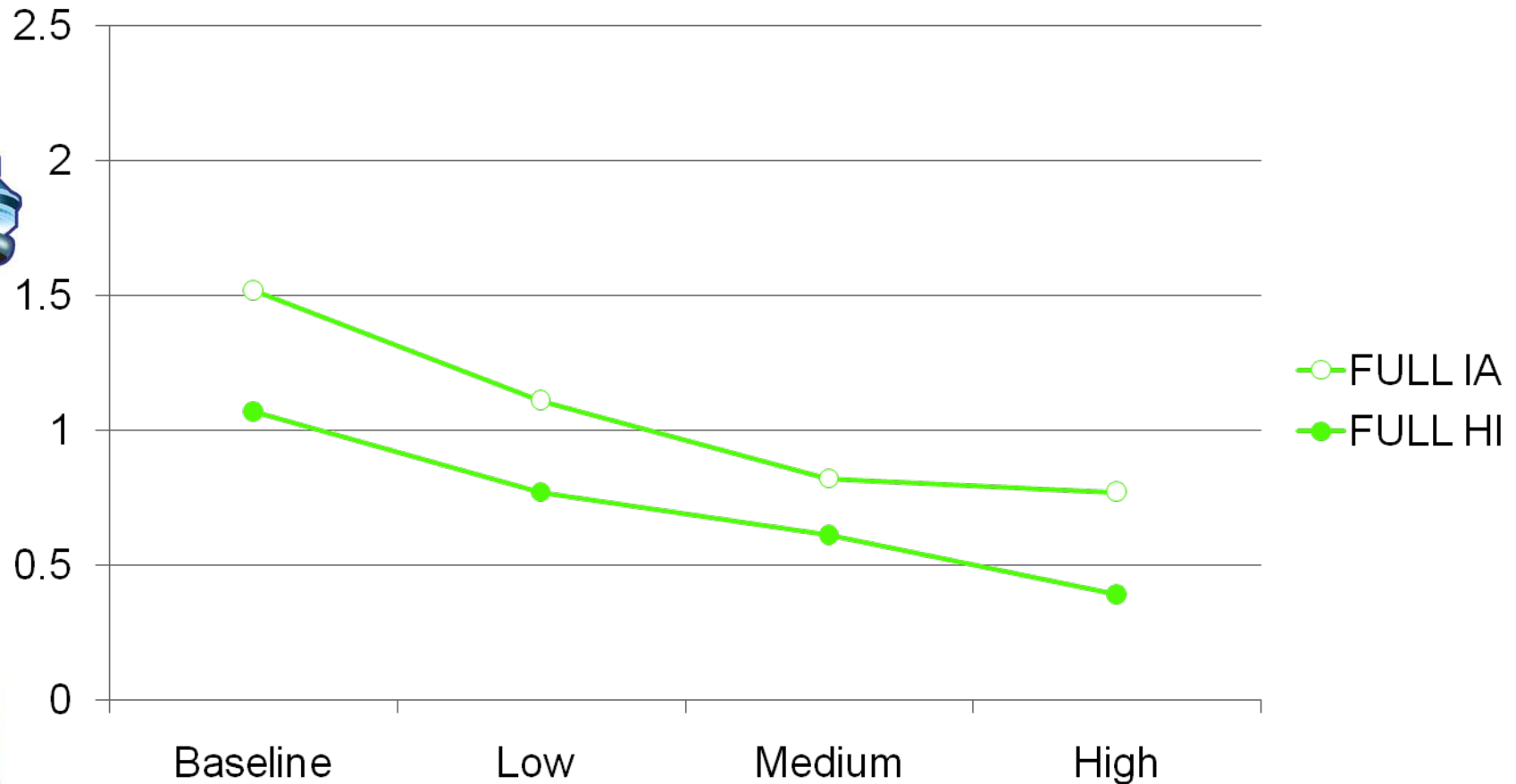


TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

FULL Across Dose Conditions

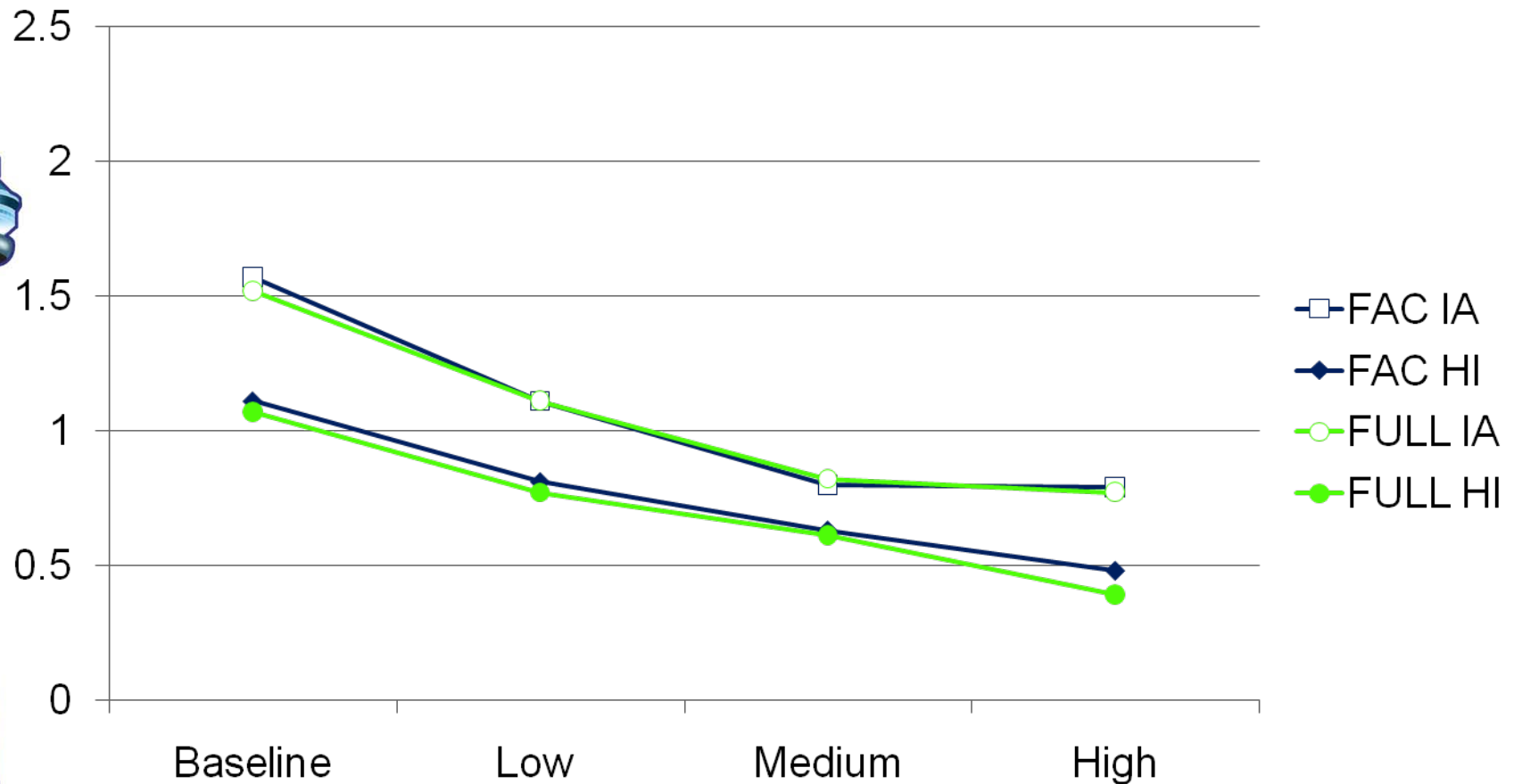


TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

FULL & FAC Across Dose Conditions

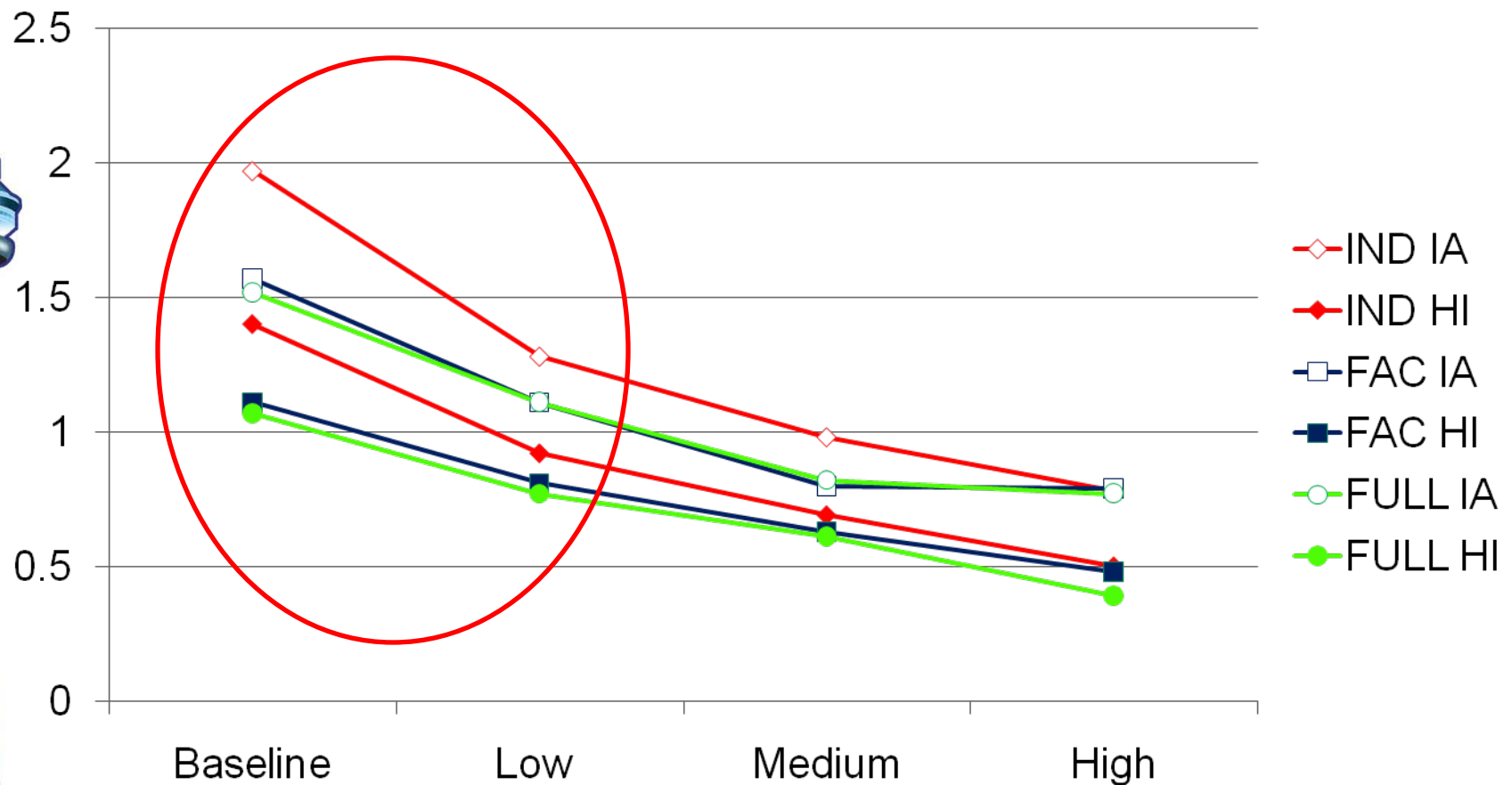


TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

All Methods Across Dose Conditions

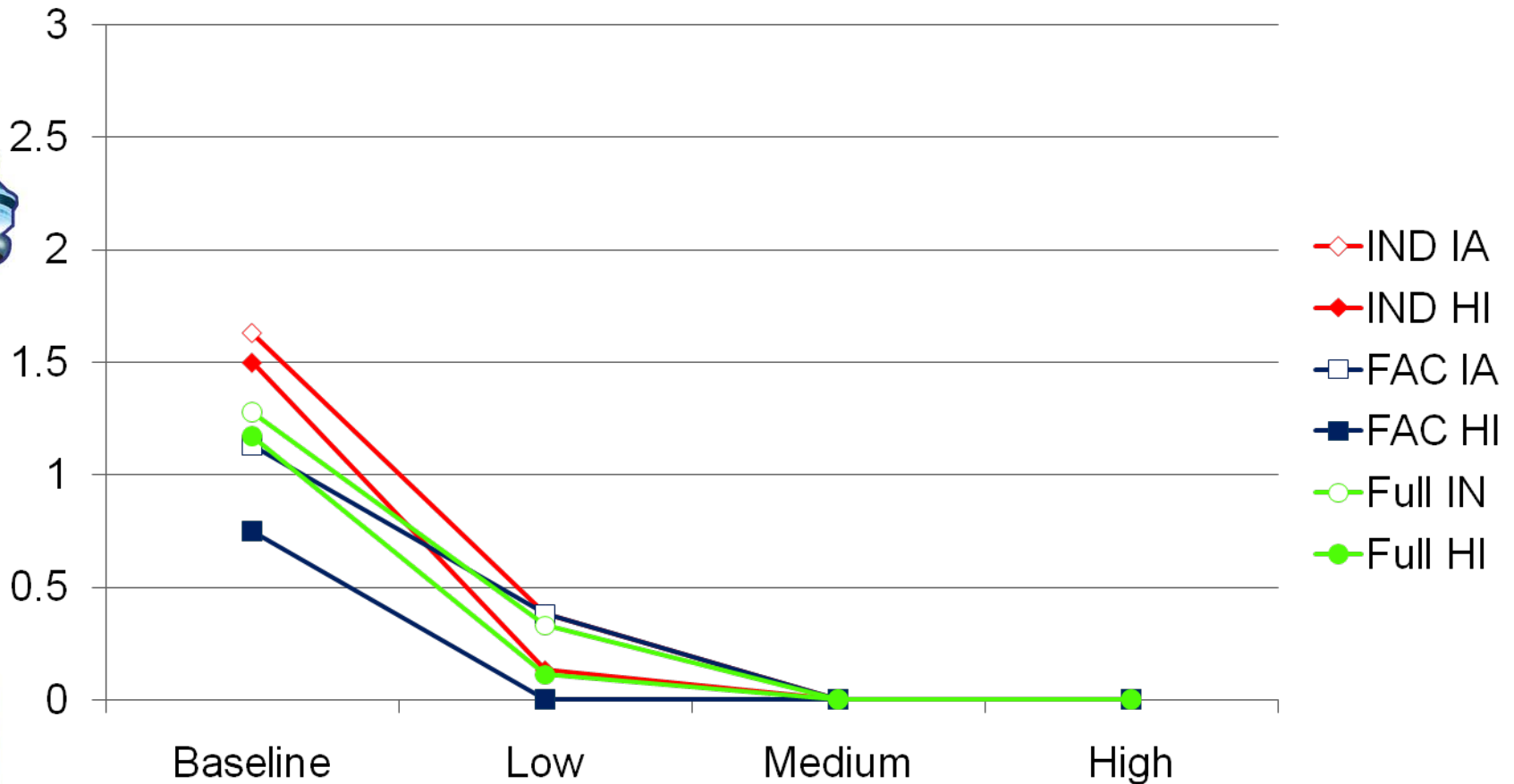


TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Nathan

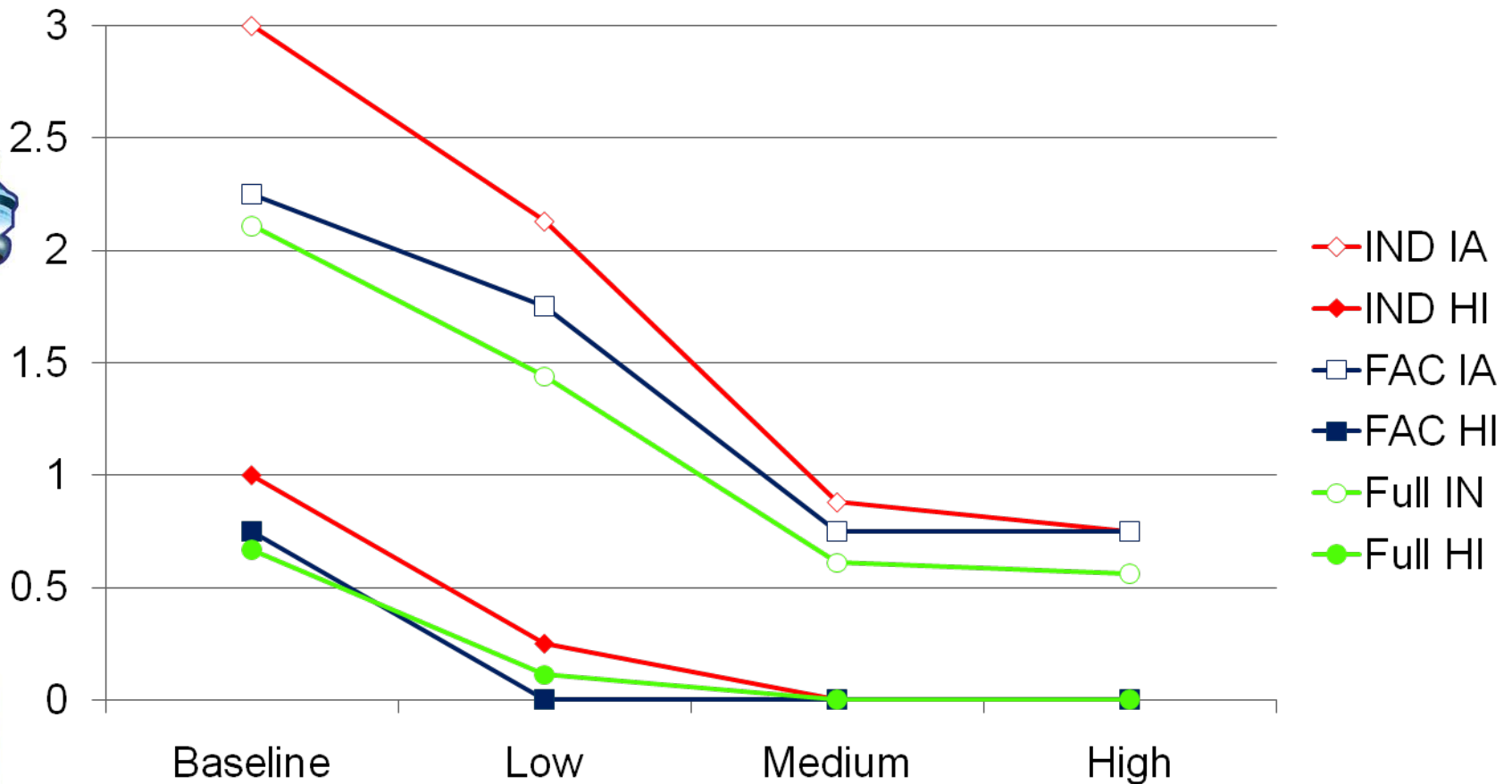


TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Henry

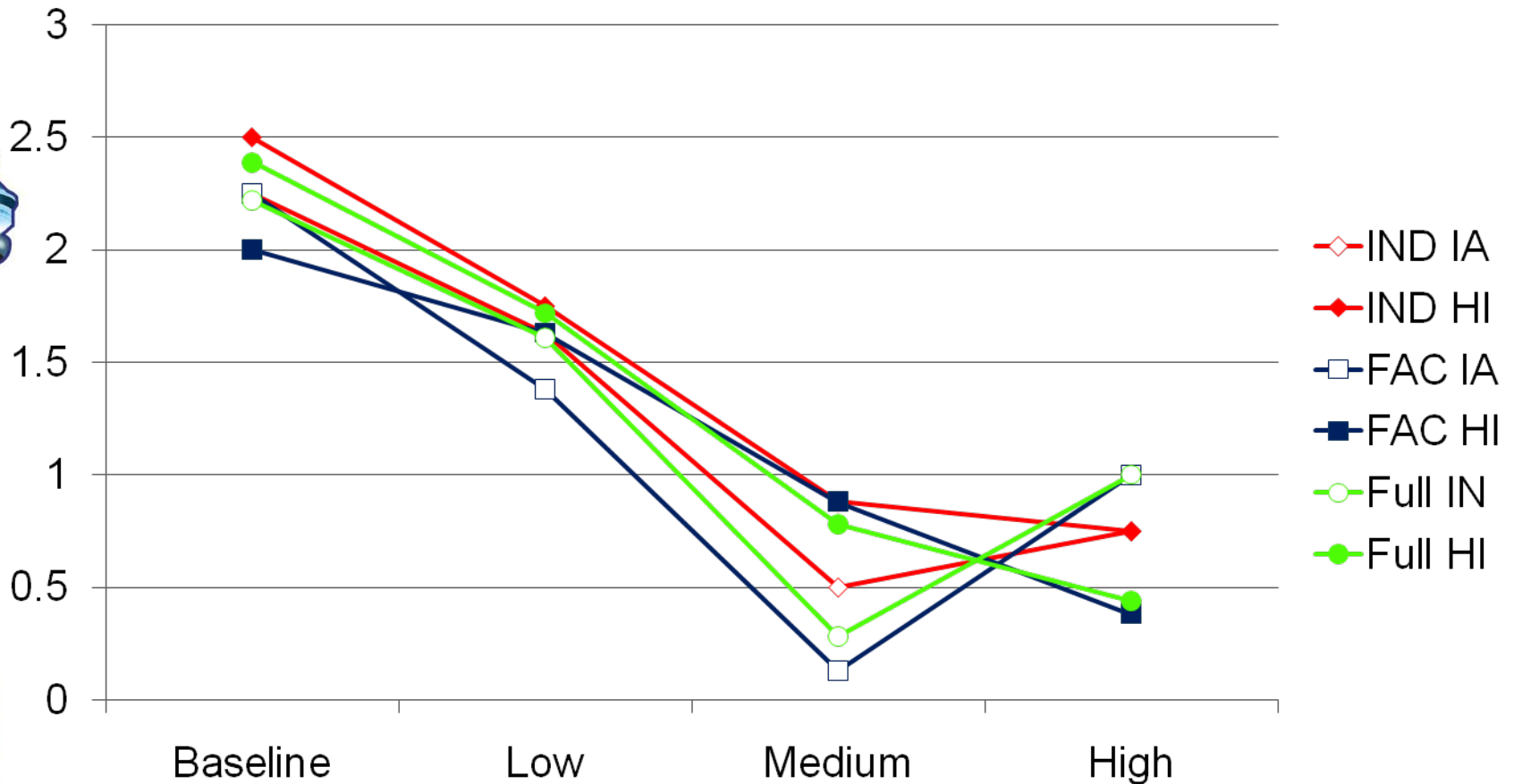


TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Sam



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

The AMBA Model

Multiple-gated Screening

Teacher
Ranking

Administer
Broad-band
Rating
Scale

Progress Monitoring

Create
DBR

Monitor
Progress
with DBR

Create
ABRS for
Each Scale
Reaching
Criterion

Monitor
Progress
with ABRs

Probe with
Broad-band
Scale

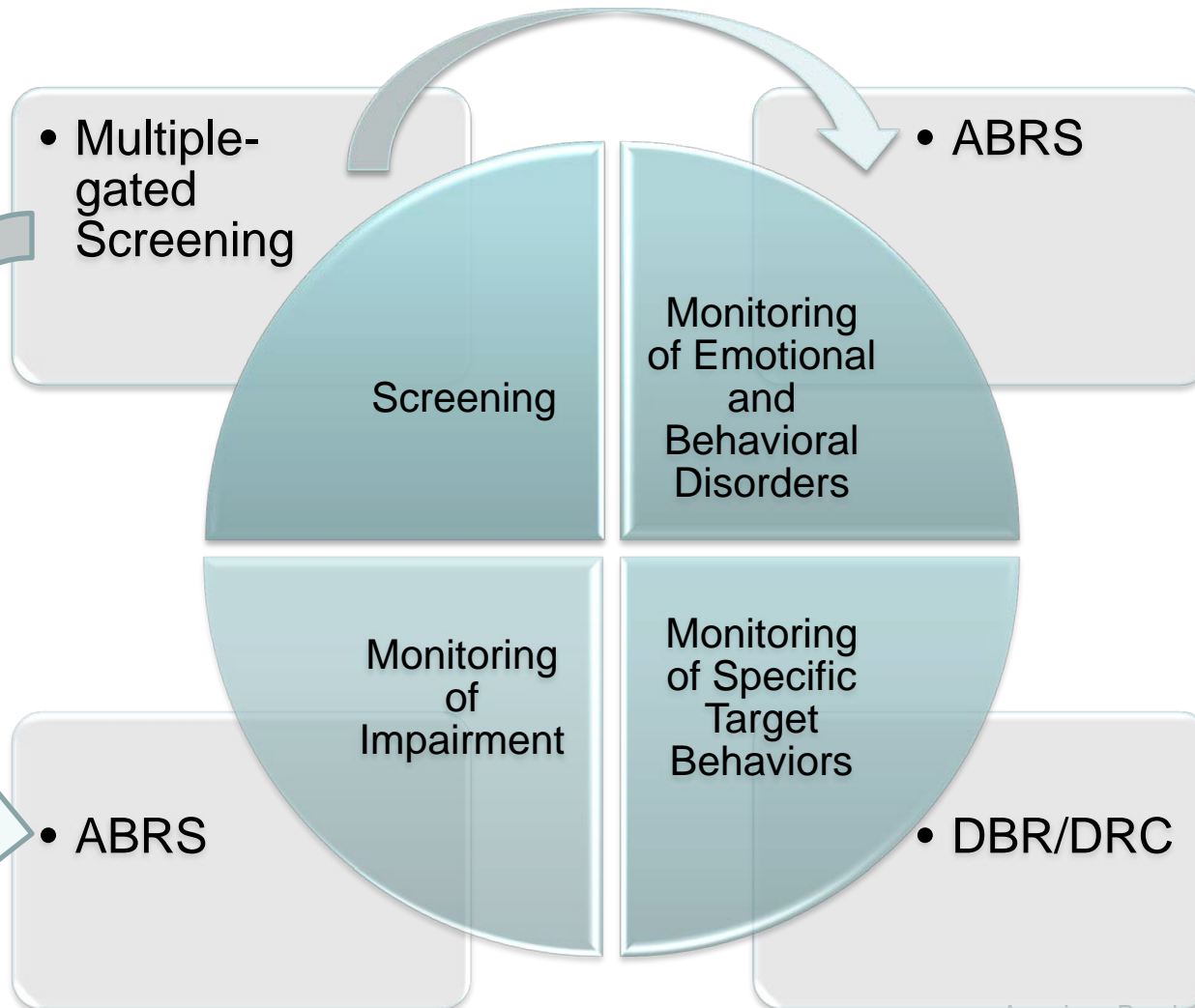


TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Adaptive Model of Behavioral Assessment





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

REVIEW OF DIRECT BEHAVIOR RATING (DBR)

Theodore J. Christ, Ph.D.

Department of Educational Psychology, University of Minnesota



COLLEGE OF EDUCATION
+ HUMAN DEVELOPMENT

UNIVERSITY OF MINNESOTA





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Agenda Slide

Historical Context

- Social-emotional & Behavior Assessment

- Daily Behavior Report Cards (DBRC)

Define Direct Behavior Ratings (DBR)

- Direct Behavior Rating

- Guiding Principles

- Applications

- Target Behaviors

- Formats

Use of DBR in Practice





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Project VIABLE

Project VIABLE:

**Validation of
Instruments for
Assessing
Behavior
Longitudinally &
Efficiently**

Research was supported in part by a grant from the **Institute for Education Sciences**, U.S. Department of Education (R324B060014).





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Reference Materials

Content from:

Christ, T. J., Riley-Tillman, T. C., & Chafouleas, S. M. (2009).
Foundation for the development and use of Direct Behavior
Rating (DBR) to assess and evaluate student behavior
Assessment for Effective Intervention, 34(4), 201-213

As part of a special series on Direct Behavior Ratings





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Historical Context: Social-emotional & Behavior Assessment

- Early 1980s Projective Testing Dominant
 - » Goh & Fuller, 1981, 1983; Goh, Teslow, & Fuller, 1981
 - Rorschach, Draw-a-Person, and Thematic Appreciation Test
 - idiosyncratic & personalized interpretation
- 1990s: Decline in Projectives
 - » Hutton, Dubes, & Muir, 1992; Kennedy, Faust, Willis, & Piotrowski, 1994; Stinnett, Havey, & Oehler-Stinnett, 1994
 - Projectives still in use
 - Greater emphasis on validity, reliability and empirical evidence
 - Ratings scales emerging as a viable alternative





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Historical Context

Recent Practice

- Surveys indicate emphasis on ecology, behavior & intervention
 - » (Demaray et al., 2003; Koonce, 2007; Shapiro & Heick, 2004)
- **76%** greater use of **behavioral assessments** (Shapiro & Heick, 2004)
- **90%** agreed that “behavioral assessments was valuable in linking assessment to intervention” (Shapiro & Heick, 2004).
- **60 to 90%** of cases included **interviews**, **rating scales** and **direct observation** (Shapiro & Heick, 2004)
- Ratings scales and interviews most valuable for
 - Diagnosis of ADHD (Demaray et al., 2003)
 - Provide the most valuable information (Cashel, 2002)



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

DBR Example (standard form)



**Observe
then
Record**

Date: _____ Time: ____ to ____

Academically Engaged (% of time)

0	1	2	3	4	5	6	7	8	9	10
0%										100%
Never						Sometimes				Always

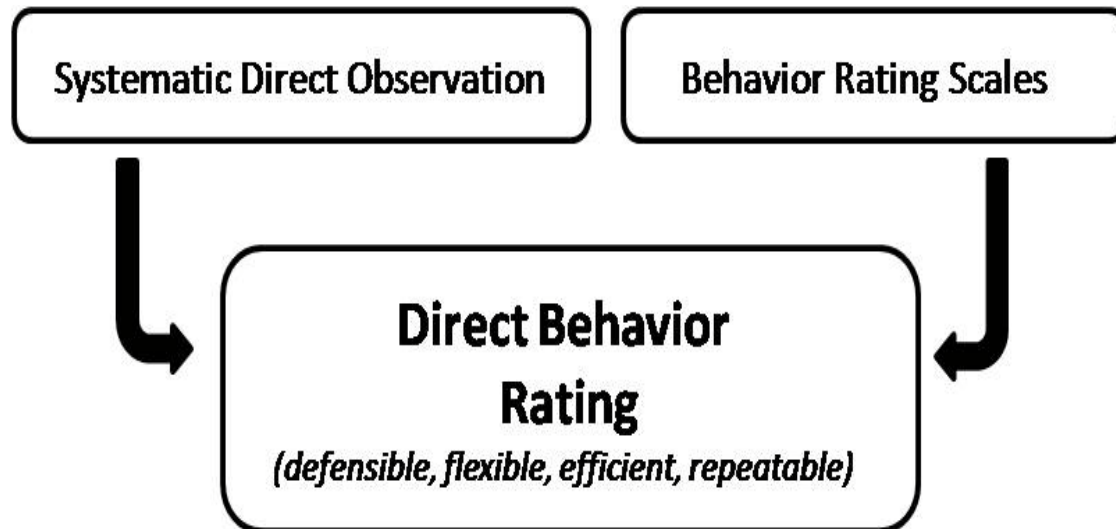
Disruptive Behavior (% of time)

0	1	2	3	4	5	6	7	8	9	10
0%										100%
Never						Sometimes				Always



Emerging Alternative: DBR

An emerging alternative to **behavior rating scales**, **systematic direct observation** and to **informal observations** is ***direct behavior ratings (DBR)*** which combines the advantages of both.



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009



Direct Behavior Ratings

Assessment • Communication • Intervention

www.directbehaviorratings.com/index.html

Development and Evaluation of Direct Behavior Ratings

**A METHOD
BY ANY
OTHER NAME**





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Good Bad or In Between

(Chafouleas, Riley-Tillman, & McDougal, 2002)

Got the ball rolling ... with a review of “Daily Behavior Report Cards”

- Observed many **communication** and **intervention** applications
- Proposed **ASSESSMENT** applications
 - Proposed DBRC as a formative assessment measure
 - Provided initial conceptualization



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Other Names for the DBR

(Riley-Tillman, Chafouleas, & Briesch, 2007)

Daily Behavior Report Card (DBRC)

- Home-School Note
- Behavior Report Card
- Daily Progress Report
- Good Behavior Note
- Check-In Check-Out Card
- Performance-based behavioral recording
- Also
 - Self Management/Monitoring Card
 - Point Card
 - Teacher Rating Form





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Who already uses the Daily Behavior Report Cards (DBRC)?

(Chafouleas, Riley-Tillman, & Sassu, 2006)

Teacher Survey about DBRC:

- Purpose
 - 60% use to **change student behavior** (Intervention)
 - 32% use to **monitor** (Assessment)
 - 32% use “routinely” for **classroom management** (Intervention)
- Types of Behaviors
 - 81% to identify positive behaviors,
 - 77% to identify negative behaviors
- For Whom?
 - 86% use with **individual students**
 - 19% with whole class
 - 9% with small groups





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Initial Studies and Findings

- *What do Daily Behavior Report Cards (DBRCs) measure? An initial comparison of DBRCs with direct observation for off-task behavior*
 - (Chafouleas, McDougal, Riley-Tillman, Panahon, & Hilt, 2005)
- Conclusion: a **moderate association** between teacher perceptions of behavior as measured by DBRC ratings and direct observation conducted by an external observer.





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Initial Studies and Findings

- *An analysis of the similarity of Direct Behavior Ratings and Systematic Direct Observation for off-task and disruptive behaviors*
 - (Riley-Tillman, Chafouleas, Sassu, Chanese, & Glazer, 2008)
- Conclusion: **replicated moderate association** between teacher perceptions of behavior as measured by DBRC ratings and direct observation conducted by an external observer.





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Initial Studies and Findings

- *Acceptability and reported use of Daily Behavior Report Cards among teachers*
 - (Chafouleas, Riley-Tillman, & Sassu, 2006)
 - Conclusion: provide support to previous claims that the **DBRC is both a used and accepted** tool in practice





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Initial Studies and Findings

- *The consistency of Daily Behavior Report Cards in monitoring interventions*
 - (Chafouleas, Riley-Tillman, Sassu, LaFrance, & Patwa, 2007)
 - Conclusion: suggested **similar conclusions** might be drawn when visually examining data collected by an external observer using either systematic direct observation or a DBRC





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

The End of the DBRC Line

Daily Behavior Report Cards

(Description was limiting)

- “Daily”
 - Predefines schedule
 - Precludes alternatives
 - Atheoretical
- “Report Cards”
 - Communication emphasis
 - Precluded other uses

What other description might have stronger - and theoretically consistent - implications for what we are trying to do?



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

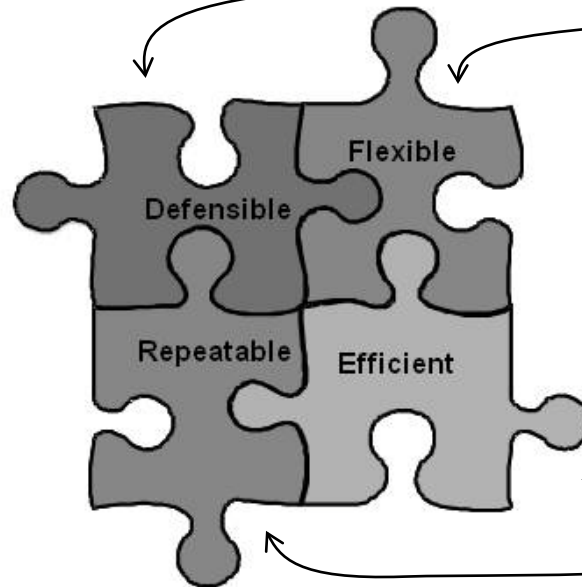
The Emerging Method

Descriptive
Theoretical
Foundation

DIRECT BEHAVIOR RATING

Guiding Principles
for development and
evaluation

ASSESSMENT



INTERVENTION

COMMUNICATION

Uses

Applications



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009



Direct Behavior Ratings

Assessment • Communication • Intervention

www.directbehaviorratings.com/index.html

Development and Evaluation of Direct Behavior Ratings

DIRECT BEHAVIOR RATINGS



The logo for the Toronto 2009 APA convention. It features a stylized illustration of the CN Tower on the left side of the slide. The word "TORONTO" is written in a bold, white, sans-serif font. To its right, the year "2009" is written in a large, white, cursive script font. Below "TORONTO", the text "APA 117TH ANNUAL CONVENTION • AUGUST 6-9" is written in a smaller, white, sans-serif font. The background is a light blue with a pattern of white maple leaves.

TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

What is Direct Behavior Rating?

- DBR is an **evaluative rating** that is generated at the **time and place** that behavior occurs (typically) by those persons who are **naturally occurring** in the **context of interest**
 - Single or (brief) Multi-Item



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Prior DBR-like Examples

Numeric Rating Scale ^a

McCaffery & Beebe (1993)

What does your pain feel like:

-0-	-1-	-2-	-3-	-4-	-5-	-6-	-7-	-8-	-9-	-10-
None	Mild			Moderate			Severe			

Wong-Baker Faces Rating Scale ^b

Wong & Whaley (1986)



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

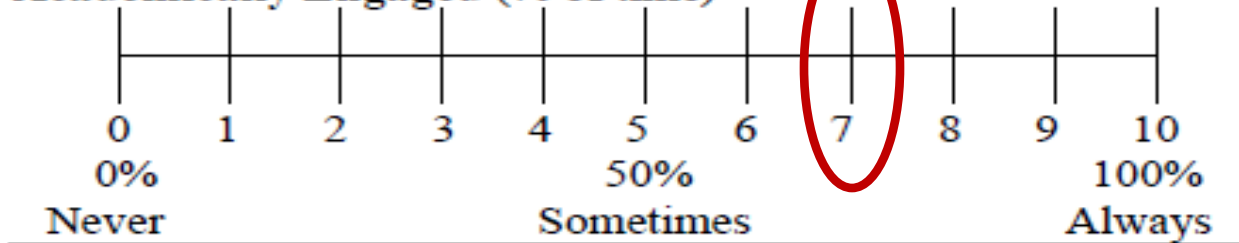
2009

DBR Example (standard form)

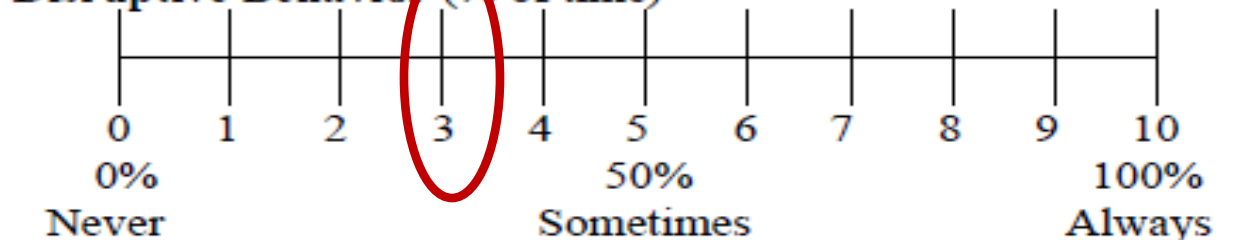
Date: _____

Time: _____ to _____

Academically Engaged (% of time)



Disruptive Behavior (% of time)



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Direct Behavior Rating

DIRECT BEHAVIOR RATING



Direct

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



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Direct Behavior Rating

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).



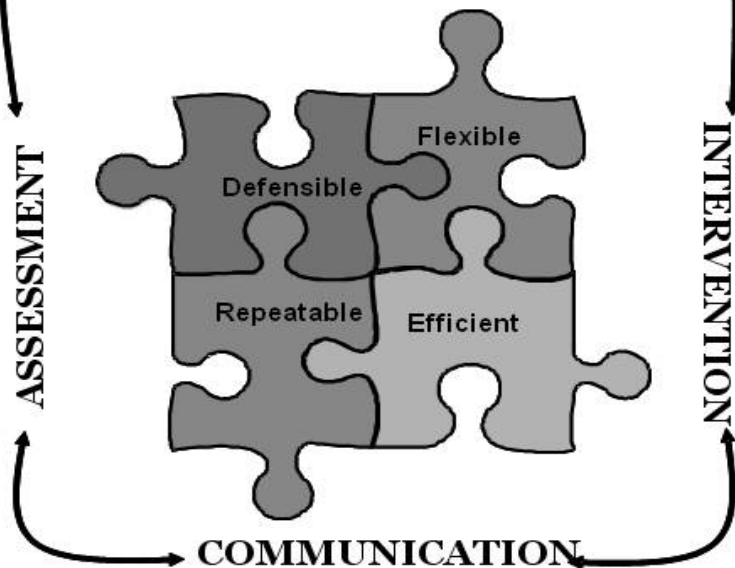
TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Direct Behavior Rating

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 assessments
 - SDO
 - Interviews
 - behavioral rating scales

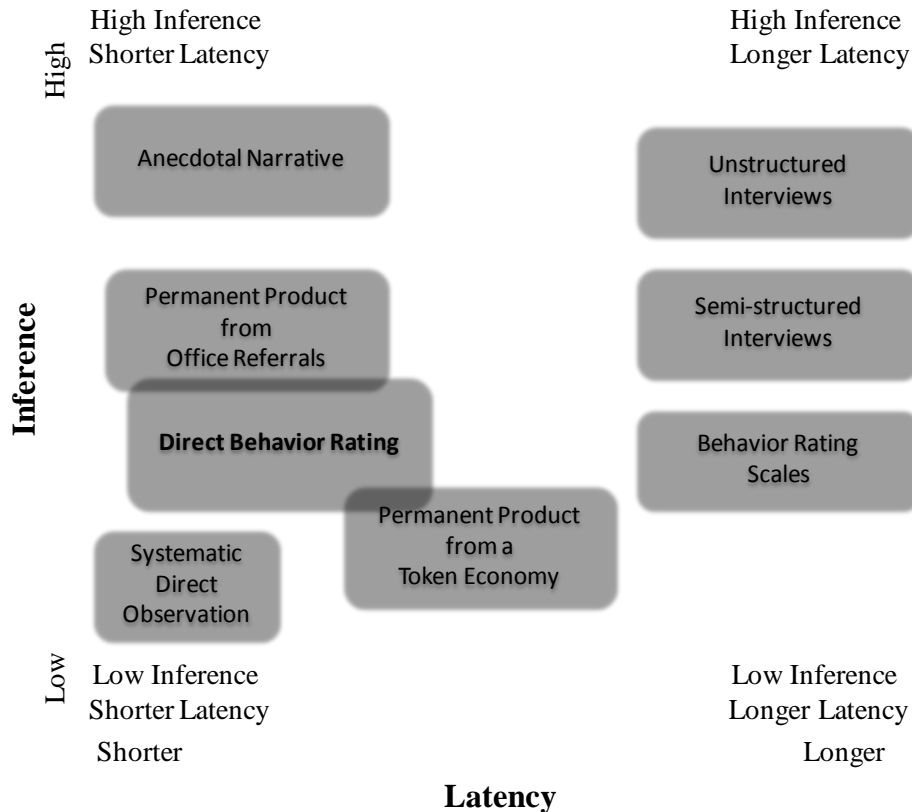


TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

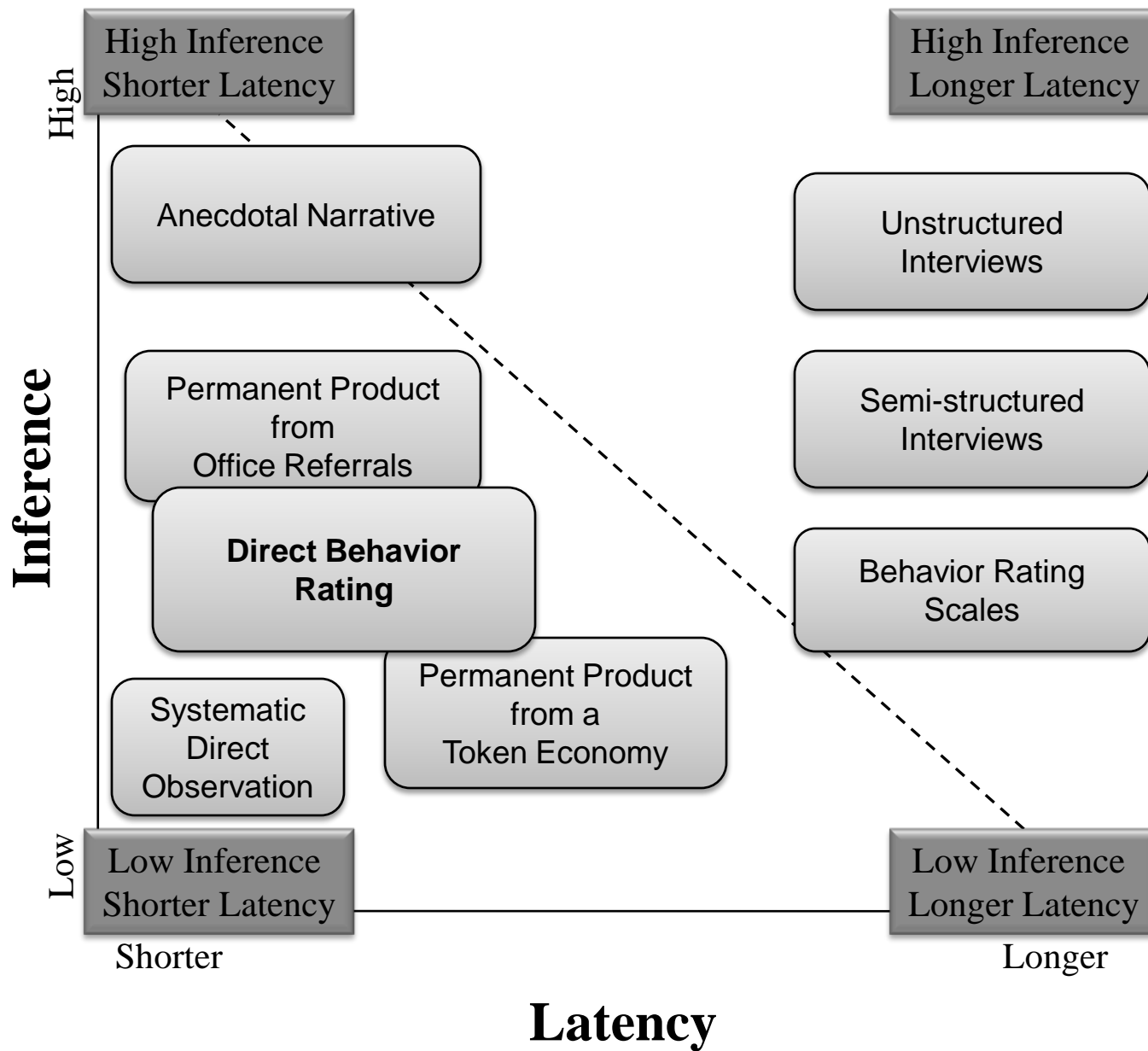
Direct Behavior Rating & Other Methods of Social/Emotional & Behavioral Assessment



Inference - relative objectivity required to generate data

Latency - relative immediacy and proximity between the occurrence of behavior and ratings/reporting





TORONTO

APA 117TH ANNUAL CONVENTION



09



Direct Behavior Ratings

Assessment • Communication • Intervention

www.directbehaviorratings.com/index.html

Development and Evaluation of Direct Behavior Ratings

GUIDING PRINCIPLES



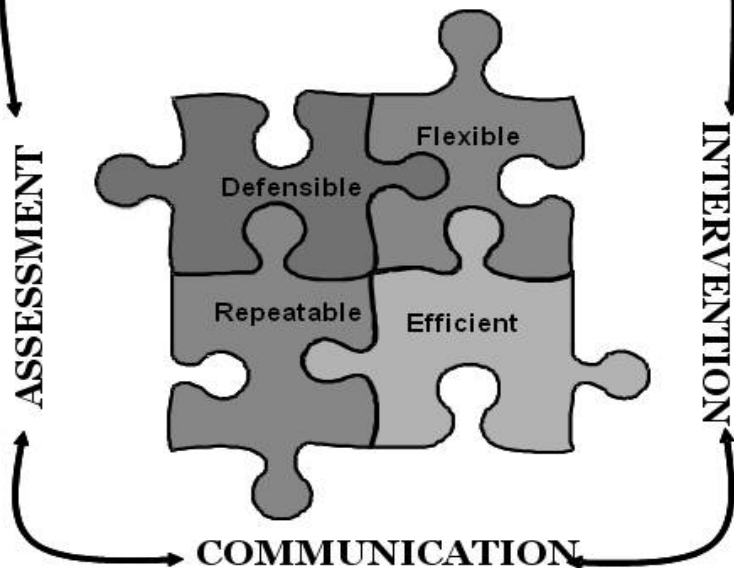
TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Guiding Principles

DIRECT BEHAVIOR RATING



Defensible

- **standardization** and demonstrable technical adequacy (e.g., accuracy, reliability, validity).
- A systematic line of research is necessary and ongoing to evaluate and **develop both DBR procedures and instrumentation.**



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Guiding Principles

DIRECT BEHAVIOR RATING



Flexible

- a wide variety of purposes, contexts and behaviors.
 - Variety of **instruments**
 - Variety of **behaviors**
 - Variety of **purposes**
 - screen and identify behavior problems,
 - define the magnitude of problems,
 - monitor progress and intervention
 - evaluate problem solutions
 - part of a multi-method approach to diagnostic and classification decisions.



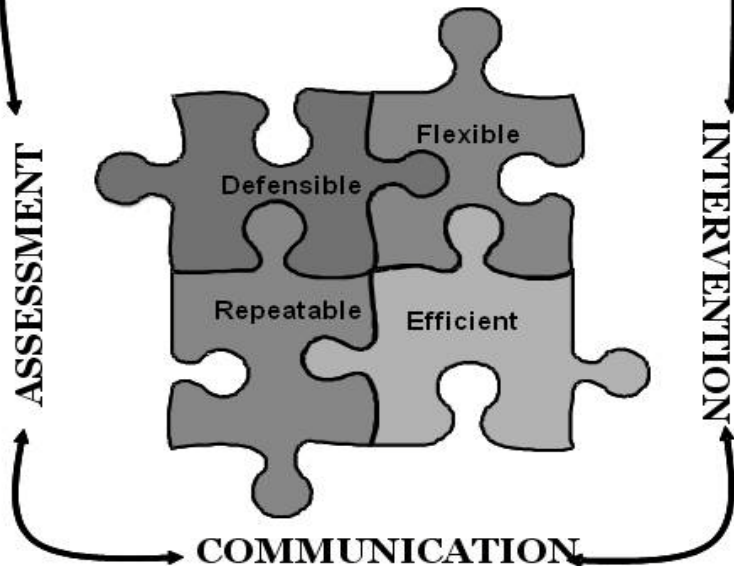
TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Guiding Principles

DIRECT BEHAVIOR RATING



Efficient

- a) ratings are completed by those **persons who are naturally occurring** in the context of interest, and
- b) ratings are collected in brief periods of time (i.e. **few seconds**), resulting in minimal disruption.



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Guiding Principles

DIRECT BEHAVIOR RATING



Repeatable

- facilitates **ongoing data collection** within and across occasions



TORONTO

APA 117TH ANNUAL CONVENTION



09



Direct Behavior Ratings

Assessment • Communication • Intervention

www.directbehaviorratings.com/index.html

Development and Evaluation of Direct Behavior Ratings

APPLICATIONS



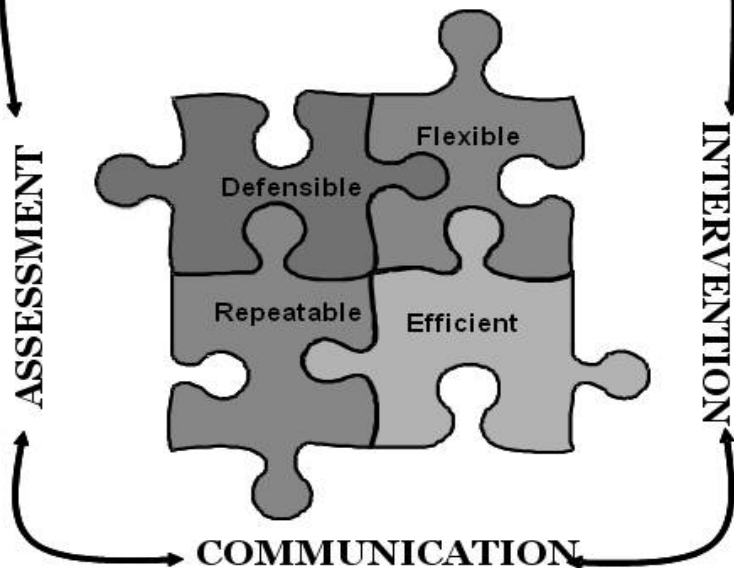
TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Applications of DBR

DIRECT BEHAVIOR RATING



Assessment

- DBR provides information to evaluate child behavior and guide decisions
 - "What percentage of time is Sarah disruptive during math class?" or
 - "What percentage of the time is Immanuel compliant with adult instructions?"



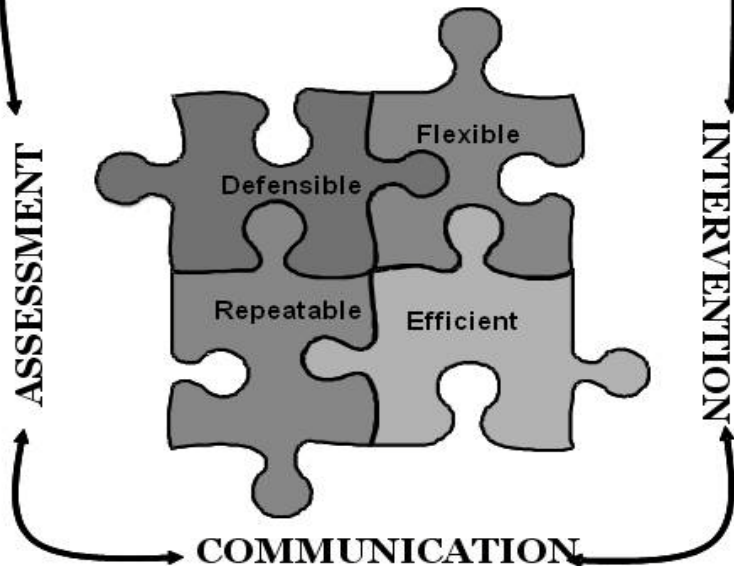
TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Applications of DBR

DIRECT BEHAVIOR RATING



Communication

- to share information among those key persons in a child's life (e.g. teacher-child, home-school, teacher-teacher).
 - immediate and consistent feedback about student behavior
 - fosters shared responsibility for student welfare
 - establish shared behavior goals across settings and persons
 - Increases opportunities for feedback & positive attention



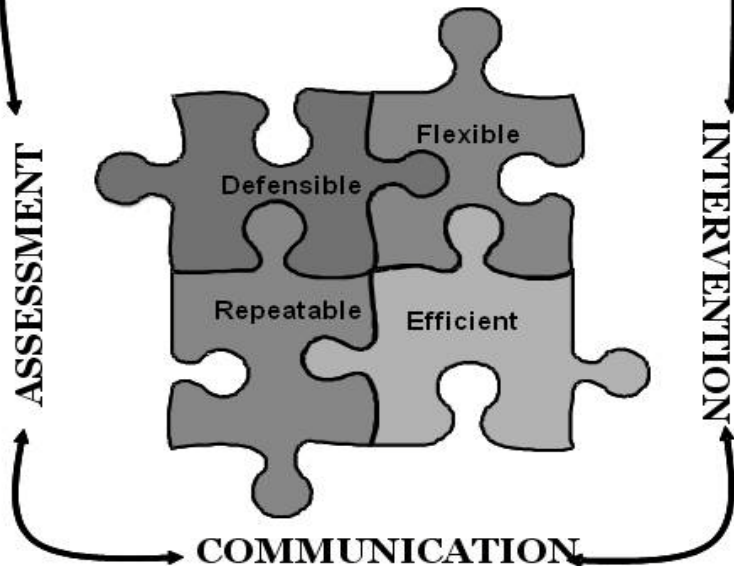
TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Applications of DBR

DIRECT BEHAVIOR RATING



Intervention

- substantial body of research exists to demonstrate the effectiveness of interventions that include DBR as one component
 - Incentive Programs
 - Self Management



TORONTO

APA 117TH ANNUAL CONVENTION



09



Direct Behavior Ratings

Assessment • Communication • Intervention

www.directbehaviorratings.com/index.html

Development and Evaluation of Direct Behavior Ratings

INSTRUMENTATION & PROCEDURES





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

What, When, Where to Observe

Instrumentation

- What bx
 - General
 - Specific
- Definitions
- Rating item(s)

Procedures

- When
 - Where
 - Who
 - How often
- ... that data are collected





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

What behaviors do I rate?

- **Academically Engaged** is defined as actively or passively participating in the classroom activity.
 - For example: writing, raising his/her 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 Behavior** is defined as a student action that interrupts regular school or classroom activity.
 - For example: out of his/her seat, fidgeting, playing with objects, acting aggressively, talking/yelling about things that are unrelated to classroom instruction.

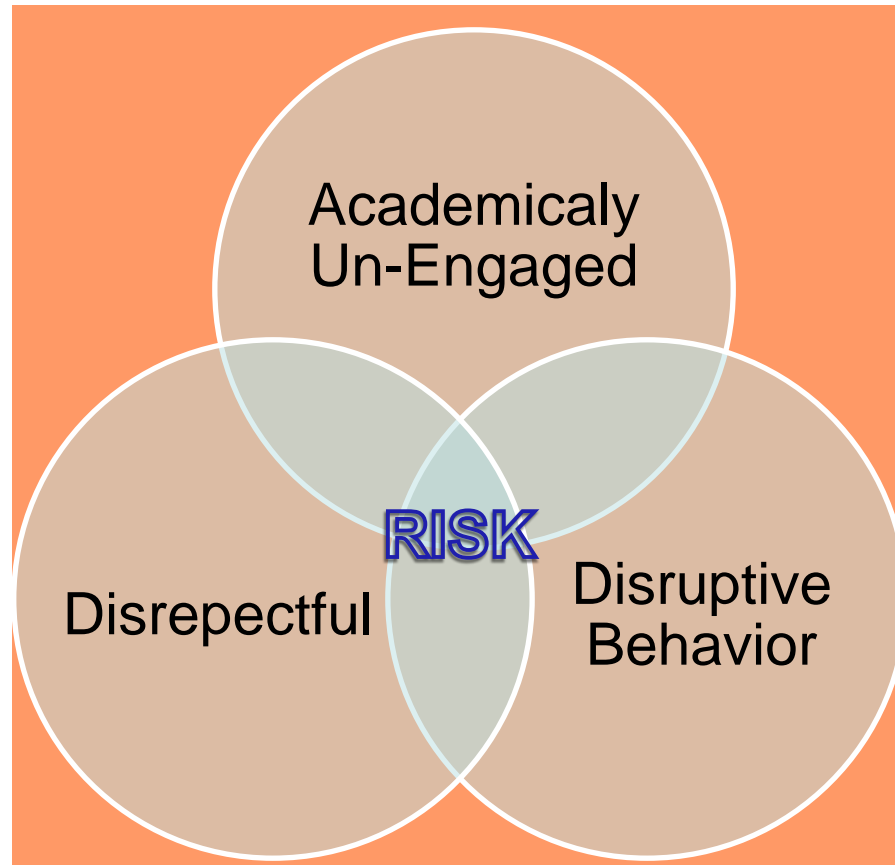


TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

What behaviors do I rate?



*Working Risk-Resilience Model
for School-Based Behavior



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009



Insert standard form

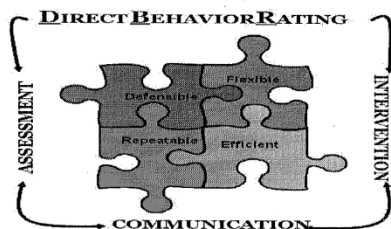


TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

DBR Booklet



Student Name

Rater Name

to _____
Date Range

*Setting Information: _____

**The setting (time and activity) must be consistent across all ratings.
For example, setting could be all day, morning, or math class.*

Behavior Definitions

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

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

Optional Behaviors are chosen by the rater and can be used to track behaviors that are specific to the student being rated. For example, if the student throws objects in the classroom, the rater may write a definition for throwing objects that is concrete, observable, and measureable. Additionally, the definition should include examples and non-examples of the behavior.

Optional Behaviors (be sure to clear define)

- Interaction with Peers
- Physical Behavior toward Others
- Physical Behavior toward Self
- Respectful
- Compliance
- Verbal Behavior
- Out of Seat
- Tantruming

© Chafouleas, Riley-Tillman & Christ

© Chafouleas, Riley-Tillman & Christ



TORONTO

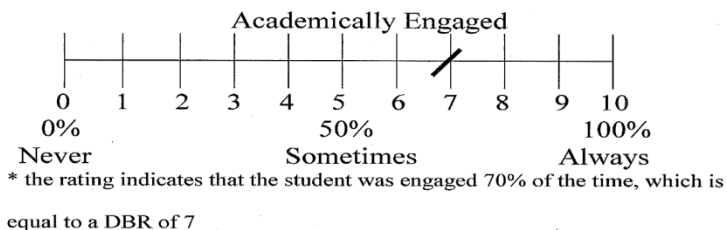
APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

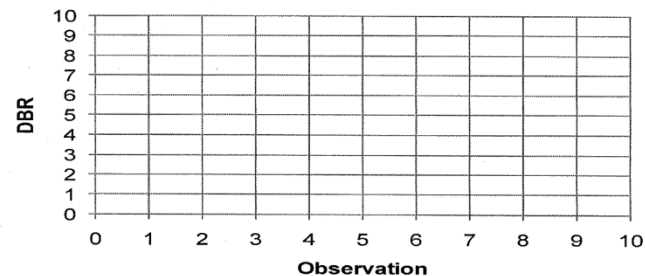
DBR Booklet

Directions for Direct Behavior Rating (DBR)

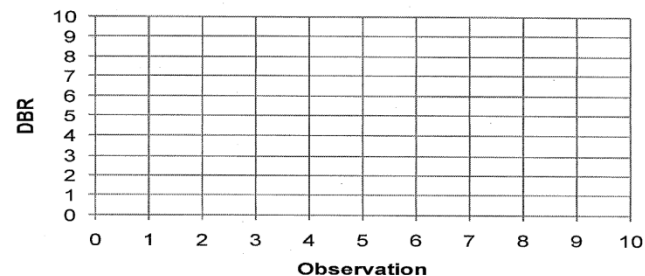
1. Determine the behaviors of interest, either by selecting from among the possible pre-defined target behaviors or identify your own target behavior.
2. Decide who, where, and how often to collect behavior ratings with DBR (e.g., daily, AM, PM). Ratings can be completed in a matter of seconds.
3. Observe and estimate the amount of time that the behavior occurs during an observation period (e.g., full day, half day, class period).
4. Collect multiple ratings across multiple occasions (see below).
5. Plot data graphically, and evaluate child behavior.



Behavior: _____



Behavior: _____



© Chafouleas, Riley-Tillman & Christ

© Chafouleas, Riley-Tillman & Christ



TORONTO

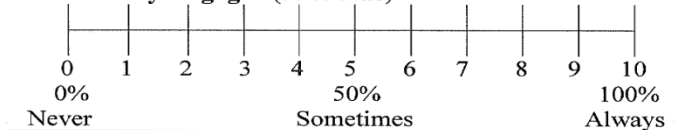
APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

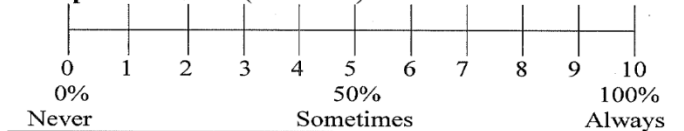
DBR Booklet

Date: _____ Time: _____ to _____

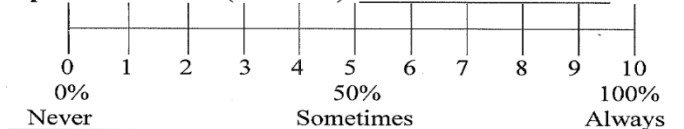
Academically Engaged (% of time)



Disruptive Behavior (% of time)



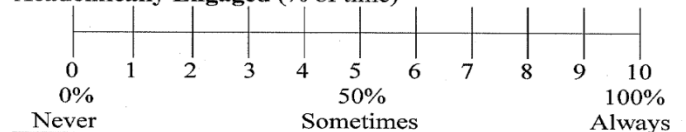
Optional Behavior (% of time):



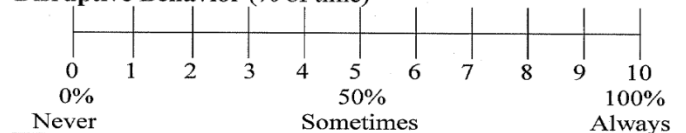
Notes:

Date: _____ Time: _____ to _____

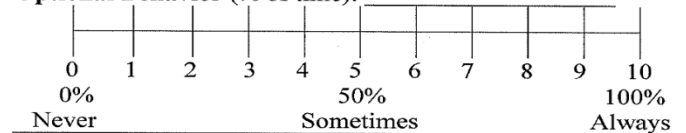
Academically Engaged (% of time)



Disruptive Behavior (% of time)



Optional Behavior (% of time):



Notes:



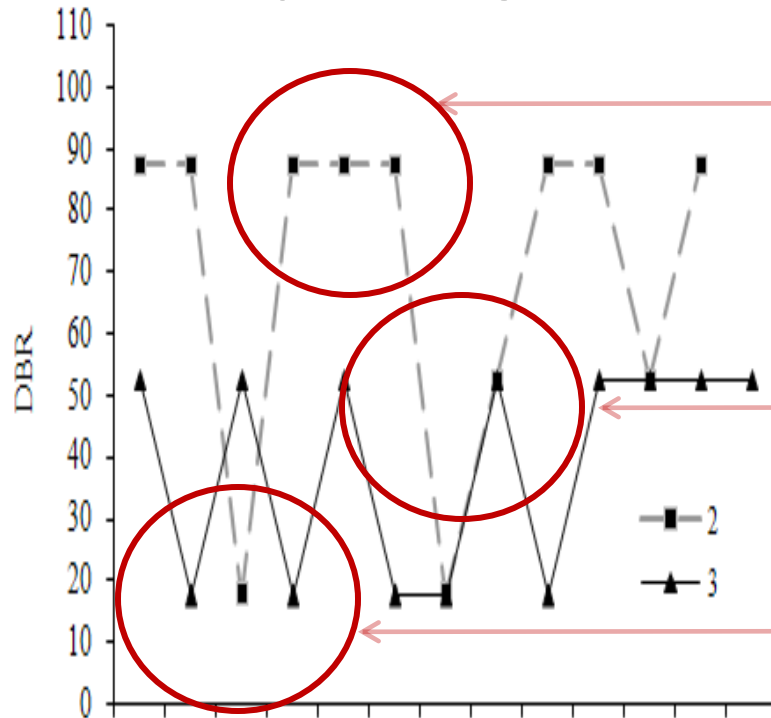
TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

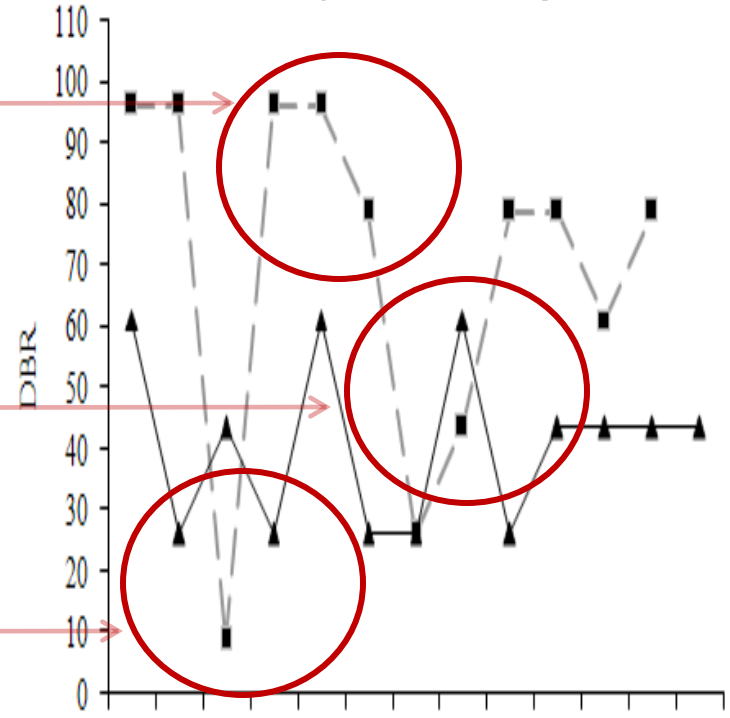
2009

DBR Format: 10 Gradients

3 Point Scale



10 Point Scale



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Progress Monitoring Booklet



Student Name

Rater Name

to _____
Date Range

*Setting Information: _____

**The setting (time and activity) must be consistent across all ratings.
For example, setting could be all day, morning, or math class.*

© Chafouleas, Riley-Tillman & Christ

Behavior Definitions

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

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

Optional Behaviors are chosen by the rater and can be used to track behaviors that are specific to the student being rated. For example, if the student throws objects in the classroom, the rater may write a definition for throwing objects that is concrete, observable, and measurable. Additionally, the definition should include examples and non-examples of the behavior.

Optional Behaviors (be sure to clearly define)

- Interaction with Peers
- Physical Behavior toward Others
- Physical Behavior toward Self
- Respectful
- Compliance
- Verbal Behavior
- Out of Seat
- Tantruming

© Chafouleas, Riley-Tillman & Christ





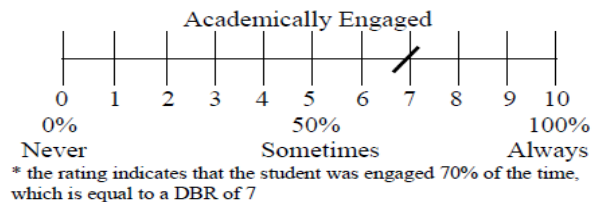
TORONTO 2009

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

Progress Monitoring Booklet

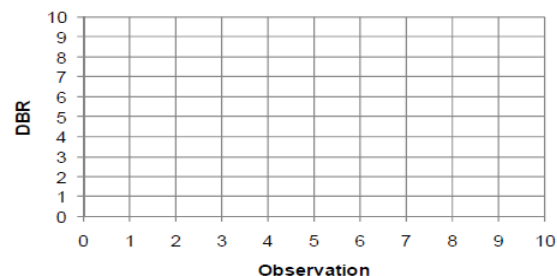
Directions for Direct Behavior Rating (DBR)

1. Determine the behaviors of interest, either by selecting from among the possible pre-defined target behaviors or identify your own target behavior.
2. Decide who, where, and how often to collect behavior ratings with DBR (e.g., daily, AM, PM). Ratings can be completed in a matter of seconds.
3. Observe and estimate the amount of time that the behavior occurs during an observation period (e.g., full day, half day, class period).
4. Collect multiple ratings across multiple occasions (see below).
5. Plot data graphically, and evaluate child behavior.

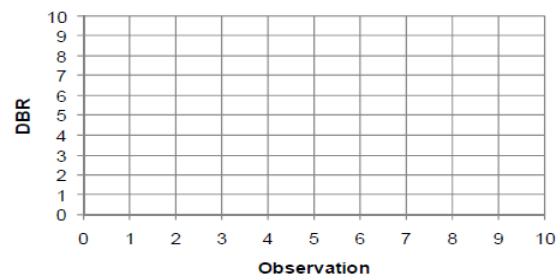


© Chafouleas, Riley-Tillman & Christ

Behavior: _____



Behavior: _____



© Chafouleas, Riley-Tillman & Christ





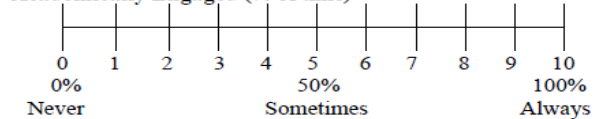
TORONTO 2009

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

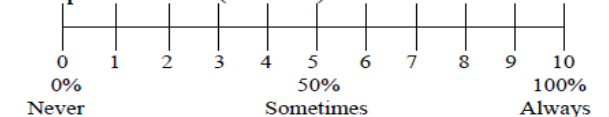
Progress Monitoring Booklet

Date: _____ Time: ____ to ____

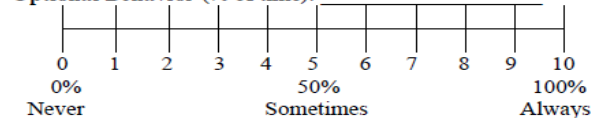
Academically Engaged (% of time)



Disruptive Behavior (% of time)



Optional Behavior (% of time):

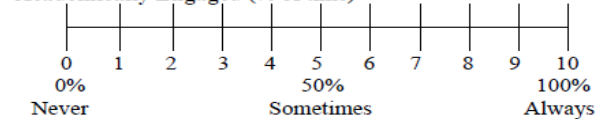


Notes:

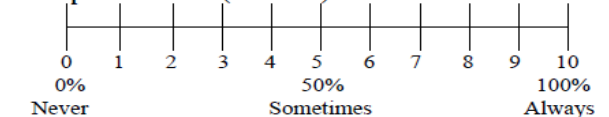
© Chafouleas, Riley-Tillman & Christ

Date: _____ Time: ____ to ____

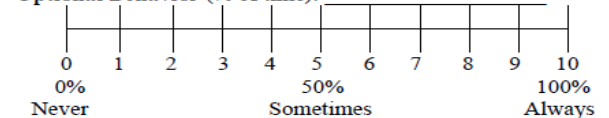
Academically Engaged (% of time)



Disruptive Behavior (% of time)



Optional Behavior (% of time):



Notes:

© Chafouleas, Riley-Tillman & Christ



TORONTO

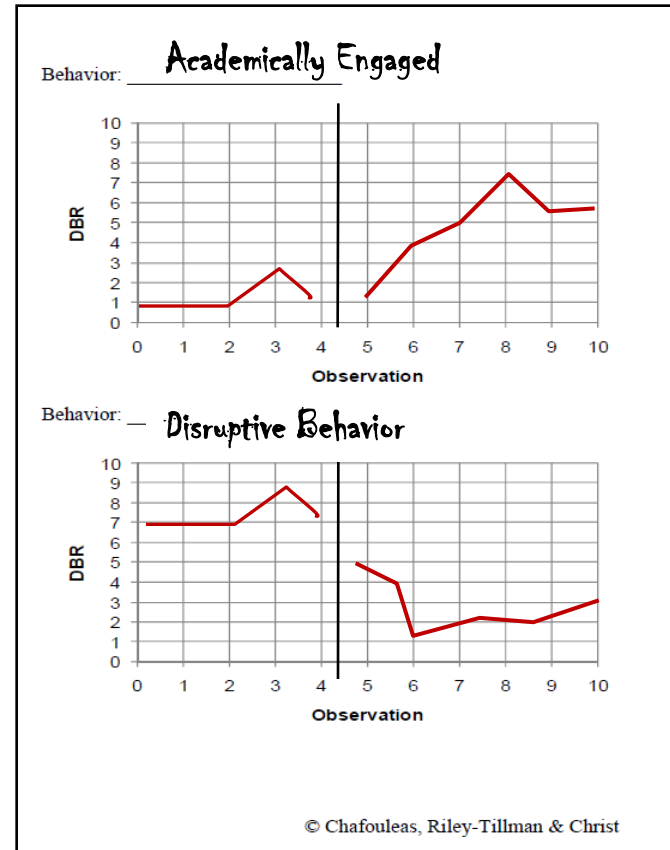
APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Progress Monitoring Booklet

How Often?

We recommend (5 to) **10 datapoints** per phase, but the emphasis is on ideographic analysis *and* high/low stakes decisions



TORONTO

APA 117TH ANNUAL CONVENTION



09



Direct Behavior Ratings

Assessment • Communication • Intervention

www.directbehaviorratings.com/index.html

Development and Evaluation of Direct Behavior Ratings

ONLINE SUPPORT



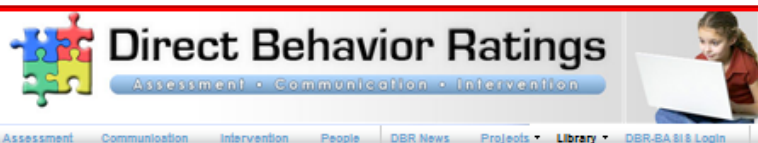
TORONTO



Direct Behavior Ratings

Assessment • Communication • Intervention

www.directbehaviorratings.com/index.html



News

Upcoming Special Issue of the Journal
Assessment for Effective Interventions:

- Direct Behavior Rating (DBR): An Emerging Method for Assessing Social Behavior within a Targeted Intervention System

Upcoming DBR Presentations

Current/Recent Research Studies:

- Preschool intervention study that uses Direct Behavior Ratings as a communication tool
- Direct Behavior Ratings direct training study looking at the effects of direct training with corrective feedback on ratings

[click for more](#)

About Us



Sandra M. Chaffin, Ph.D.
Univ. of Connecticut
Associate Professor,
School Psychology
Research Scientist, CDBR

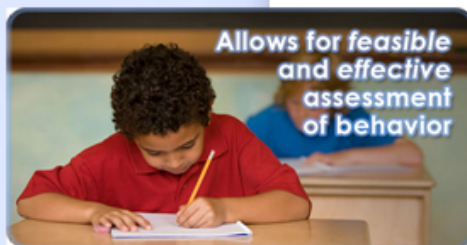


I. Chouksey, Ph.D.
East Carolina University
Associate Professor,
School Psychology
Director of the School
Psychology Program



Theodore J. Chmielewski, Ph.D.
University of Minnesota
Associate Professor,
School Psychology
243 Education Sciences
Building

[click for more](#)



"I was surprised at how easy it was to complete the Direct Behavior Rating forms. This information is really valuable in helping me understand what's happening in my classroom." Sue, Kindergarten teacher

What is Direct Behavior Rating (DBR)?

DBR involves rating of behavior following a specified observation period, and then sharing of that information to inform decisions. As an example, a teacher might use DBR to rate how well Johnny paid attention in math class. Then, that teacher might share that rating with Johnny and, as part of an intervention, link a consequence (e.g., sticker) to that rating. DBR tools have a long history of use as a component of a behavior support plan (e.g., self-management, behavior contract), as well as the method for collecting information about behavior change over time (e.g., monitoring effects of medication for ADHD). Other common terms for DBR tools have included home-school note, good behavior note, behavior report card, etc...

Why use Direct Behavior Rating?

DBR can facilitate communication among students, parents, and teachers because ratings can provide a simple, inexpensive, and flexible way to provide frequent feedback about behavior. DBR is also appealing given a connection between data collection and intervention – DBR may serve both purposes! For example, DBR can be used to monitor behavior in response to an intervention while at the same time serving as an intervention tool to teach and reinforce expectations regarding behavior.



DBR for Assessment

DBR use in assessment provides information to evaluate child behavior and guide decisions related to behavior supports. For instance, a DBR may be used to answer the question, "What percentage of time is Sarah disruptive during math class?" or "What percentage of the time is Immanuel compliant with adult instructions?"

How can I use a DBR in assessment?

It's simple and quick! Print out a DBR form and complete the top section.

1. Determine the behaviors of interest, either by selecting from among the possible pre-defined target behaviors or identify your own target behavior.
2. Decide who, where, and how often to collect behavior ratings with DBR (e.g., daily, weekly). Ratings can be completed in a matter of seconds.
3. Collect multiple ratings across different occasions (e.g., periods, days) (see [DBR Standard Instructions](#)).
4. Plot data graphically, and evaluate child behavior (see [DBR Graphing and Interpretation](#)).

Who can use a DBR for assessment?

DBR can be used by parents, teachers, students, administrators, and intervention teams to collect information and make decisions regarding a child's behavior. It's a great tool for everyone because it is quick, flexible, and [evidence-based](#).



Additional Resources

- PowerPoint: DBR for Assessment
- DBR Standard Form
- DBR Standard Form Instructions
- DBR Graphing and Interpretation
- Other Resources



TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Current - Future Directions

• 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.





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Conclusion Slide

- DBR is consistent with current practice
- DBR can supplement other methods
- DBR is highly efficient
- Substantial research basis for use
- Technology-based supports available (Spring 2010)





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

REVIEW OF PERMANENT PRODUCTS: OFFICE DISCIPLINE REFERRALS

Sandra M. Chafouleas, Ph.D.

Department of Educational Psychology, Neag School of Education
Center for Behavioral Education and Research
University of Connecticut





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

What are Permanent Products?

“the tangible items or environmental effects that result from a behavior”
(p. 62)

(Alberto & Troutman, 2006)





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Potential Positive Features in Formative Assessment

- Already available AND often collected in a formative fashion
- Highly contextually relevant
- Natural occurrence can reduce/limit reactivity

(Adapted from Chafouleas, Riley-Tillman, & Sugai, 2007)





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Potential Negative Features in Formative Assessment

- Must be easily accessible in a timely manner
- Organized system for easy summarization must be in place
- Resulting data must be trustworthy

(Adapted from Chafouleas, Riley-Tillman, & Sugai, 2007)





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Examples of “child” permanent products

- grades
- scores on state-mandated tests
- work samples
- curriculum-based assessments
- attendance
- suspension/expulsion
- data from classroom behavior plans
- office discipline referral (ODR)



What is an ODR?

“an event in which (a) a student engaged in a behavior that violated a rule/social norm in the school, (b) a problem behavior was observed by a member of the school staff, and (c) the event resulted in a consequence delivered by administrative staff who produced a permanent (written) produce defining the whole event” (p. 96)

Sugai, Horner, & Walker (2000)

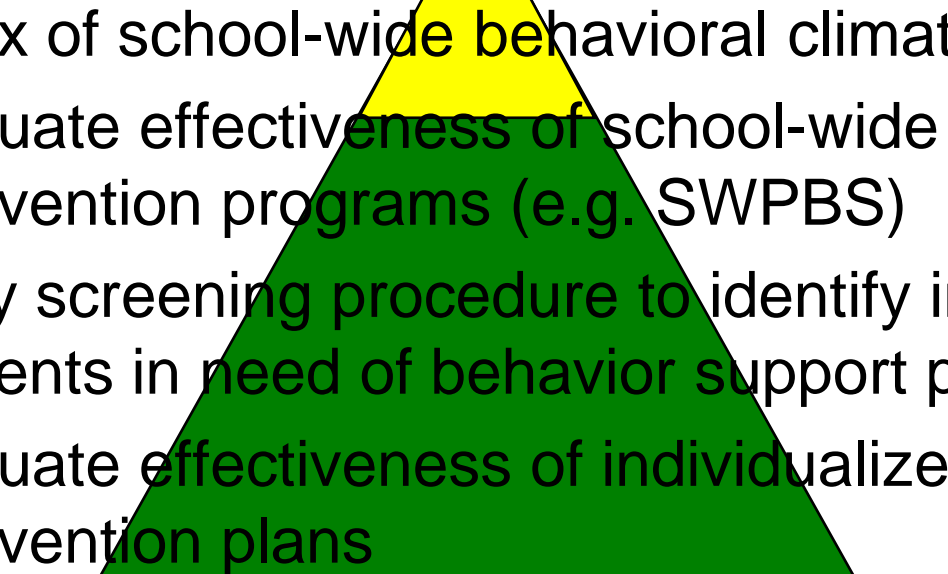


TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

What are the purported uses of ODR?

- 
- Index of school-wide behavioral climate
 - Evaluate effectiveness of school-wide behavioral intervention programs (e.g. SWPBS)
 - Early screening procedure to identify individual students in need of behavior support planning
 - Evaluate effectiveness of individualized intervention plans



Example

- Staff at West High School note concern about the number of fights occurring among students.
- ODRs over the past 2 months are reviewed.
- Review revealed a) most fighting incidents are occurring outside cafeteria and in bus loading area AND b) Johnny and Sam are the most likely culprits.
- Staff are re-assigned to increase levels of active supervision in those areas at key times.
- “Johnny and Sam” are brought to Behavior Support Team for additional support planning.





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

What is the current evidence for ODR?

- Defensible
- Flexible
- Efficient
- Repeatable



DEFENSIBLE

“Overview Snapshot”

PsychInfo – 29 articles hit with “ODR” and 155 with “discipline referral”

- Majority used ODR as a key DV/criterion
 - Only 4 detailed primary purpose was investigation of aspects of validity
 - Of those, use related to *school-wide indices* most clearly studied



ODR Defensibility

- ODRs are related to poor outcomes
e.g. school failure, juvenile delinquency

Is this surprising?

Martens (1993) – the implicit normative comparisons made by teachers suggests that by the time you get an ODR, your behavior is far from “acceptable”



ODR Defensibility

Irvin, Tobin, Sprague, Sugai, & Vincent (2004)

- Authors reviewed evidence related to DBR use and utility in a) research on school-wide discipline/juvenile delinquency, b) assessment of intervention effects, c) in program evaluations

Their conclusions:

- Evidence available to support **construct validity** for interpretation and use of ODR
- Focus of establishing “validity” of ODR should be on utility for informing decision making



ODR Defensibility

Walker, Cheney, Stage, & Blum (2005)

Investigation of tools for identifying student behavioral functioning within SWPBS

- Schools using SWPBS for at least 3 years
- Tracked students nominated through Gate 2 of SSBD using ODR (0-1, 2+) and rating scale (SSRS)

Results suggested:

- ODR alone under-represented problem behavior
- More ODRs meant referral for more intense services
 - However, majority referred to teams did not have multiple ODRs



ODR Defensibility

Nelson, Benner, Reid, Epstein, & Currin (2002)

Investigation of the convergent validity of ODR with CBCL-TRF in an elementary sample

- Used liberal criteria of 1 or more ODR and borderline or clinical on TRF

Results suggested: high levels of false negatives, low to moderate agreement (especially for Internalizing scales)



REPEATABLE / DEFENSIBLE

Yes, ODR can be administered repeatedly, but...
there is a **base rate** problem - *INDIVIDUALS*

McIntosh, Horner, Chard, Boland, Good (2006)

Investigation of reading and behavior screening
measures to predict non-response to SWPBS

Results suggested:

- Under-representation of problem behavior
(e.g. NO k students received any!)
- ODR does not provide indicator of prosocial
functioning



REPEATABLE / DEFENSIBLE

Yes, ODR can be administered repeatedly, but...
there is a **base rate** problem – *SCHOOL-WIDE*

Wright & Dusek (1998)

Investigated base rates of ODRs for *physical aggression*
across subgroups over 3 years across 2 elementary schools

Results suggested:

- % of total students receiving at least 1 ODR (11%, 33%)
variable across but stable within school
- male and sped = higher probability of ODR
- stable rate of recidivism

Thoughts: Are local norms best for understanding
base rate?

What are implications for subgroups?



DEFENSIBLE / FLEXIBLE / EFFICIENT

Irvin, Horner, Ingram, Todd, Sugai, Sampson, & Boland (2006)

Investigation of electronic ODR use and perceptions
- surveyed participants & tracked SWIS use (3 mths)

Results suggested:

- Data accessed and reported to be used at least monthly
- Elementary users slightly more positive than middle
- Users requested improved flexibility to customize
- Use for “schoolwide” purposes more often than for “individual”



A graphic for the Toronto 2009 APA Convention. It features a blue background with a pattern of white maple leaves. On the left, a stylized illustration of the CN Tower is shown in blue and green. At the top, the word "TORONTO" is written in large, white, bold, sans-serif capital letters. To its right, the year "2009" is written in a large, white, stylized script font. Below "TORONTO", the text "APA 117TH ANNUAL CONVENTION • AUGUST 6-9" is written in a smaller, white, sans-serif font.

TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

DEFENSIBLE / FLEXIBLE / EFFICIENT

Complexities of defensibility...

Suggestions to enhance defensibility can limit flexibility while enhancing efficiency

- Need to create operational definitions that are mutually exclusive
- Consistently implement defined consequences





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

ODR Categories within SWIS

MAJOR

- Abusive Language/ Inappropriate Language/ Profanity
- Arson
- Bomb Threat/ False Alarm
- Defiance/Disrespect/ Insubordination/
- Non-Compliance
- Disruption
- Dress Code Violation
- Fighting/ Physical Aggression
- Forgery/ Theft
- Gang Affiliation Display
- Harassment/Bullying
- Inappropriate Display of Affection
- Inappropriate Location/
- Lying/Cheating
- Other Behavior
- Property Damage/Vandalism
- Skip class
- Truancy
- Tardy
- Technology Violation
- Unknown
- Use/Possession of Alcohol
- Use/Possession of Combustibles
- Use/Possession of Drugs
- Use/Possession of Tobacco
- Use/Possession of Weapons

MINOR

- Defiance/Disrespect/ Non-compliance
- Disruption
- Dress Code Violation
- Inappropriate Language
- Other
- Physical Contact/ Physical Aggression
- Property Misuse
- Tardy
- Technology Violation
- Unknown

Source: 2009-10 Referral Definitions
www.swis.org



FLEXIBLE

Most common reasons for ODR within the SWIS database:

- Aggression/Fighting (major)
- Disrespect (minor and major)
- Physical Contact (minor)
- Disruption (minor and major)

Note: Patterns fairly consistent across elementary and middle although aggression drops in middle. However, most common High School categories include *disrespect*, *skip class*, and *tardy*





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

FLEXIBLE / EFFICIENT

Suggestions that can limit *flexibility* yet enhance *efficiency*:

- Create operational definitions that are mutually exclusive
- Consistently implement defined consequences
- Set up a usable system for ODR reporting
- Regularly summarize and use data



FLEXIBLE / EFFICIENT

Returning to...

Irvin, Horner, Ingram, Todd, Sugai, Sampson, & Boland (2006)

Results suggested:

- Generally data entered at least weekly, by support staff
- Data entry required 10-60 min/wk
- Respondents endorsed effort as “low” to “medium”
- Respondents perceived SWIS ODR use as increasing efficiency and effectiveness of decision making
- Overall, elementary users slightly more positive than middle



Conclusion Slide

- **Defensibility** of ODR has received less direct attention than other methods
 - Especially related to individual monitoring
 - Much of “validity” work has been under auspices of SWPBS - Would results generalize?
- Extant nature coupled with electronic systems offer high **efficiency & repeatability**
- **Flexibility** likely best restricted to enhance defensibility and generalized comparisons



Conclusion Slide

Our Goal: Develop and establish evidence-based assessment for formative purposes focused on promotion of social, behavioral, emotional functioning

“Caveats” to ODR Defensibility

- limited in “individual” prevention
 - lack: sensitivity, prosocial
- similar to other methods, role of “perception” to be considered





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

DISCUSSION & QUESTIONS

Frank Gresham, Ph.D.

Department of Psychology, Louisiana State University



General Outcome Measures (GOMs)

- Technically adequate (reliability & validity evidence)
- Sensitive to short-term changes in behavior
- Time-efficient to monitor performance 1 or 2 times per week
- Reflect important construct of interest (GOM)
 - CURRICULUM BASED MEASUREMENT
- Based on 25 years of research
- CBM is gold standard GOM for academic performance
- Used to measure rate of growth (slope) & level of performance
- Data used to maintain, change, intensify or terminate intervention
 - WE HAVE NO ACCEPTED GOMs FOR SOCIAL BEHAVIOR



Why are GOMs Important for Progress Monitoring & Decision-Making?

- Need to determine student rate of progress
- Need to decide if acceptable level of performance realized in specified period of time
- Need to identify "adequate" & "inadequate" responders using empirical data
- Need to establish valid decision rules for the above
 - The Problem
- We don't have well-established, empirically
- defensible criteria to make the above decisions



Review of Progress Monitoring Tools

• Systematic Direct Observations

- Considered by many to be gold standard in behavioral assessment
- Highly sensitive in detecting intervention effects
- Direct measures of behavior (time/place of occurrence)
- Measures multiple dimensions of behavior (frequency, duration, intensity)
- Repeated measurement of behavior over time
- Idiographic (individual)
- Treatment validity

- DRAWBACKS

- Ensuring representativeness of observations
- No benchmarks or normative standards to assist judgments
- Sources of error (setting, observer, time, etc.)
- Generalizability of observations (Hintze & Matthews, 2004)
 - Showed that acceptable reliability ($r=.90$) can only be obtained when students are observed 4 times per day, for 4 school weeks (20 days) constituting 40 hours of observation



Review of Progress Monitoring Tools

• Behavior Rating Scales

- Frequently used in schools & clinics to assess social behavior
- Quantifies information about behavior
- Extensive psychometric data available
- Multiple informants can assess same individual
- Extensive normative criteria upon which to base decisions

- DRAWBACKS

Indirect form of assessment

Measures perceived frequency of behavior

Decontextualized nature of behavior

Does not identify causes (antecedents/consequences) of behavior

CANNOT BE USED AS SHORT-TERM PROGRESS MONITORING TOOLS



Review of Progress Monitoring Tools

- Daily Behavior Reports

- Hybrid assessment tools (direct observations/behavior ratings)
- Collected multiple times per day (repeatable)
- Time- and resource-efficient tools
- Can be used as progress monitoring tools

- DRAWBACKS

- Reliability and validity evidence currently lacking
- Correlation between observations & DBRs about .50-.60
- No normative data or benchmarks upon which to make a decision
- Poor interrater reliability (based on *G* Studies)
- DBRs may not be generalizable across assessment conditions (raters, settings, times)





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Review of Progress Monitoring Tools

- Permanent Products

- Efficient
- Easily understood
- Flexible
- Repeatable
- Relevant
- Low reactivity

DRAWBACKS

Not a measure of actual behavior (indirect)
Subject to observer biases/expectations
ODRs can be manipulated by school policy changes
Limited validity evidence





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

An Alternative Progress Monitoring Tool

- Brief Behavior Ratings

- Change sensitive items on traditional rating scales
- Used frequently in ADHD literature (medication effects)
- Based on *clinimetric principles* (APGAR score)
- 3 characteristics of clinimetrics
 - Sensitivity to change
 - Stability over time (test-retest reliability)
 - Interrater reliability (interobserver agreement)





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Brief Behavior Ratings:

Some Examples Across Response Classes

Rater the following on a 1-10 Scale: 1-Never and 10 Almost Always

- Cooperation
 - Follows your directions
 - Pays attention to instructions
 - Follows classroom rules
- Self-Control
 - Stays calm when teased
 - Makes compromises during conflicts
 - Stays calm when disagreeing with others
- Externalizing
 - Has temper tantrums
 - Verbally abusive with others
 - Disobeys rules or requests
- Hyperactivity/Inattention
 - Fidgets or moves too much
 - Is inattentive
 - Breaks into or stops group activities



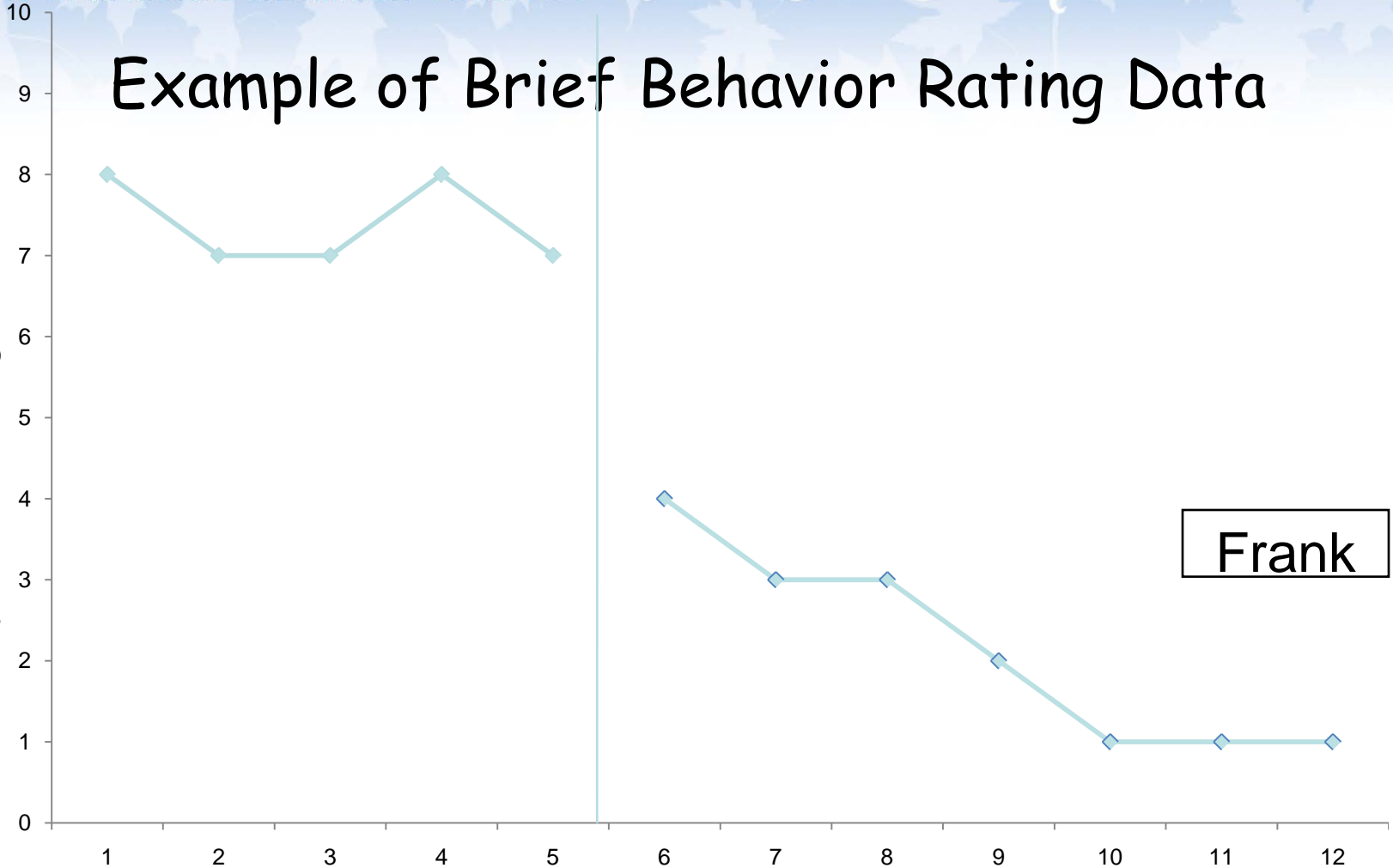
TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Example of Brief Behavior Rating Data

Disobeys/Challenges Teacher



Frank

Days





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Conclusions

- No universally accepted GOMs for social behavior
- No well-established valid decision rules for judging RTI
- SDOs considered **gold standard** but have problems (representativeness, expensive, multiple sources of error, absence of benchmarks)
- Behavior ratings psychometrically well established, but have problems (indirect, de-contextualized, insensitive to short-term changes in behavior)
- DBRs time, resource-efficient but have problems (psychometric inadequacies, lack of normative, benchmark data, multiple sources of error)
- Permanent products efficient, flexible, repeatable, but have problems (not direct measure of behavior, subject to biases, limited validity evidence)
- BBRs have potential & established in other areas (ADHD), but currently lack psychometric data & use across other domains of social behavior





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Take Away Message

There is a lot of work to be done!





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Data-Based Decision Making: What Tools Do We Have?

- Visual Analysis
 - Interocular test of significance
 - Conservative (low Type I error & higher Type II error rates)
 - Low interrater agreements using visual analysis
 - No standard by which comparisons can be made ($p < .05$)
 - The "eyes" don't necessarily have it
- Reliable Changes in Behavior
 - Absolute change indices
 - Amount of change from baseline to post-intervention levels
 - Individual no longer meets diagnostic criteria
 - Total elimination of behavior problems
 - Reliable change index (RCI) Post-Pre/ S error of difference (using stability)
 - Percent change from baseline (compares median baseline to median intervention)
 - Percent nonoverlapping data (not really index of strength of effect)
 - Effect size (modification of Cohen's D)
- Changes on Social Impact Measures (dropout, arrest rates, suspensions)
- Social Validation



A vertical illustration of the CN Tower, showing its distinctive glass and steel structure, with a blue and white color scheme. It is positioned on the left side of the slide, extending from the bottom to the top.

TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Conclusions

- No universal standard exists for data-based decision making
- Different metrics have advantages & disadvantages
- Visual analysis is not necessarily the *gold standard*
- Reliable change indices require reliability estimates (stability)
- Effect size estimates often inflated & uninterpretable
- PND does not really index strength of response
- No extant benchmarks for social behavior





TORONTO

APA 117TH ANNUAL CONVENTION • AUGUST 6-9

2009

Thank You!

Sandra M. Chafouleas, Ph.D.

sandra.chafouleas@uconn.edu

Amy M. Briesch, Ph.D.

abriesch@gmail.com

T. Chris Riley-Tillman, Ph.D.

rileytillmant@ecu.edu

Theodore J. Christ, Ph.D.

tchrist@umn.edu

Robert J. Volpe, Ph.D.

r.volpe@neu.edu

Frank Gresham, Ph.D.

frankgresham@yahoo.com

