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. Reference to the National Center on Response to Intervention (2009) and the Standards for Educational and Psychological Testing (1999) is essential to systematically highlight the complexities 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 instrucional argument, it is consistent with the concept 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, & Inns, 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 Maffoos 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, Bithara ‘Jasmine’ Path, Paul Vovosoff, 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 (NCER).
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."


**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 of which may require different types of evidence."


**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)."
Concepts Behind the Model

- Measurement Sufficiency
  - Reliable: ‘Reproducible’
  - 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
At noon, dad pulled the car into a park for lunch. We ate lunch and took a quick walk. Soon we were back on the road. From the front seat, I could see much beier. We were getting closer to the city. There were a lot more houses; we didn’t see any more farms. There were so many cars that Dad had to slow down. Up ahead there were flashing lights. “What’s that?” Anthony asked. “That’s a toll booth,” said Dad. “We have to pay 75 cents to use this part of the road.”

The First Link: Measurement and Instruction

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

Emphasis in Research

<table>
<thead>
<tr>
<th>Measurement Sufficiency</th>
<th>Instructional Adequacy</th>
<th>Decision Making</th>
</tr>
</thead>
<tbody>
<tr>
<td>37</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
</tbody>
</table>
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.
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 of 10 data points
- Every 9 weeks
- 20 data points collected across 3 months

**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.

<table>
<thead>
<tr>
<th>Researchers</th>
<th>Teachers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norm-references</td>
<td>Norm-references</td>
</tr>
<tr>
<td>OLS Regression</td>
<td>Professional judgment</td>
</tr>
</tbody>
</table>

**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**

(Sources: Arreola, Hess, & Schumaker, 2010; Davis-Lemke & Deno, n.d.; as cited in Jenkins & Terjesen, 2011; L. Fuchs, Fuchs, Ham, & Ham, 2000; Marston and Tindal, 1996, as cited in Ardoin, 2004; Mirkin, Deno, Tindal, & Kuehnle, 1982; Shinn, 1989.)

**Progress Monitoring in Action**

<table>
<thead>
<tr>
<th>Grade 3 Reading Comprehension</th>
<th>Number of Testing Occasions</th>
<th>Frequency</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>553</td>
<td>37.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>881</td>
<td>21.3</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>533</td>
<td>12.9</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>484</td>
<td>11.7</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>217</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>130</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>119</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>41</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>51</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>4137</td>
<td>100</td>
</tr>
</tbody>
</table>
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

- 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
District Demographics

<table>
<thead>
<tr>
<th>District Location &amp; Population</th>
<th>Student Race/Ethnicity</th>
<th>2009 Median Income</th>
</tr>
</thead>
<tbody>
<tr>
<td>TX (2,366)</td>
<td>Asian: 0%</td>
<td>Black: &lt;1%</td>
</tr>
<tr>
<td>IL (19,651)</td>
<td>Asian: 2%</td>
<td>Black: &lt;1%</td>
</tr>
<tr>
<td>OR-A (153,231)</td>
<td>Asian: 3%</td>
<td>Black: 2%</td>
</tr>
<tr>
<td>OR-B (15,325)</td>
<td>Asian: &lt;1%</td>
<td>Black: &lt;1%</td>
</tr>
<tr>
<td>WA (42,046)</td>
<td>Asian: 6%</td>
<td>Black: 5%</td>
</tr>
</tbody>
</table>

A Tale of 3 Districts: Tier 1 Policy

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of “scientific-based reading instruction curriculum emphasizing the 5 critical elements of reading”</td>
<td>Provide &quot;all students access to the general education curriculum focusing on the 5 essential components of reading&quot;</td>
<td>To &quot;teach skills and strategies which address the &quot;Big Five&quot;”</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>90 min per day (K-3) 60 min per day (4-5)</td>
<td>90 min per day (60 min whole class, 30 min small group)</td>
<td>90 min per day (60 min whole class, 30 min small group)</td>
</tr>
<tr>
<td>Assm’t</td>
<td>&quot;Universal screening&quot; 3x a year</td>
<td>&quot;Benchmark testing&quot; 3x a year</td>
<td>&quot;Universal screening&quot; 3x a year</td>
</tr>
</tbody>
</table>
A Tale of 3 Districts: Tier 2 Policy

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Focus</strong></td>
<td>Use of &quot;research-based interventions&quot; to support lowest 20th percentile</td>
<td>Provide &quot;differential and skill focused&quot; instruction lowest 20th percentile</td>
<td>Use of &quot;supplemental targeted instruction&quot; for &quot;additional practice&quot; to students performing @ lowest 20th percentile</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>+30 minutes</td>
<td>+30 minutes (120 min total)</td>
<td>+30 to +60 extra minutes</td>
</tr>
<tr>
<td><strong>Assmt</strong></td>
<td>PM every 2 weeks</td>
<td>PM for at least 6 weeks, with 3 data points (although not required)</td>
<td>PM every 1-6 weeks, with minimum of 4 data points needed to make decisions</td>
</tr>
<tr>
<td><strong>Check for Non-response</strong></td>
<td>4-6 points below aimline; slope is flat or decreasing</td>
<td>If &quot;measured achievement falls below aimline&quot;, or &quot;flat progress trend&quot;</td>
<td>After 4 data points, &quot;make an instructional change or continue to monitor&quot;</td>
</tr>
</tbody>
</table>

A Tale of 3 Districts: Tier 3 Policy

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Focus</strong></td>
<td>Provide &quot;sustained, intensive scientifically based instruction for students with marked difficulties&quot; with or w/out IEPs</td>
<td>Provide comprehensive core + 60 minutes of targeted instruction</td>
<td>Use of &quot;supplemental targeted instruction&quot; to provide &quot;additional practice&quot; to students performing @ lowest 20th percentile</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>+ 30 minutes per day</td>
<td>+ 60 minutes per day over 6 weeks</td>
<td>+30 to +60 extra minutes</td>
</tr>
<tr>
<td><strong>Assmt</strong></td>
<td>PM every week</td>
<td>PM every 2 weeks</td>
<td>PM every 1-6 weeks, with minimum of 4 data points needed to make decisions</td>
</tr>
<tr>
<td><strong>Check for Non-response</strong></td>
<td>4-6 points below aimline; slope is flat or decreasing</td>
<td>If &quot;measured achievement falls below aimline&quot;, or &quot;flat progress trend&quot;</td>
<td>After 4 data points, &quot;make an instructional change or continue to monitor&quot;</td>
</tr>
</tbody>
</table>

Reported Intervention Frequency
(N = 547 4th grade students)

<table>
<thead>
<tr>
<th>Total Number of Reported Reading Interventions Implemented Across School Year</th>
<th>Student N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>355 (64.9%)</td>
</tr>
<tr>
<td>2</td>
<td>138 (25.2%)</td>
</tr>
<tr>
<td>3</td>
<td>25 (4.6%)</td>
</tr>
<tr>
<td>4 or more</td>
<td>29 (5.3%)</td>
</tr>
</tbody>
</table>
Reported 1st Intervention Curricula

- CARAS/STARS (Comprehensive Assessment of Reading Strategies/Strategies to Achieve Reading Success)
- Corrective Reading
- Explode the Code
- Horizons
- Open Court Intervention Kit
- Reading Mastery
- Reading with All Kids
- Reading with Style
- Science and Driving
- SPAN (Structural Language Training)
- Soar to Success
- Study Island
- Step Up to Writing
- Various trade consumable workbooks
- Various leveled texts
- Nauken Houghton Mifflin
- Treasures
- Soar to Success
- Study Island
- Step Up to Writing
- Various trade consumable workbooks
- Various leveled texts
- Nauken Houghton Mifflin
- Treasures

1st Intervention Skill Emphasis

- RC 21.2%
- VOCC 7.8%
- Fluency 38.8%
- Word ID 32.2%

Reported Intensity of 1st Intervention

<table>
<thead>
<tr>
<th>Number of Days Per Week</th>
<th>Percent of Students</th>
<th>Number of Minutes Per Day</th>
<th>Percent of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>23.4%</td>
<td>More than 60</td>
<td>4.4%</td>
</tr>
<tr>
<td>4</td>
<td>51.1%</td>
<td>60</td>
<td>13.9%</td>
</tr>
<tr>
<td>3</td>
<td>9.7%</td>
<td>30-59</td>
<td>59.3%</td>
</tr>
<tr>
<td>2</td>
<td>15.5%</td>
<td>Less than 30</td>
<td>21.9%</td>
</tr>
<tr>
<td>1</td>
<td>1.1%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 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; Homer, Carr, Hall, 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, Koscielek, & 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).

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.

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, Koscielek, Livingston, 2010; L. Fuchs, Fuchs, Hintze, & Lembke, 2006; Mirkin, Deno, Tindal, & Kauffman, 1982; Shinn, 1989).

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, 1988), weekly or biweekly (Deno, Lambke, & 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 (Chirls, 2009).

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

Data-Based Decisions

Example 2: One intervention followed by flat growth, no further interventions
Data-Based Decisions

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

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

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

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

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/