Reporting With Purpose: Current Approaches to Promoting Test Score Meaning

April L. Zenisky
University of Massachusetts Amherst

NCME 2015, Chicago
The Landscape

- Charged political and social climate for us
  - “Psychometric-industrial complex”
  - “Test publisher cartels”

- Spotlight on tests, test use, and test data is very different than in past
Democrats and Republicans Agree: It's Time To Rewrite No Child Left Behind

States Seek Guidance in Face of 'Opt-Out' Push
A flurry of parents seeking to opt their children out of new assessments has some states seeking guidance from federal officials on test-participation mandates.

Students Around The Country Are Opting Out Of Tests, Even If State Law Doesn't Allow It
Libertyville student opts out of PARCC, given 'sit and stare' time
Revolt against high-stakes standardized testing growing — and so does its impact

New state standardized tests begin after rocky trial run
States Prepare Public for Common-Core Test Results
Scores from tests aligned with the new standards are widely expected to be lower than the results of previous assessments, and states want to head off a backlash.

ATLANTA — Eleven former Atlanta Public Schools educators accused Wednesday of conspiring to cheat on standardized tests have been convicted, a federal jury ruled.

Politicians Introduce Bills That Would Work To Reduce The Role Of School Standardized Testing
“Promoting Test Score Meaning”

• Going to spend my allotted time today unpacking that phrase

• How can we do this?

• Three main threads through the talk
  – Needs
  – Content
  – Format
One key perspective on promoting meaning in test scores

- Evolution in emphasis on reporting responsibilities from 1966 to 1974 to 1985 to 1999 to 2014
- Primacy and forcefulness characterize the 2014 language
- Substantive changes in reporting expectations as well
My Favorite Standard

- 1985 15.10: Those responsible for testing programs should provide appropriate interpretations when test score information is released to students, parents, legal representatives, teachers, or the media. The interpretation should describe in simple language what the test covers, what scores mean, common misinterpretations of test scores, and how the test will be used.

- 2014 6.10: When test score information is released, those responsible for testing programs should provide interpretations appropriate to the audience. The interpretations should describe in simple language what the test covers, what scores represent, the precision/reliability of the scores, and how scores are intended to be used.
What did that standard remind us about promoting meaning?

1. Audience Matters
2. Define the Domain
3. Context, Context, Context
4. Results Beyond Numbers (*there’s a range!*)
5. It’s All About Intent (Validity!)
What did that standard remind us about promoting meaning?

In a word: Context.
Seems obvious, but...

- Think for a moment about how you understand data and results
  - How do data stories unfold?

- Enhancing context helps to clarify meaning
- Better understanding of meaning improves relevance
- Greater perceived relevance = greater likelihood of action, use, and/or value
Organization for remainder

• Reporting Strategies and Sources of Context
  – Reporting Frames and Themes
  – Locus of Control
  – Systems for Support
I. Reporting Frames and Themes

- Hattie (2009): *Visibly Learning from Reports*

- Principle 8: Each Report needs to have a major theme.

- Principle 11: A Report should be designed to address specific questions.
<table>
<thead>
<tr>
<th><strong>Console Report</strong></th>
<th>How am I going relative to others?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Individual or Group Learning Pathway</strong></td>
<td>How am I going relative to my strengths and weaknesses?</td>
</tr>
<tr>
<td><strong>Curriculum Skyline</strong></td>
<td>How is this group distributed in their achievement?</td>
</tr>
<tr>
<td><strong>Progress</strong></td>
<td>How am I going over time?</td>
</tr>
<tr>
<td><strong>Expectation</strong></td>
<td>What are the student/teacher targets over time?</td>
</tr>
<tr>
<td><strong>Tabular Report</strong></td>
<td>Downloadables for further analyses and storage</td>
</tr>
<tr>
<td><strong>What Next</strong></td>
<td>What do I do next?</td>
</tr>
</tbody>
</table>

(Adapted from Hattie, 2009)
| Group Students | What are areas of weakness?  
How difficult were troublesome items? |
| Help individuals | Who just barely made it into a performance level?  
What does X need help in within [area]? |
| Communicate with parents | How do schools in town compare?  
How has X’s performance changed? |
| Understand the standards | What standards come under [area]?  
What standard(s) is item linked to? |
| Planning and Validating | What standards in [area] did my students have trouble with?  
How did my limited English students do? |

(Underwood, Zapata, and Hester, 2006)
• Policymakers - adopt a question-based approach (Van Winkle, Vezzu, & Zapata-Rivera, 2011)

• Build interactive report around specific question
  – Graphical representation of answer
  – Results description
  – Highlights of results
  – Limitations
  – Statistical caveats
Reports for Administrators

What are the results for: my district for the 8th grade tests in all subjects?

My District 8th Grade Test Results*

<table>
<thead>
<tr>
<th>District</th>
<th>Total % Below</th>
<th>What's that?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Writing</td>
<td>56</td>
<td>30</td>
</tr>
<tr>
<td>Math</td>
<td>55</td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School 1</th>
<th>Total % Below</th>
<th>What's that?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>35</td>
<td>10</td>
</tr>
<tr>
<td>Writing</td>
<td>58</td>
<td>35</td>
</tr>
<tr>
<td>Math</td>
<td>41</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School 2</th>
<th>Total % Below</th>
<th>What's that?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>39</td>
<td>10</td>
</tr>
<tr>
<td>Writing</td>
<td>47</td>
<td>17</td>
</tr>
<tr>
<td>Math</td>
<td>54</td>
<td>31</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>School 3</th>
<th>Total % Below</th>
<th>What's that?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reading</td>
<td>36</td>
<td>7</td>
</tr>
<tr>
<td>Writing</td>
<td>60</td>
<td>34</td>
</tr>
<tr>
<td>Math</td>
<td>54</td>
<td>31</td>
</tr>
</tbody>
</table>

* These are cumulative results based on all tests taken during the school year.

Key findings
About half of the students in all district schools performed at the Not yet at Basic and Basic levels on the Writing and Mathematics tests.
• Goodman & Hambleton, 2004:
  - What would you do next? How would you navigate around report ‘screen’? [paper or electronic] What are the potential uses for this report?

• NEGP, 1998:
  - How did my child do? What does this performance reflect? How did my child perform in comparison? What can I do to help my child improve?

• Hattie & Timperly, 2008:
  - Where am I going? How am I going? Where to next?
Reporting Frames and Themes

- Reflect on report purpose and specific report uses
- Draw on your stakeholders - research!
  - Learn about their needs
  - Then, see what users think!
II. Locus of Control

• What’s a report?
  – Document generated by agency charged with reporting and disseminated to users intact
  – ‘Document’ generated by user from agency database and ...

• Two very different propositions

• From “Here’s the Story” to “You’re in Charge”
  – Interesting twist to report development processes and use cases
“It’s interpret-your-own-test-results day today.”
Why is this important?

- Reflective of evolution in how people interact with information today
  - Explanatory and Instructive
  - Exploratory and Constructive

- As the agents of reporting, these orientations inform the choices we make in developing and evaluating reporting systems
Paradigm shift for results reporting

• Allow users to customize which results to display based on interests / needs
  – Some blurring of line between reporting and secondary analysis

• Caveat
  – Generally, practices for individual reporting vs. group-level reporting remain quite different
    • Function of current tests, intended users, information needs, and data

• Emerging approaches from formative assessment, gamification, e-learning, dashboards...
**STEP 4:** View each report table by selecting the report name from the drop-down menu. Create report types to edit and preview, each tab created represents one report type to export.

**Subject, Grade:** Reading, Grade 8  
**Jurisdictions:** National, National public, Massachusetts  
**Measure:** Composite scale  
**Variable:** All students  
**Year:** 2013

**Select Report:** Report 1

---

**Chart Options**

Choose an available chart type based on selected data. Choose data dimensions from the drop-down menus and name the chart, then preview the chart. To select different data return to Data Options.

- **Bar Chart**
- **Column Chart**
- **Line Chart**

**Bar Values:** All students  
**Values Grouped by:** None  
**Chart Name:** Chart 1

---

[Export Reports]
III. Systems for Support

• Focus of many conversations about score reporting is how to develop a report
  – What do we put on there?
  – What color for the bars?

• Great, but...

• If goal really is promoting meaning:
  – What else can agencies do to support interpretation and action?
Upping the ante

• Research has identified various specific challenges and knowledge disconnects between and within various user groups, such as:
  – Technical language, use of statistics, jargon
  – Score scales
  – Measurement error
  – Primary language and language of report
  – Unfamiliarity with test or test purpose
  – Time
  – Lack of resources
  – Lack of access to resources
A Modest Proposal

• (Further) reframe report development
  – Reporting responsibility is more than just the document with performance data

• In reporting systems:
  – Allocate resources to develop interpretive resources
  – Be creative about formats and access
A Few Promising Practices

- MCAS: Family guides in English and 10 other languages
- SAT Skills Insight
- Growth Model Analogies & Videos
  - Georgia, Virginia, Hawaii
  - [https://vimeo.com/47277587](https://vimeo.com/47277587)
- Sunday afternoon at NCME
- What about non-online resources?
  - Outreach at school level
Concluding Thoughts

- Promoting test score meaning
- How do we do this?
- Multifaceted construct of context
  - Themes and Frames
  - Locus of Control
  - Support for Users
<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advanced</strong></td>
<td>Students go beyond understanding what they read to explain what they like or do not like about a story.</td>
</tr>
<tr>
<td><strong>Accelerated</strong></td>
<td>Students understand what they read (e.g., fairy tales, folk tales, poetry) and compare stories with other stories.</td>
</tr>
<tr>
<td><strong>Proficient</strong></td>
<td>Students understand what they read. They try to list important ideas in the texts they read.</td>
</tr>
<tr>
<td><strong>Basic</strong></td>
<td>Students understand some of what they read. They use strategies (e.g., clues in sentences and paragraphs, knowledge of word parts) to learn new words.</td>
</tr>
<tr>
<td><strong>Limited</strong></td>
<td>Students may struggle with simple reading tasks (e.g., fairy tales, folk tales, poetry).</td>
</tr>
</tbody>
</table>

Kyree’s spring score is 395.
She has performed at the basic level and does not meet standards for Grade 3 reading.

Kyree’s Fall Score is 390
The 3 R’s

- Reporting systems that promote meaning are not built as one-and-done propositions
  - Hambleton & Zenisky, 2012

- Return
- Re-evaluate
- Revise
Prioritize intended users

- Identify main use cases
- Develop reports to accomplish known tasks
- Where feasible, develop systems to aid users with task-specific data analysis needs
- Provide supports to address known issues for various user groups
Thanks!

questions, comments, or a copy of the slides:

azenisky@educ.umass.edu