Validating Progress Monitoring in the Context of RTI

Gerald Tindal • Joe Nese • Leilani Saez • Julie Alonzo
Behavioral Research and Teaching
University of Oregon

Abstract

- The purpose of this panel is to provide an open forum and debate-type discussion on measurement components for response-to-intervention (RTI) models. A key focus of the session will be criteria to consider in evaluating the appropriateness of RTI measurement systems, referencing the National Center on Response to Intervention and the *Standards for Educational and Psychological Testing* (1999). This session will systematically highlight the complexity of RTI and the tension among measurement, student progress, and appropriate interventions that represents the very core of the validation process that is iterative and argument-based.
- Although CBM preceded the most recent *Standards* (1999), where the focus has systematically been on establishing validity as an instructional argument, it is consistent with the conception of validity promulgated by Messick (1995). "Validity is an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the *adequacy* and *appropriateness* of *interpretations* and *actions* based on test scores or other modes of assessment…Validity is not a property of the test or assessment as such, but rather of the meaning of the test scores…and a function of the persons responding as well as the context of the assessment" (p. 5). Much of the early research on curriculum-based measures directly addressed this perspective through a series of studies on decision-making. This perspective is largely absent from the current focus on the validation process with the use of measurement systems.
- In the end, "fully determining the validity of an assessment process transcends what any one researcher can accomplish. It is a task for a community of researchers and practitioners to consider meanings and utility of assessment procedures in relation to current thinking about how to improve instructional practice and issues raised by studies of implementation" (Gersten, Keating, & Irvin, 1995, p. 512).

Acknowledgements

- The legacy of research begun with pioneers Stan Deno, Phyllis Mirkin and their colleagues Naomi Zigmond and Joe Jenkins, that has continued in immeasurable ways (irony intended) from Doug and Lynn Fuchs with the Vanderbilt contingency, and the Mafioso from Florida, Minnesota, Texas (several families), and Oregon, as well as many, many others over the decades of research.
- Researchers from BRT including Julie Alonzo, Joe Nese, Leilani Saez, Daniel Anderson, Shawn Irvin, Cheng Fei Lai, Bitnara 'Jasmine' Park, Paul Yovanoff, and Aki Kamata as well as programmers and staff: Kirt Ulmer, Aaron Glasgow, Raina Megert, and Denise Swanson.
- Office of Special Education Programs and Institute of Education Sciences (NCSER).

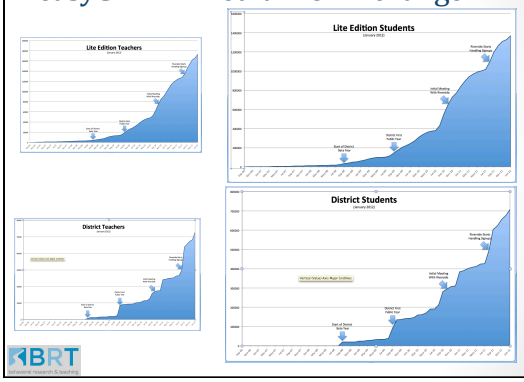
Original Conditions and Emphases of CBM

- Technically adequate: Must be valid…
- Sensitive to relatively small adjustments made in: instructional methods and materials, motivational techniques, administrative arrangements
- Easy to develop and administer
- Alternate forms available to administer frequently
- Time efficient
- Inexpensive
- Unobtrusive
- Simple to teach

Three Contestable Statements

- The *Standards for Educational and Psychological Testing* are frequently ignored in our curriculum-based measurement systems or their evaluation.
- Even in an argument-based approach to validation, privilege is given to measurement over decision-making and within our measurement research, we underplay scaling and standards (as common core standards).
- We have a research-to-practice perspective with little attention to uptake; if we begin with practice and understand it well, our research may have a better shelf life.

easyCBM: A Medium of Exchange



Messick ala 1995

- "Validity is an overall **evaluative judgment** of the degree to which **empirical evidence** and **theoretical rationales support** the ***adequacy*** and ***appropriateness*** of ***interpretations*** and ***actions*** based on test scores or other modes of assessment.
- **Validity is not a property of the test** or assessment as such, **but** rather of the **meaning of the test scores**…and a **function of the persons** responding as well as the **context of the assessment.**

Messick, S. (1995). Standards of validity and the validity of standards in performance assessments. *Educational Measurement: Issues and Practice. Winter*, 5—8.

Educational Standards ala 1999

- "Validity refers to the degree to which **evidence and theory support the interpretations** of test scores entailed by the **proposed uses** of tests."
- "The **proposed interpretation refers to** the **construct** or concepts the test is intended to measure."
- Clarified by **propositions** that **support proposed interpretations**, each of which may require different types of evidence.

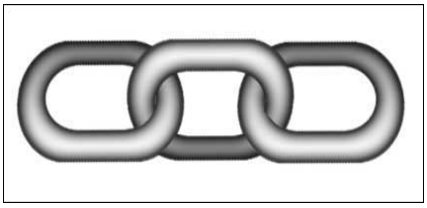
AERA, APA, NCME (1999). *Standards for Educational and Psychological Testing* (Page 9).

Gersten ala 1995

- "Fully determining the **validity** of an assessment process **transcends what any one researcher can accomplish**. It is a task for a **community of researchers and practitioners** to **consider meanings** and **utility** of assessment procedures in relation to current **thinking about how to improve instructional practice and** issues raised by studies of **implementation**" (p. 512).

Gersten, R., Keating, T., & Irvin, L. K. (1995). The burden of proof: Validity as improvement of instructional practice. *Exceptional Children*, *61*(6), 510-519.

A Model with a Dilemma



Measurement Sufficiency | Instructional Responsiveness | Decision Making

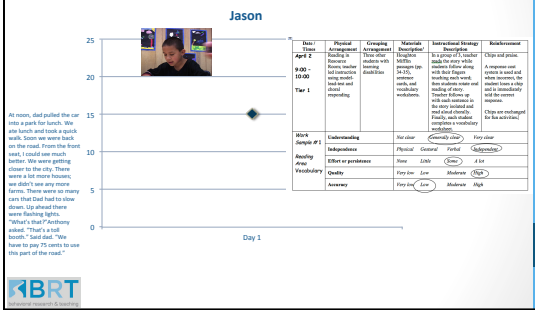
Concepts Behind the Model

- **Measurement Sufficiency**
  - Reliable: 'Reproducable'
  - Sensitive level of difficulty
  - Adequate number of occasions
  - Appropriate distribution of occasions
- **Instructional Responsiveness**
  - Sufficient components like grouping, curriculum, instructional presentation (strategies and models), error correction, reinforcement (with guided and independent practice), etc.
  - Fidelity ('reliability' is not just a measurement term)
- **Decision-making**
  - Catalyst (slope, variability, level, overlap)
  - Consequence (shift happens)

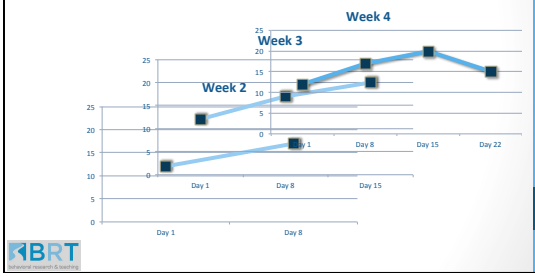
Delimiters of the Model

- **Response to Intervention (RTI) is not the issue**
  - Several authors have developed comprehensive models
  - Systems issues are part of most RTI models
- **The quality of research on CBM is not the issue**
  - Considerable high quality research has been done on measurement and instruction
  - The variables that have been addressed are important
- **The direction of research on CBM is an issue**
  - Understanding the reading (mathematics) learning process and all of its components (in reading, variables like morphology, syntax, grammatical structures, language, word meaning, etc.)
  - How teachers use (multiple) measures and connect the dots is critical
- **The research (designs) we employ with CBM is an issue**
  - Very few ideographic designs have been used
  - Little attention has been devoted to scaling or quasi-experimental threats to validity in the context of teacher decision making in practice

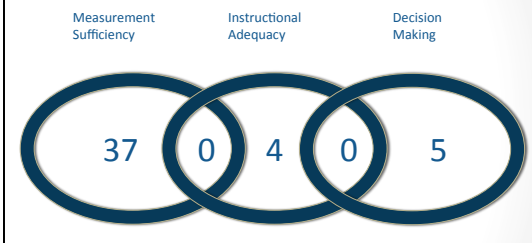
The First Link: Measurement and Instruction



The Second Link: Instruction and Change = Data-Based Decision-Making



Emphasis in Research



List of Studies

Author	Mvmt	Instruct	Decision	Note 1
Albano & Rodriguez (2011)	√			
Ardoin (2006)		√		Single subject study
Ardoin and Christ (2008)	√			
Ardoin and Christ (2009)	√			
Barth, Stuebing, Anthony, Denton, Mathes, Fletcher, & Francis (2008)	√			
Betts, Pickart, & Heistad (2009)	√			
Boscardin, Muthe'n, Francis, & Baker (2008)	√			
Bradley-Klug, Shapiro, Lutz, & DePaul (1998)	√			
Burns & Senesac (2005)	√			
Burns, Scholin, Kosciolek, Livingston (2010)	√			
Capizzi & Fuchs, 2005			√	
Christ (2006)	√			
Christ & Silberglitt, (2007)	√			
Christ and Ardoin (2009)	√			
Christ, Silberglitt, Yeo, and Cormier (2010)	√			
Clements, Bolt, Hoyt, Kratochwill (2007)	√			
Codding, Skowron, & Pace (2005)			√	
Deno, Fuchs, Marston, Shin (2001)	√			
Francis, Santi, Barr, Fletcher, Varisco, Foorman	√			
Fuchs, Fuchs, Hamlett, Walz, & Germann, (1993)	√			
Good, & Shinn (1990)	√			
Graney et al., (2009)	√			
Griffiths, VanDerHeyden, Skokut, Lilles (2009)	√			
Hintze and Christ (2004)	√			
Hintze, Daly, Shapiro, (1998)	√			
Hintze, Owen, Shapiro, Daly (2000)	√			
Jenkins, Graff, & Miglioretti	√			
Jenkins, Zumeta, & Dupree (2005)	√			
Johnson, Jenkins, & Petscher (2010).	√			
Kerr, Marsh, Ikemoto, Darilek, & Barney (2006)			√	
Kim, Petscher, Schatschneider, & Foorman (2010)	√			
McMaster, Fuchs, Fuchs, & Compton, D. (2002)		√		Controlled intervention
Mellard, McKnight, Woods, (2009)	√			
Nese, Biancarosa, Anderson, Lai, & Tindal (In press)	√			
Petscher & Kim (2011)	√			
Pansy, Skinner, & Axtell (2005)	√			
Reschly, Busch, Betts, Deno, & Long (2009)	√			
Roehrig, Duggar, Moats, Glover, & Mincey (2008)			√	No student outcomes
Schatschneider, Wagner, & Crawford (2008)	√			
Shin, Deno, & Espin, (2000)	√			
Shinn (2007)	√			
Silberglitt & Hintze (2005)	√			
Stecker, Fuchs, & Fuchs, (2005)			√	Critical review study; dated
Vaughn, Linan-Thompson, & Hickman (2003)		√		Conference presentation
Wagner, McComas, Bollman, & Holton, (2006)		√		Single subject design
Wei, Blackorby, & Schiller (2011)	√			

The Dilemma

- The technical template to judge curriculum-based measurement emphasizes a traditional measurement model for validation rather than a construct/ interpretation model.
- The research on curriculum-based measurement articulates ideal and controlled conditions with limited understanding of practice in use and almost no ideographic studies in the literature.
- Practice in use is undisciplined with insufficient data on the intersection of measurement sufficiency, instructional responsiveness, and decision making.

The Link Between Measurement Sufficiency & Instructional Decision-Making

Response to Intervention (RTI)

- Components
  - Assessment
  - Intervention
- Schools use assessments for:
  - Screening
  - Progress monitoring
  - Evidence-based instruction

What We Know We Need

- Good Reliability
  - Alternate form / test-retest
  - Generalizability
  - Measurement error
- Good Validity
  - Content
  - Construct
  - Convergent / Discriminant
  - Generalization
  - Consequences of testing

National Center on Response to Intervention (NCRTI)

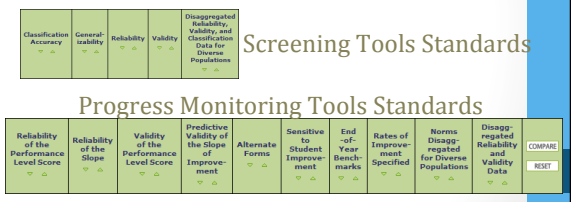
www.rti4success.org

- Technical review process to "determine which tools are deemed scientifically valid and appropriate."
- Clearinghouse for assessment systems.

NCRTI Review System

www.rti4success.org

Chart for tool review, displaying the technical standards.



What We Know

- Messick (1995)

  The importance of "the appropriateness, meaningfulness, and usefulness of score-based inferences" whose "power is derived from empirically grounded score interpretation."

What We Know

- VanDerHeyden (2011)

  "To demonstrate that RTI implementations are technically adequate, each assessment must be technically adequate for the purpose for which it is used."

What We Need to Know

- Measurement Sufficiency
- Growth Trajectories
- Instructional/Intervention Information

Measurement Sufficiency

- Number of testing occasions
- Time between testing occasions
- Different initial achievement levels
- On/off grade-level
- Assessment onset

Frequency of Progress Monitoring

- Daily
- At least three times per week
- Twice per week
- Weekly or biweekly
- Biweekly data collected across 10 weeks
- Every 3 or 4 weeks
- At least 10 data points
- Every 9 weeks
- 20 data points collected across 3 months

(Ardoin, 2004; Christ, 2006; Deno, 1985; Deno, Lembke, & Reschly, n.d.; Fuchs, Fuchs, & Hamlett, 1989; Jenkins et al., 2009; Mirkin et al., 1981; Shinn, 2002)

Instructional Aim Lines

- Aim line (or goal line): the projected amount of growth across time that is established as a minimum for ***adequate progress***.
  - Generally scaled as weekly growth.

Researchers	Teachers
Norm-references	Norm-references
OLS Regression	Professional judgment

Rules of Thumb

- Monitor progress for 4 weeks and at least 9 data points
  - *4 consecutive points below the aim line*
- Monitor progress weekly or biweekly for 6 weeks
  - *3 consecutive data points below the aim line*
- Sufficient progress = 3-5 consecutive data points above the aim line
- Insufficient progress = 3-5 consecutive data points below the aim line

(Burns, Scholin, Kosciolek, Livingston, 2010; Deno Lembke, & Reschly, n.d., as cited in Jenkins & Terjeson, 2011; L. Fuchs, Fuchs, Hintze, & Lembke, 2006; Marston and Tindal, 1996, as cited in Ardoin, 2004; Mirkin, Deno, Tindal, & Kuehnle, 1982; Shinn, 1989).

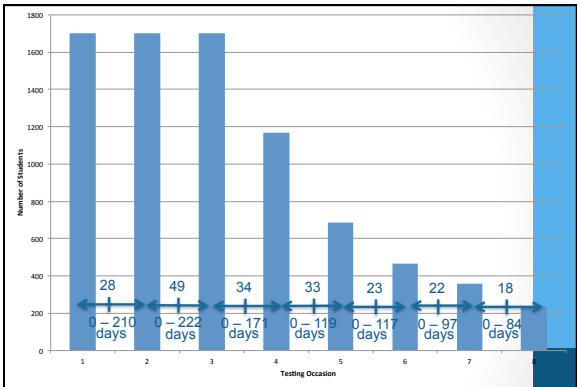
Progress Monitoring in Action

Grade 3 Reading Comprehension

Number of Testing Occasions	Frequency	Percent
1	1553	37.5
2	883	21.3
3	533	12.9
4	484	11.7
5	217	5.2
6	110	2.7
7	119	2.9
8	41	1.0
9	51	1.2
…	…	…
23		
Total	4137	100

Progress Monitoring in Action

- **1371** – students test only ON grade 3 level
- **330** – students tested OFF grade level
  - 22% tested at grade 2
  - 78% took tests from multiple grades
- **40%** scored **ABOVE** the 50th percentile on their first assessment



Measurement Sufficiency

- Empirical evidence to determine appropriateness.
  - Different constructs and measures (e.g., letter sounds, fluency, comprehension, math)
  - Frequency and distribution of testing occasions
  - On- and off-grade level testing
  - Different for initial achievement levels
  - Considering measurement error

Growth Trajectories

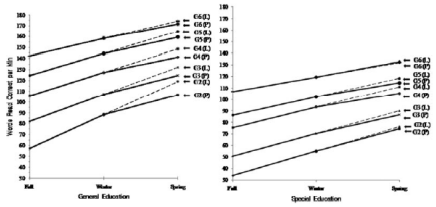
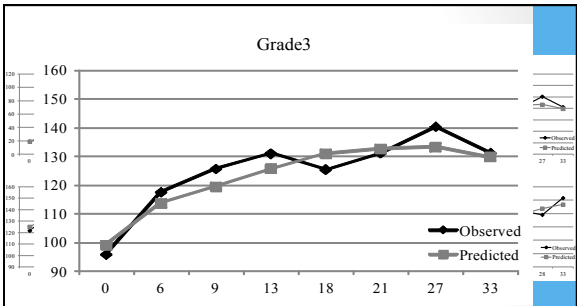


Figure 1. Differences for linear (L, solid line) and piecewise (P, dashed line) models of growth (G) for students in second through sixth grade within either the general (left) or special education population (right)

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-based measurement of oral reading: An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review, 39*, 447–462.



Within-year mean oral reading fluency (ORF) scores by grade

Observed mean

Predicted mean

Nese, Biancarosa, Cummings, Kennedy, Alonzo, & Tindal

Growth Trajectories

- Non-linear
- But our growth rates are from OLS or HLM, which estimates a **mean growth** for the sample, and assumes:
  1) all students within the sample are from the same population, and
  2) a single mean growth trajectory describes an entire population
  3) covariates that affect growth influence each student the same way
- Trajectories for different groups of students based on initial starting point or instructional program.
- What does growth look like for students receiving Tier 2 and Tier 3 intervention?

Instructional/Intervention Information

- Instructional information to enhance growth models
  - Variables for effective instruction
- Sensitive to student improvement
- Reliability
- Classification Accuracy

Merely *detecting* low achievement is not sufficient. Low achievement demonstrates a spectrum of behavior. To address the behavioral heterogeneity of low achievement, the screening test must go further. It must increase survival.

- Siddhartha Mukherjee, 2010

Summary

- What do we know about students' development of academic competence across different skills?
- Do we know enough about the instruction and the effects of intervention on skill development?
- How does assessment information link to intervention?
- How meaningful are growth results unless we know the instruction that a student or group is receiving?

Instructional Responsiveness: What are Teachers Doing?

Leilani Sáez, Ph.D.
Behavioral Research and Teaching
College of Education – University of Oregon

Assessment-to-Instruction Relation

"What is less clear is exactly how much instruction must occur, how contextualized skills instruction needs to be, and the level of intensity at which it must occur in order for struggling readers to succeed."

(Mathes, Denton, Fletcher, Anthony, Francis, & Schatschneider, 2005, p.151)

Issues Raised

- Timing & Use of data for measurement
- Criterion for learning as inadequate
- Nature of the intervention (intensity, duration, delivery format, group size, instructor)
- Nature of Tier 1 instruction
- Treatment Fidelity

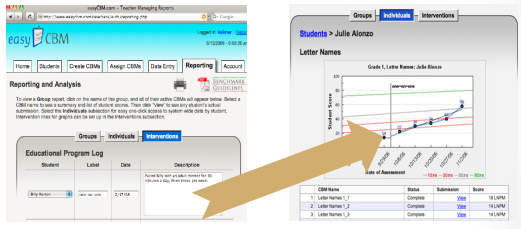
(D. Fuchs & Deshler, 2007; L Fuchs, 2003; Griffiths, VanDerHeyden, Parsons, & Burns, 2006; Kovaleski, 2007)

"Although we know much about early intervention for many students at risk for reading difficulties, the question remains:
What instruction is used to assist students who have demonstrated low response

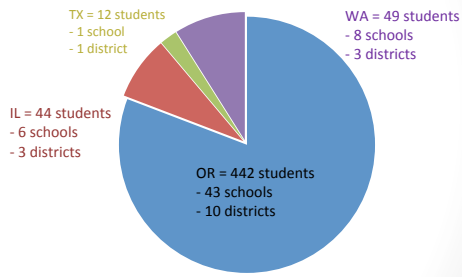
??

A Descriptive Report Of Responsive Instruction Practices
Among 17 Districts

Background



2010-2011 Sample



District Demographics

District Location & Population	Student Race/Ethnicity				2009 Median Income
	Asian	Black	Hispanic	White	
TX (2,366)	0%	<1%	26%	62%	$58,057
IL (19,651)	2%	<1%	12%	84%	$91,156
OR-A (153,231)	3%	2%	11%	70%	$39,640
OR-B (15,325)	<1%	<1%	31%	47%	$46,831
WA (42,046)	6%	5%	13%	57%	$52,868

A Tale of 3 Districts: Tier 1 Policy

	District A (2010) ✓	District B (2008) ✓	District C (2010) ✓
Focus	Use of "scientific-based reading instruction curriculum emphasizing the 5 critical elements of reading"	Provide "all students access to the general education curriculum focusing on the 5 essential components of reading"	To "teach skills and strategies which address the 'Big Five'"
Time	90 min per day (K-3) 60 min per day (4-5)	90 min per day (60 min whole class, 30 min small group)	90 min per day (60 min whole class, 30 min small group)
Assm't	"Universal screening" 3x a year	"Benchmark testing" 3x a year	"Universal screening" 3 x a year

A Tale of 3 Districts: Tier 2 Policy

	District A (2010) ✓	District B (2008)	District C (2010)
Focus	Use of "research-based interventions" to support lowest 20th percentile	Provide "differentiated and skill focused" instruction lowest 20th percentile	Use of "supplemental/ targeted instruction" for "additional practice" to students performing @ lowest 20th percentile
Time	+30 minutes	+30 minutes (120 min total)	+30 *to* +60 extra minutes
Assmt	PM every 2 weeks	PM for at least 6 weeks, with 3 data points (*although not required*)	PM every 1-6 weeks, with minimum of 4 data points needed to make decisions
Check for Non-response	4-6 points below aimline; slope is flat or decreasing	If "measured achievement falls below aimline"; or "flat progress trend"	After 4 data points, "make an instructional change or continue to monitor"

A Tale of 3 Districts: Tier 3 Policy

	District A (2010)	District B (2008)	District C (2010)
Focus	Provide "sustained, intensive scientifically based instruction for students with marked difficulties" with or w/out IEPs ✓	Provide comprehensive core + 60 minutes of targeted instruction	Use of "supplemental/ targeted instruction" to provide "additional practice" to students performing @ lowest 20th percentile
Time	+ 30 minutes per day	+ 60 minutes per day over 6 weeks	+30 *to* +60 extra minutes
Assmt	PM every week	PM every 2 weeks ✓	PM every 1-6 weeks, with minimum of 4 data points needed to make decisions
Check for Non-response	4-6 points below aimline; slope is flat or decreasing	If "measured achievement falls below aimline"; or "flat progress trend"	After 4 data points, "make an instructional change or continue to monitor"

Reported Intervention Frequency

(*N* = 547 4th grade students)

Total Number of Reported Reading Interventions Implemented Across School Year	Student *N*
1	**355 (64.9%)**
2	138 (25.2%)
3	25 (4.6%)
4 or more	29 (5.3%)

Reported 1st Intervention Curricula

CARS/STARS (Comprehensive Assessment of Reading Strategies/ Strategies to Achieve Reading Success)

Corrective Reading	Explode the Code
Harcourt Intervention Station	Horizons
Open Court Intervention Kit	Triumphs
Read Naturally	Reading Mastery
Rewards	Score4Reading
SLANT (Structured Language Training)	

Intensive = 66.5%

Soar to Success
Study Island
Step Up to Writing
Various trade consumable workbooks
Various leveled texts

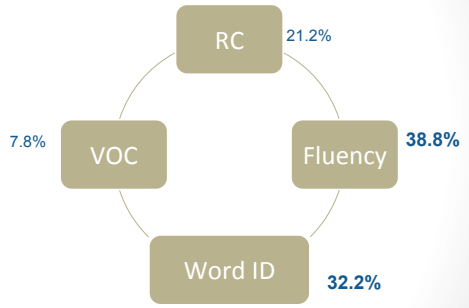
Strategic = 30.9%

Houghton Mifflin
Treasures

Core =2.6%

KBRT

1st Intervention Skill Emphasis



KBRT

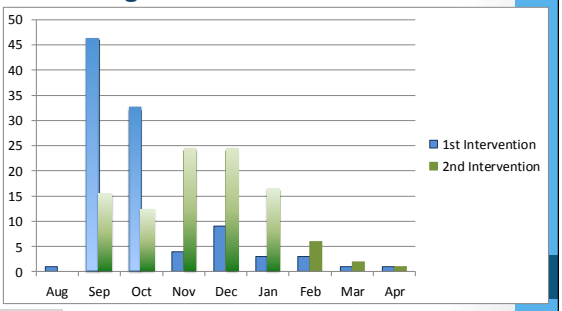
Reported Intensity of 1st Intervention

Number of Days Per Week	Percent of Students
5	23.4%
4	**51.1%**
3	9.7%
2	15.5%
1	1.1%

Number of Minutes Per Day	Percent of Students
More than 60	4.4%
60	13.9%
30-59	**59.3%**
Less than 30	21.9%

KBRT

Percent of Reported Interventions Occurring Across the School Year



Reported Instructional Changes Between 1st and 2nd Interventions

Intervention Change	Percent of Total Intervention Changes
Instructional Program/Curricula	**50.0%**
Intensity (duration and/or frequency)	18.7%
Tier or Teacher	6.1%
Group Size	6.1%
Addition of Progress Monitoring	1.3%

Summary

- Limited evidence of reading intervention change across the school year in 4th grade
- 1st intervention most likely to be intensive and focused on building word identification & fluency
- Interventions were implemented within first 5 months of schooling
- Most frequent changes to the 1st intervention were curricular/ program based

Some Lingering Questions

What do we *really* know about students' opportunity to learn within a typical classroom?

- To what extent are interventions implemented as intended?
- How well do we understand what it means to "intensify" instruction? What's driving intervention "change"?
- Do we have a good handle on expected rates of skill development under different instructional conditions?
- How well do we understand what's going on in the 2nd half of the school year and its impact on expectations for growth?

Data-based Decision Making: Practice to Research

Julie Alonzo, Ph.D.
Behavioral Research and Teaching
College of Education – University of Oregon

Data-Based Decisions

Practice in use is undisciplined with insufficient data on the intersection of measurement sufficiency, instructional integrity, and data-based decision making.

What is actually happening out in the schools?

Solid, research-based decision rules?

Clear understanding of what to do and why?

Commitment to measurement principles?

Data-Based Decisions

Julie, I'm an aide and also a parent and recognized that my child's teacher is printing out progress monitoring to give to parents to practice. I told her that it wasn't meant for that type of use. I continued to say that these progress monitoring is to see if the kids are internalizing what they learned and see if they can apply it. The teacher's response was this .... "We have heard that after a progress monitoring has been administered it is acceptable (even encouraged) to use it as a teaching tool. Based on that information we decided to use a probe that will not be used as a progress monitoring measure as a teaching tool (we chose PM-16...we won't get even close to 16 progress monitors in each area). We of course make sure NOT to show students a probe that we would actually be using to assess their progress. The reason for sending home anything at all is to allow students to become familiar with the testing format. Many students are intimidated by a one minute timing and we want them to be comfortable with the test format and procedure to get a true result of their knowledge. In addition, a visual of the testing format is very helpful to the parents (most of whom don't have a background in education and can't imagine one minute timed reading tests for kindergarten).

Note: Actual e-mail received Jan. 24, 2012

Data-Based Decisions

Practice in use is undisciplined with insufficient data on the intersection of measurement sufficiency, instructional integrity, and data-based decision making.

As a research community, we have said:

- CBM progress monitoring integrates psychometric principles with single subject research design methodology (Deno, Fuchs, Marston, & Shin, 2001).
- Data for each individual are presented on a separate line graph: data are collected repeatedly, graphed regularly, and analyzed frequently to make data-based decisions on an on-going basis (Gast, 2010).

Data-Based Decisions

Single subject research design:

- Includes a baseline phase that provides repeated measurement prior to intervention to establish a pattern that can be used to compare post-intervention trajectory (Gast, 2010; Horner, Carr, Halle, McGee, Odom, & Wolery, 2005).
- Decisions to maintain or change the intervention are made in accordance with visual-analysis guidelines to evaluate data of individuals or small groups (Gast, 2010).

Data-Based Decisions

In RTI applications:

Progress monitoring graphs should include goal lines, or aim lines. Decision rules for intervention change, including service eligibility, are based upon these goal lines in concert with an empirically sound instructional and decision-making sequence (Burns, Scholin, Kosciolek, & Livingston, 2010; Barnett et al., 2006).

Some researchers provided explicit guidelines for decision-making, such as monitoring progress weekly or biweekly for 6 weeks, and evaluating the adequacy of progress according to the rule that instructional changes should be made if three consecutive data points fall below the goal line (Deno Lembke, & Reschly, n.d., as cited in Jenkins & Terjeson, 2011).

Data-Based Decisions

In RTI applications:

Some researchers defined sufficient progress as three to five consecutive data points above the aim line, and suggested that either a more ambitious goal was needed or intervention termination should be considered, and defined insufficient progress as three to five consecutive data points below the aim line, and suggested that the intervention needed to be modified (Burns, Scholin, Kosciolek, Livingston, 2010; L. Fuchs, Fuchs, Hintze, & Lembke, 2006; Mirkin, Deno, Tindal, & Kuehnle, 1982; Shinn, 1989).

Jenkins and Terjeson (2011) suggested that the "points-below" rules, which rest on the tenuous hypothesis of linear growth, should be abandoned in favor of computational models of students' growth.

How Often To Assess?

Dear Help Desk,

We need some advice. We have administered all alternate forms of the CBMs on your site last month and need to know what to do this month. Administering all 17 forms of all the measures took a great deal of time, and many of the students were upset by how much we were testing them. We understand that we need data to make good decisions, but this seems a little overboard!

--Teacher

NOTE: Paraphrase of an e-mail we received in 2010.

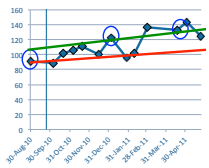
How Often To Assess?

Recommendations have included daily (Deno, 1985), at least three times per week (Mirkin et al., 1981), to twice per week (Fuchs, Fuchs, & Hamlett, 1989a), weekly or biweekly (Deno, Lembke, & Reschly, n.d.), every 3 or 4 weeks (Jenkins et al., 2009). To account for measurement error, a minimum of 10 data points is suggested as best practice (Shinn, 2002), or biweekly data collected across 10 weeks (Christ, 2006).

Jenkins, Graff, and Miglioretti (2009) compared growth slopes based on measurements taken every 1-4 and 9 weeks and found that the frequency of progress monitoring could be significantly reduced without detracting from the validity of growth estimates. One study included decision rules that the number of data points needed to make a reasonably valid estimate of a student's progress was closer to 20 data points collected across 2.5-3 months, and for a decision regarding a student's eligibility for special education, a total of 40 data points collected across 5-6 months (Ardoin, 2004).

Data-Based Decisions

Example 1: One intervention followed by increasing growth, no change to intervention



50th percentile

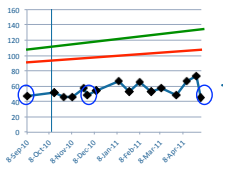
20th percentile

Note. Benchmark assessments are circled. Vertical line notes date of intervention.

Intervention 1	Rewards 2x/week for 20_x000D_ Strategic and Intensive kits for comprehension and vocabulary 2 x/week for 20 min_x000D_ Read Naturally 20 min 3x/week_x000D_

Data-Based Decisions

Example 2: One intervention followed by flat growth, no further interventions



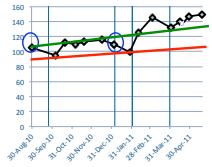
50th percentile

20th percentile

Intervention 1	Tier 3b (BEST): Skill focus: Phonics & Fluency. Using 4th grade Triumphs Materials & the Rewards program. 5 students taught by an IA. 30m/day, 4x a week.

Data-Based Decisions

Example 3: One intervention followed by flat growth, then additional intervention followed by some growth



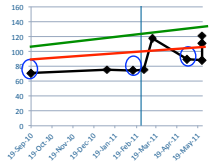
50th percentile

20th percentile

Intervention 1	Rewards 2x/week for 20_x000D_ Strategic and Intensive kits for comprehension and vocabulary 2 x/week for 20 min_x000D_ Read Naturally 20 min 3x/week_x000D_
Intervention 2	Group Change: Moved to ___'s group(smaller size). Doing Harcourt Intensive materials, Read Naturally 2x/week and Study Island 2x/week to practice fluency and comprehension.
Intervention 3	Study Skills: 20 min. of test taking practice and strategies
Intervention 4	Concerns noted with teacher about the lack of comprehension. Decided to wait a few more weeks to see if extra class helped out.

Data-Based Decisions

Example 4: Intervention perhaps inappropriately late in the year, then perhaps some growth after intervention



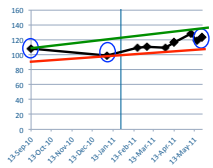
50th percentile

20th percentile

Intervention 1	Corrective Reading B-2: 40 min.4x/week;group of 6

Data-Based Decisions

Example 5: Intervention appropriately late (given high fall benchmark score), some growth after intervention.



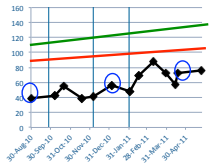
50th percentile

20th percentile

Intervention 1	Read Natural: Student goes to read naturally each morning during Tiger Time (30 mins).

Data-Based Decisions

Example 6: Many interventions, not enough PMs between each.

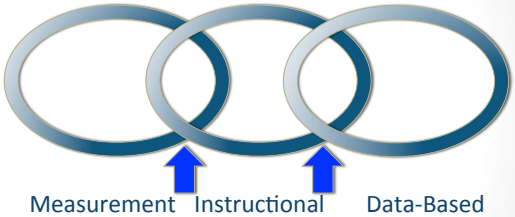


50th percentile

20th percentile

Intervention 1	Read Naturally 20 minute sessions four days a week. Harcourt Strategic comprehension skills 20 minute sessions 2 days a week.
Intervention 2	Phonics for Reading Level 3. 20 min. session four days a week.
Intervention 3	New Skill Block: Moving to __'s Skill Block. Sp. Ed.

Future Directions in Research



Measurement Sufficiency

Instructional Integrity

Data-Based Decision Making

Conclusions – Take 1

- A stronger presence is needed in the measurement community because at its root CBM is all about measurement sufficiency, a term that needs to include more clear invocation of the *Standards for Educational and Psychological Testing* in the argumentation of our research.
- A more articulate explication should provide the linkage between measurement sufficiency, instructional adequacy, and decision-making. The validation process then needs to focus on quasi-experimental research designs with stronger elements of causation and an ideographic footprint.
- Attention is warranted in the 'practice-to-research' component of the cycle, as current writing reveals a noticeable short sightedness that focuses on what teachers *should* do and nearly complete lack of attention to what they currently do or the contingencies that need to change.

Conclusions – Take 2

- Standard error is not limited to the instrument; it also applies to the environment (a measurement sufficiency editorial).
- Reliability and validity apply to both independent and dependent variables (a measurement sufficiency and instructional responsiveness editorial).
- Most of our shared instructional descriptions provide relatively vague descriptions of interventions; grouping, curriculum, and time are the most popular (an instructional adequacy editorial) .
- The vast unknown is how the behavior of students migrates to the actions of teachers (a decision making editorial).
- "We look at the present through a rear view mirror – We march backwards into the future" (from Marshall McLuhan).

The end?

http://www.brtprojects.org/