Assessment and Progress Monitoring: Applications at the System and Student Level

Leah M. Nellis, Ph.D., NCSP
Indiana State University
lnellis@indstate.edu
You can use all the quantitative data you can get, but you still have to distrust it and use your own intelligence and judgment.
We are encouraged by the potential of RTI models to improve educational opportunities for culturally and linguistically diverse students and to reduce their disproportionate representation in special education. At the same time, we are concerned that if we do not engage in dialogue about how culture mediates learning, RTI models will simply be like old wine in a new bottle, in other words, another deficit-based approach to sorting children, particularly children from marginalized communities.

NCCRESSt Position Statement on RTI
Yet, unless researchers and educators capitalize on known and evolving knowledge about appropriate assessments and interventions for culturally and linguistically diverse students, there is also concern that....those implementing the RTI model will presume that if a child does not make adequate progress, he or she must have an internal deficit of some kind or come from a deficit background (making underachievement something that cannot be helped).

Klingner, Sorrells, & Barrera, 2007
Topics for this Session

- Evaluating the effectiveness of RTI, with a culturally and linguistically diverse lens.

- Considering student need and progress through culturally responsive assessment.
The Flow of RTI

- Determine whether effective instruction is in place for *all* students, which includes culturally and linguistically diverse (CLD) students,
- Provide instruction/intervention to those in need and measure the effectiveness,
- Continue to provide additional or more intensive intervention as needed.

Requires us to disaggregate data and look critically at what it is telling us!
Guiding Questions

• What does the data tell us about ethnicity, SES, language proficiency, etc?

• Are CLD students making modest (or less) gains while majority students are making outstanding gain?

...which need be asked at a schoolwide or specific student level!

• How do “non-responders” differ from “responders” (considering both majority students and those in similar cultural and linguistic groups?)
Ways to Evaluate Effectiveness

- Levels across benchmark periods (schoolwide)
- Movement *between* tiers (schoolwide and student)
- Movement *within* tiers due to rate of improvement, i.e., Growth. (schoolwide and student)

Clemens & Shapiro, in press
Levels Across Benchmark Periods

- **Key question:** “What percentage of students scored within low-, some-, and at-risk levels across benchmark periods?”

- **Goal:**
  - Increase the percentage of students in the low risk category (Tier 1) across the year
  - Decrease the percentage of students in the at risk category (Tier 3) across the year

Clemens, Gischlar, Devlin, & Shapiro, 2009
## Comparing Fall to Winter

<table>
<thead>
<tr>
<th>2nd grade, Oral Reading Fluency</th>
<th>Fall</th>
<th>Winter</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Benchmark</td>
<td>42%</td>
<td>26%</td>
</tr>
<tr>
<td>% Strategic</td>
<td>37%</td>
<td>31%</td>
</tr>
<tr>
<td>% Intensive</td>
<td>21%</td>
<td>43%</td>
</tr>
</tbody>
</table>
Let’s look at some data....

Using the data on your tables, discuss

-What the data indicates about instruction in this school,

-What the data indicates about intervention in this school, and

-What patterns in the data you might see related to culturally and linguistically diverse learners.
Movement Between Tiers

- Key question: How many students moved to a less intensive tier between benchmark periods?
- Assumption: Improved instruction and intervention will enable students to accelerate their learning and move from higher to lower risk levels (e.g., Tier 3 to Tier 2; Tier 2 to Tier 1)
- To what degree are our students moving to lower risk levels?
  - Are we “losing” kids from Tier 1 or 2?
  - If so, who are they and why is this happening?
## What Changed from Fall to Winter?

### Benchmark in Fall

<table>
<thead>
<tr>
<th>Winter ORF</th>
<th>20 Benchmark</th>
<th>10 Strategic</th>
<th>4 Intensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>34 (42%)</td>
<td>Students</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Strategic in Fall

<table>
<thead>
<tr>
<th>Winter ORF</th>
<th>1 Benchmark</th>
<th>11 Strategic</th>
<th>18 Intensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 (37%)</td>
<td>Students</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Intensive in Fall

<table>
<thead>
<tr>
<th>Winter ORF</th>
<th>0 Benchmark</th>
<th>4 Strategic</th>
<th>13 Intensive</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 (21%)</td>
<td>Students</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Who are they?

<table>
<thead>
<tr>
<th>Winter ORF</th>
<th>20 Benchmark (59%)</th>
<th>10 Strategic</th>
<th>4 Intensive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>8 African American Boys, F/R</td>
<td>1 African American Girl</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 White Boys, F/R</td>
<td>1 White Girl, F/R, Lunch</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 Latino Girl, F/R, ESL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1 White Boy</td>
</tr>
</tbody>
</table>
### Who are they?

17 (21%) Students at **Intensive** in Fall

<table>
<thead>
<tr>
<th>Winter ORF</th>
<th>0 Benchmark</th>
<th>4 Strategic</th>
<th>13 Intensive (76%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>6 Latino Girls, F/R, ESL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2 White Girls, F/R</td>
<td>4 Latino Boys, F/R, ESL</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 White Girl</td>
<td>2 African American Boys, F/R</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1 White Boy</td>
<td>1 White Boy, F/R</td>
</tr>
</tbody>
</table>

- 2 White Girls, F/R
- 1 White Girl
- 1 White Boy
- 6 Latino Girls, F/R, ESL
- 4 Latino Boys, F/R, ESL
- 2 African American Boys, F/R
- 1 White Boy, F/R
Looking at More Data..............

Using the data on your tables, discuss

- What “successes” you see in the data,
- What concerning patterns you see in the data,
- What additional information/data you might like to have,
Movement within tiers due to rate of improvement (ROI)

- Key questions: *What is the rate of improvement over time? How does it compare to the expected ROI for the same group?*
- Expected ROI between Fall and Winter for 2nd graders on ORF is 1.33 WRC per week.
- Informs decision making about groups of students (e.g., Tier 2, specific intervention programs, subgroups) and individual students.
Are students on average meeting, exceeding, or falling below their targeted ROI for progress monitoring?
Are **individual** students meeting, exceeding, or falling below their targeted ROI for progress monitoring?

![Graph showing individual student progress with ROI values]

ROI = 1.73

ROI = 0.67
What does Tim's progress data suggest?

ROI = 1.33
ROI = 3.60
ROI = 1.40
ROI = 0.67
Culturally Responsive Assessment
What do we know about Assessment Practices?

• Barrera Metz (1988), Harry et al., (2002), Maldonado-Colon (1986), and Ochoa et al., (1997) found that language and culture were not adequately considered during the assessment process.

• English language tests were often used exclusively even when the student’s background warranted bilingual testing.

• Whether or not the unexpected underachievement of ELLs could be explained by their limited English proficiency was not given adequate consideration.

• We still need to develop better ways of assessing students’ language and literacy skills.

Klingner, Artiles, & Barletta
Stage Model for Nondiscriminatory Assessment (Rhodes, Ochoa, & Ortiz, 2005)

Assess for purpose of intervention
Assess initially with authentic and alternative procedures
Assess and evaluate the learning ecology
Assess and evaluate language proficiency
Assess and evaluate opportunity for learning
Assess and evaluate relevant cultural and linguistic factors
Evaluate, revise, and re-test hypotheses
Determine the need for and language(s) of formal assessment
Reduce bias in traditional assessment practices
Support conclusions with multiple indicators and converging data
“Unless measurement methods used in RTI, whether CBM or otherwise, account for the differential rates of development that are occurring in the processes related to native language acquisition, English acquisition, and acculturation to the mainstream, there is no guarantee that results will be any more “fair”.

Ortiz, 2009
Recommended Assessment Procedures

- Portfolios and authentic assessment procedures with rubric evaluation,
- Curriculum based measurement,
- Dynamic assessment,
- Language reduced assessments (e.g., nonverbal),
- Native language assessments, and
- Modified and adapted assessment procedures.
Additional Suggestions

- Cultural and linguistic differences should be taken into consideration when tools selected, developed, and interpreted,
- Students should be active participants in the progress monitoring process,
- Comparisons should be made between “true peer” groups (e.g., other ELL students),
- Use both quantitative and qualitative information,
- Consider students' accents and pronunciations when scoring; do not penalize students for dialect features,
- Raters should be similar to students in terms of cultural and linguistic background.

Ortiz; Brown & Doolittle; Vaughn & Ortiz
Reducing Bias in Traditional Practices

- Evaluate the impact of cultural and linguistic demands on assessment data.
- One approach, by Flanagan, Ortiz, & Alfonso (2007), is the Culture-Language Test Classifications (C-LTC) and Culture-Language Interpretive Matrix (C-LIM)
- The C-LIM was designed to answer the question: “Is the measured performance a reflection primarily of actual ability or simply one of cultural or linguistic difference?”
- The results of an individual’s C-LIM can be interpreted within the following expected pattern of performance for diverse individuals.
Cultural and Linguistic Classification of Tests Addressing Bias in Test Validity and Interpretation

Pattern of Expected Performance of Culturally and Linguistically Diverse Children

<table>
<thead>
<tr>
<th>DEGREE OF LINGUISTIC DEMAND</th>
<th>LOW</th>
<th>MODERATE</th>
<th>HIGH</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEGREE OF CULTURAL LOADING</td>
<td>LOW</td>
<td>MODERATE</td>
<td>HIGH</td>
</tr>
<tr>
<td>LOW</td>
<td>PERFORMANCE LEAST AFFECTED</td>
<td>INCREASING EFFECT OF LANGUAGE DIFFERENCE</td>
<td></td>
</tr>
<tr>
<td>MODERATE</td>
<td>INCREASING EFFECT OF CULTURAL DIFFERENCE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HIGH</td>
<td>PERFORMANCE MOST AFFECTED (COMBINED EFFECT OF CULTURAL &amp; LANGUAGE DIFFERENCES)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Suggested Websites and Resources


- Cross Battery Assessment, crossbattery.com.

- Iris Center at Vanderbilt University, iris.peadbody.vanderbilt.edu


Thank You!

Please leave the data packets for the next session.