TIME TO CHANGE THINGS UP: EVOLVING THE CONVERSATION ABOUT RESULTS REPORTING

April L. Zenisky, Ed.D.
College of Education / Center for Educational Assessment
University of Massachusetts Amherst

February 1, 2019: UMN Educational Measurement Speaker Series
Reporting has been an interest of mine for about 20 years now, dating back to my first year in grad school.

Salient because of a conversation at my first NCME/AERA in Montreal, April 1999.

It was suggested to me that reporting was a “dead-end topic” and I might look elsewhere to contribute to our field.
Clearly, I disagreed.

- Reporting *does* matter.
- Communicating test results well is not easy.
- Progress is being made, but we haven’t solved everything.
PATHWAY FOR TODAY

• Briefly trace ways in which research and practice of score reporting has progressed in recent years (and in some ways, not)

• Reflect on challenges and opportunities for reporting in 2019 and beyond

• Talk about ideas I’m exploring in my work
ART
AND
SCIENCE

Science: Psychometrics and statistics

- What data should be communicated?
  - Scores, percentiles, grade equivalents, errors, written diagnostic information, skill profiles, etc.

Art: Public relations, graphic design

- What is the most effective way to communicate that data?
  - Graphs, tables, color, layout and format, etc.
• Wainer, Hambleton, & Meara (1999) Study

- Worked with 1994 NAEP Reading data, and carried out experimental study

- Identified five problematic displays and redesigned them

- Engaged 100 educators and policy makers
Figure 3. Average Reading Proficiency by Grade and by Region — NAEP 1992 and 1994

*Significant decrease between 1992 and 1994
Average Reading Proficiency by Grade and Region
NAEP 1992 and 1994

NAEP Proficiency Scale

Year
PERSPECTIVES ON REPORTING AS A NON-“DEAD END” TOPIC

• *External*: Role of tests and testing has evolved over the years, particularly in education context
  - Public perceptions, increasingly polarizing language

• *Internal*: the actions of reporting were historically viewed an add-on to the test development process
  - Often it’s up to test developers, but we’re perhaps not best equipped for that
BACKGROUND - INDIVIDUAL REPORTS

• Purpose: To provide information about individual performance

• Relevant stakeholders:
  - Examinees themselves
  - Families
  - Educators (teachers, profs, administrators, etc.)
  - Decisionmakers (employers, admissions, etc.)

• Relevance of granular data decreases with distance
BACKGROUND - GROUP-LEVEL REPORTS

• Purpose: To provide information about performance of groups of test-takers

• Often, groups are defined as:
  - Proximity/Geography: Class, school, district, state, etc.
  - Demographics: Race/ethnicity, sex, special needs status

• Stakeholders
  - More proximal (educators, admins., institutions)
  - Less proximal (general public, policymakers)

• Action versus information orientation shifts with distance
BACKGROUND - CONTENT ELEMENTS

- REPORT PURPOSE AND/OR TEST PURPOSE
- OVERALL SCORE
- PERFORMANCE LEVEL (BINARY OR >2 CATEGORIES)
- RELEVANT REFERENCE GROUP COMPARISON DATA
- INDIVIDUAL TEST QUESTION PERFORMANCE RESULTS
- SUBCONTENT AREA PERFORMANCE DETAILS
- GROWTH PERCENTILES
BACKGROUND - DESIGN ELEMENTS

• Personalization
• Highlight main/key results
• Clear sectioning of report elements: Logical flow
• Mix of text, tables, and graphics
• Compartmentalized layout / Everything has a place
• Balance data/information with white space
• Minimal jargon, long technical explanations
• Color (judiciously used)
• Length: 1 vs. 2 (or more!!) pages
BACKGROUND - AUDIENCES

- Educators
- Educational Administrators
- Parents
- Students
- Politicians/Policymakers
- Media
- Business Community
- Researchers
- General Public

All of these groups have different interests and needs regarding test data and results.
Likewise, they may vary in assessment needs and interests.
FROM WHENCE WE CAME...
HISTORICAL CHALLENGES IN REPORTING
(Or, ”The World According to Ron Hambleton”,
when I got into this stuff)

• Scales: Reporting scales
  • Raw scores, scale scores; Percents & percentiles; IQ scores...

• Disjointed or meaningless information on some reports

• Lack of consideration of users

• Lack of error bands (and good explanations of what that means for interpretation)

• Lack of actionable diagnostic information
• Happily, things have changed greatly since then

• Reporting is indeed part of the TD conversation
  - Design processes: work by Hambleton & Zenisky; Zapata-Rivera
  - Validity: Thinking (and talking!) about actual reporting activities not at end, but beginning
  - Ongoing, multidisciplinary work on contents and formats
  - Engaging stakeholders in research and practice
Jane’s ELA/Literacy Score: 2680±10

Level 4: The student has exceeded the achievement standard and demonstrates advanced progress toward mastery of the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.

Level 3: The student has met the achievement standard and demonstrates progress toward mastery of the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.

Level 2: The student has nearly met the achievement standard and may require further development to demonstrate the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.

Level 1: The student has not met the achievement standard and needs substantial improvement to demonstrate the knowledge and skills in English language arts/literacy needed for likely success in entry-level credit-bearing college coursework after high school.

Jane’s ELA/Literacy scale score is 2680. This score is higher than the average score of eighth graders in her school, higher than that of eighth graders in her district, and higher than that of eighth graders statewide.

A student’s test score can vary if the test is taken several times. If your child was tested again, it is likely that Jane would receive a score between 2670 and 2690.
United States Medical Licensing Examination®

Step 3 Score Report

FOR EXAMINEE USE ONLY. THIRD-PARTY USERS OF USMLE SCORES SHOULD RELY SOLELY ON OFFICIAL TRANSCRIPTS RECEIVED DIRECTLY FROM THE EXAMINEE'S USMLE REGISTRATION ENTITY.

NAME: Last, First
USMLE ID: 0-000-000-0
TEST DATE: Month 1, 2018

Your Performance

Test Result
PASS

Test Score
205

Your Performance Compared to Other Examinees

The chart below represents the distribution of scores for recent examinees from US and Canadian medical schools taking Step 3 for the first time. Reported scores range from 1-300 with a mean of 226 and a standard deviation of 15.

If you tested repeatedly under the same conditions on a different set of items covering the same content, without learning or forgetting, your score would fall within one standard error of the estimate (SEE) of your current score two-thirds of the time. The SEE on this exam is 8 points.

Your score +/- SEE: 197 – 213
Your scores indicate you are close to being on track for college readiness, but you need to continue to strengthen your skills.

Let's get you back on track, so you won't have to take noncredit courses in college. You have free, personalized recommendations waiting for you on khanacademy.org/sat.
Reading Achievement Results

State Average Score: 420
School Average Score: 410
District Average Score: 404

Advanced - Students go beyond understanding what they read to explain what they like or do not like about a story.

Accelerated - Students understand what they read (e.g., fairy tales, folk tales, poetry) and compare stories with other stories.

Proficient - Students understand what they read. They try to list important ideas in the texts they read.

Basic - 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.

Limited - Students may struggle with simple reading tasks (e.g., fairy tales, folk tales, poetry).

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
Jacqueline scored at Level 3 on the English language arts/Literacy test and scored at Level 2 on the Mathematics test.

<table>
<thead>
<tr>
<th>ELA/Literacy</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
<th>Level 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics</td>
<td></td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**ELA/Literacy Results**

Jacqueline’s Total Scale Score = 2651  
(Score Scale Range 2288-2769)

**Level 3: Meets the Achievement Level**

Jacqueline has met the achievement level for English language arts and literacy expected for this grade. Students performing at this level are demonstrating progress toward mastery of English language arts and literacy knowledge and skills. Students performing at this level are on track for likely success in high school and college coursework or career training.

<table>
<thead>
<tr>
<th>Student’s Score</th>
<th>2651</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Average</td>
<td>2602</td>
</tr>
<tr>
<td>District Average</td>
<td>2571</td>
</tr>
</tbody>
</table>

A student’s test score can vary if the test is taken several times. If your child were tested again, it is likely that Jacqueline would receive a score between 2641 and 2661.
FOR SUMMATIVE ASSESSMENT PURPOSES...

• Report development efforts are coalescing around several key elements
  - Prioritizing status as information goal - snapshot view
  - Being purposeful with space and structure
  - Using layout, font size, color, and element design to communicate key data
  - Communicating error! (Zapata-Rivera, Zwick, & colleagues)

We’re good at status reporting in summative context!
BUT...
SANDS ARE SHIFTING AND AND SEISMIC CHANGES ARE AFOOT

Assessments Themselves
Information Needs and Interests
Technology and Information Graphics
1. NEXT-GEN ASSESSMENTS

• Domains are getting harder to define and/or measure
  - 21st Century Skills; Collaborative Problem Solving, etc.
  - Even traditional subjects are evolving
• Multimodal data streams and natural interfaces
  - Cross-cutting traits that play out in >1 context
• Focus on Process, not just Outcomes
  - Activity patterns within a task
• KSAs to be assessed involve interactions and dependencies
• Lack of stability at grain size of interest, often
2a. TECHNOLOGICAL UNDERPINNINGS

- Increased use of the Web for everything, including reporting
- More interactive/self-directed tools
  - Significant departure for score reporting
  - Public relations aspect of large-scale testing
  - Control of statistical ‘story’ shifts from testing agency to consumer of data
- Allow users to customize which results to display based on interests / needs
  - Blurring of lines - reporting and secondary analysis
2b. DATA VIZ AND INFORMATION GRAPHICS

• We sometimes forget why visualization works!
  - Organize data by indexing it spatially
  - Offloading cognitive processes onto automatic perceptual processes
  - Free up working memory, to help guide audiences to certain patterns / data stories

• Exploration vs. communication purposes
  - Visualizations for different goals / aims
2b. DATA VIZ AND INFORMATION GRAPHICS

• Cognitive science and related research on information processing can be quite instructive for reporting

• Inspiration from outside of testing
  - Originated with Playfair, Tufte, etc., as cited by Wainer
  - Now: apps, activity trackers, consumer product & service companies, media...
ComEd
An Exelon Company

SERVICE FROM 1/11/16 THROUGH 2/11/16 (31 DAYS)
Residential - Single

Customer Name
Service Address
City, State, Zip
000.000.0000

Total Amount Due by 3/4/16 $71.29
Thank you for your payments totaling $77.44.

TOTAL USAGE (kWh)

<table>
<thead>
<tr>
<th>2015</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEB</td>
<td>354</td>
</tr>
<tr>
<td>MAR</td>
<td>365</td>
</tr>
<tr>
<td>APR</td>
<td>374</td>
</tr>
<tr>
<td>MAY</td>
<td>312</td>
</tr>
<tr>
<td>JUN</td>
<td>277</td>
</tr>
<tr>
<td>JUL</td>
<td>1132</td>
</tr>
<tr>
<td>AUG</td>
<td>1290</td>
</tr>
<tr>
<td>SEP</td>
<td>596</td>
</tr>
<tr>
<td>OCT</td>
<td>361</td>
</tr>
<tr>
<td>NOV</td>
<td>484</td>
</tr>
<tr>
<td>DEC</td>
<td>565</td>
</tr>
<tr>
<td>JAN</td>
<td>491</td>
</tr>
</tbody>
</table>

CURRENT CHARGES SUMMARY
See reverse side for details

SUPPLY $31.98
DELIVERY $30.96

ComEd provides your energy.
ComEd.com
1.800.334.7661

TAXES & FEES $8.35

ComEd delivers electricity to your home.

For Electric Supply Choices visit pluginillinois.org
Want to change behavior?
Show impact!
Hegarty example – the width of the cone is the uncertainty at a point in time!!
My daughter's sleeping patterns for the first 4 months of her life. One continuous spiral starting on the inside when she was born, each revolution representing a single day. Midnight at the top (24 hour clock). [OC]
Weighing their words

We compared word counts in the key parties' manifestos to work out what really mattered to them.

**How it works:** Each manifesto is broken into coloured blocks which proportionally represent the percentage of words devoted to key topics and individual policies.

- **Brexit**
- **Democracy and British values**
- **Economy and employment**
- **Education and childcare**
- **Environment**
- **Equality**
- **Healthcare**
- **Housing**
- **International development and foreign aid**
- **Ireland**
- **Local communities**
- **Migration**
- **Military and defence**
- **Miscellaneous**
- **Multiculturalism**
- **Online security**
- **Policing**
- **Scotland**
- **The elderly and disabled**

**Sample policies**

Conservatives
- Propose for adult social care, later dubbed the 'camilla tax': 981 words
- Reduce not migration to tens of thousands: 345 words
- Cap on energy prices, 256 words

Labour
- Nationalise the railways: 377 words
- £30bn more for NHS: 247 words
- Scrap tuition fees: 132 words
- Increase income tax for highest earners: 130 words

Lib Dems
- Referendum on final Brexit terms: 459 words
- Legalise cannabis and end prison terms for drug possession: 187 words
- Ban the niqab/burqa: 189 words
- Not pay any Brexit divorce settlement to EU: 105 words

UKIP
- Close the Department for International Development: 925 words
- Scrap Trident: 283 words
- Restore devolution and avoid direct rule: 1,014 words
- 50,000 new Northern Ireland jobs: 247 words

SNP
- Keep Scotland in the EU single market: 933 words
- Second Scottish independence referendum: 254 words
- Scrap Trident: 283 words
- Reduce hospital waiting lists: 194 words

DUP
- 7,667 words: "Standing strong for Northern Ireland"
TURNING BACK TO TESTING...
3. INFORMATION INTERESTS AND NEEDS

Primary Needs
Formal test purposes and validated test uses

Secondary Wants
What users hope to glean from results

These are determined \textit{a priori}, typically by the client, NOT necessarily many (most?) users.

Status is good. But action is where it’s at (for many stakeholders).
HEIGHTENED EXPECTATIONS

• Quite simply, many users expect more.

• More:
  - Detail
  - Guidance
  - Actions
  - Help
  - Precision

More than what our current tests, Standards and models typically permit, at present
TIME MANAGEMENT BY SECTION

INTEGRATED REASONING

VERBAL

QUANTITATIVE

YOUR MEAN RESPONSE TIME PER QUESTION

MEAN TIME FOR ALL TEST TAKERS:

1.87 minutes

Definition

For all GMAT Exam scores recorded in the past three years, the mean time to complete a question in the Integrated Reasoning, Verbal, and Quantitative sections is 1.87 minutes. Use this number to evaluate how well you managed your time.

SUMMARY

- Your GMAT Total score of 640 is higher than 73% of GMAT Exam scores recorded in the past three years.
  - Your Integrated Reasoning score of 3 is higher than 75% of GMAT Exam scores recorded in the past three years.
  - Your Verbal score of 34 is higher than 71% of GMAT Exam scores recorded in the past three years.
  - Your Quantitative score of 64 is higher than 50% of GMAT Exam scores recorded in the past three years.
- Your Quantitative score of 44 and your Verbal score of 34 were used to calculate your GMAT Total score. Especially strong or weak scores on either section significantly impacts the GMAT Total score.
- Time management is an important factor in the GMAT Exam scoring. Failing to complete all questions can impact your final score.
THE WICKEDEST REPORTING CHALLENGE

• The difficult, murky, thorny, troublesome, gray area of subscores, subdomains, and next steps
  - Great research by Sinharay, Haberman, et al.

• Our Unreliability Problem
  - Typically, relatively few items from each KSA
  - How reliable are granular results?
  - How useful are these results?

• Tension of “right proper” psychometrics and user perceptions of actionable results at useful grain size
Your child’s performance will also be broken down into subcategories within each subject. In each subcategory, your child will be provided a rating of one to three stars, with a description of your child’s performance in this specific area and the support needed moving forward.

### Reading Performance

<table>
<thead>
<tr>
<th>★★★</th>
<th>Moderate Performance</th>
<th>★★★</th>
<th>Strong Performance</th>
<th>★★★</th>
<th>Weak Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>30%</td>
<td></td>
<td>50%</td>
<td></td>
<td>20%</td>
<td></td>
</tr>
</tbody>
</table>

#### Literary Text

- ★★★ Strong Performance
  
  In this area, your student is able to read and analyze grade-appropriate fiction, drama, and poetry very well and is prepared for further studies.

#### Informational Text

- ★★★ Moderate Performance
  
  Your student can read and analyze grade-appropriate non-fiction, including texts about history, science, art, and music. Your student may need additional support to be fully prepared for further studies.

#### Vocabulary

- ★★★ Weak Performance
  
  Your student will need significant support in using context to determine what words and phrases mean in grade-appropriate texts.

### Writing Performance

<table>
<thead>
<tr>
<th>★★★</th>
<th>Weak Performance</th>
<th>★★★</th>
<th>Strong Performance</th>
<th>★★★</th>
<th>Moderate Performance</th>
<th>★★★</th>
<th>Weak Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td></td>
<td>45%</td>
<td></td>
<td>50%</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Written Expression

- ★★★ Moderate Performance
  
  Your student can compose well-developed, organized, and clear writing, using details from what he/she read, but may need additional support to be fully prepared for further studies.

#### Knowledge & Use of Language Conventions

- ★★★ Strong Performance
  
  Your student can compose writing using the rules of standard English (including those for grammar, spelling, and usage) and is prepared for further studies.

### Legend

- ★★★ Strong Performance
  - Prepared for further studies
- ★★★ Moderate Performance
  - May need additional support to be fully prepared for further studies
- ★★★ Weak Performance
  - Will need significant support for further studies
### READING

<table>
<thead>
<tr>
<th>Reading score range: 10 to 90</th>
<th>Average of students just meeting expectations</th>
<th>School average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your child's score: 44</td>
<td>50</td>
<td>52</td>
</tr>
</tbody>
</table>

#### LITERARY TEXT

In this area, your child did not do as well as students who met the expectations.

Students meet expectations by showing they can read and analyze grade-appropriate fiction, drama, and poetry.

#### INFORMATIONAL TEXT

In this area, your child did almost as well as students who met the expectations.

Students meet expectations by showing they can read and analyze grade-appropriate non-fiction, including texts about history, science, art, and music.

#### VOCABULARY

In this area, your child did as well as or better than students who met the expectations.

Students meet expectations by showing they can use context to determine what words and phrases mean in grade-appropriate texts.

### WRITING

<table>
<thead>
<tr>
<th>Writing score range: 10 to 60</th>
<th>Average of students just meeting expectations</th>
<th>School average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Your child's score: 30</td>
<td>35</td>
<td>39</td>
</tr>
</tbody>
</table>

#### WRITING EXPRESSION

In this area, your child did not do as well as students who met the expectations.

Students meet expectations by showing they can compose well-developed, organized, and clear writing, using details from what they have read.

#### KNOWLEDGE AND USE OF LANGUAGE CONVENTIONS

In this area, your child did as well as or better than students who met the expectations.

Students meet expectations by showing they can compose writing using the rules of standard English, including those for grammar, spelling, and usage.

### LEGEND

- **Below Expectations**
- **Nearly Meets Expectations**
- **Meets or Exceeds Expectations**

To see selected questions from the test, visit [understandthescore.org](http://understandthescore.org).
The New York State Learning Standards for English Language Arts & Literacy describe what Jane should know and be able to do at each grade level. This section shows the number of points earned on the Reading and Writing from Sources test sections.

**Reading**
In this section, students were asked to:
- comprehend the explicit and implicit ideas present in grade-level text and recognize how they develop over the course of the text;
- describe characters in depth, and connect ideas and events;
- understand the meaning and effects of words and figurative and connotative phrases;
- analyze the text's structure and organization, and identify how point of view is developed;
- integrate the knowledge and ideas conveyed in text using illustrations and logical connections between parts of the text.

**Writing from Sources**
In this section, students were asked to:
- make claims that show valid reasoning and sufficient textual evidence;
- write about characters and describe how they relate to each other in the text;
- compare and contrast points of view and themes in texts;
- determine the effect of language on aspects of a text and describe the importance of story and structural elements of the text;
- communicate understanding by producing coherent writing that demonstrates grade-level English grammar and use.

---

**Helpful next steps**

**Supporting Jane's Success**

- Contact Jane's teacher and ask which skills are the most challenging for Jane.
- Ask your school how you will receive regular reports on Jane's progress.
- If appropriate, seek student support services such as counseling or school guidance to help Jane be most successful.
- Review classwork and homework with Jane to see how Jane's skills are progressing.

For more information about this test, the New York State standards, and how you can help Jane, go to: www.engageny.org/parent-and-family-resources
Jane’s score is 706. She has performed at the proficient level and meets standards for Mathematics.

School Average Score: 725
District Average Score: 721
State Average Score: 717
Your Evidence-Based Reading and Writing Score

380 | 160 to 760

Your Nationally Representative 18th Sample Percentile

Your Total Score

860 | 320 to 1520

Your Nationally Representative 28th Sample Percentile

Your Math Score

480 | 160 to 760

Your Scores: Next Steps

Reading Test

18
8 to 38

Improve your skills by focusing on:

- Maintain or add information or ideas to a piece of text to support claims or points in challenging text
- Use a variety of sentence structures to accomplish a rhetorical purpose such as persuading an audience
- Revise text as needed to improve the exactness or content appropriateness of word choices within challenging texts

The College and Career Readiness Benchmark

Reaching your grade-level Benchmark means that you are likely on track to be ready to succeed in select first-year, credit-bearing college courses.

- Need to strengthen skills
- Approaching Benchmark
- Meets or exceeds Benchmark

*The red, yellow, and green ranges in the test scores and subscores reflect your areas of strengths and weaknesses compared to the typical performance of students in your grade.*
**FAQs**

**What is the Smarter Balanced ELA/Literacy Assessment?**
This assessment is aligned to new academic standards for ELA/literacy. These standards identify what students should know and be able to do to graduate high school ready for college and the workplace. They challenge students to develop a deeper understanding of subject matter, learn how to think critically, and apply what they are learning to the real world.

**What is a claim?**
ELA claims are broad statements that identify the knowledge and skills students should know and be able to do as they progress toward college and career readiness. How your child performed on each claim is found in the columns to the right.

### How did Jane perform on the different claims of the test?

<table>
<thead>
<tr>
<th>Claim 1: Reading</th>
<th>Claim 2: Listening and Speaking</th>
<th>Claim 3: Writing</th>
<th>Claim 4: Research/Inquiry</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="%E2%9A%A0%EF%B8%8F" alt="Below Standard" /></td>
<td><img src="%E2%9C%94%EF%B8%8F" alt="Above Standard" /></td>
<td><img src="%E2%9C%94%EF%B8%8F" alt="Above Standard" /></td>
<td><img src="%E2%9E%96" alt="At/Near Standard" /></td>
</tr>
</tbody>
</table>

- **Claim 1: Reading**
  - Student has difficulty reading closely and analytically to comprehend a range of increasingly complex literary and informational texts.

- **Claim 2: Listening and Speaking**
  - Student can employ effective speaking and listening skills for a range of purposes and audiences.

- **Claim 3: Writing**
  - Student can produce effective and well-grounded writing for a range of purposes and audiences.

- **Claim 4: Research/Inquiry**
  - Student may be able to engage in research and inquiry to investigate topics, and to analyze, integrate, and present information.
TO RECAP THE BIG CHANGES:

Tests are changing

Technology and data visualizations are evolving

Data ‘wants’ are growing

... so how do we move forward?
CHANGING THINGS UP, #1

- Let’s think about how we think about data and reports
  - What we’re good at is only part of what some stakeholders want
  - Pursuit of psychometric perfection is paralyzing progress

- What balance can we strike?
- For lower-stakes uses, can we think about being more forthcoming about data, acknowledging limitations?
CHANGING THINGS UP, #2

• New work by Roduta Roberts, Gotch, & Lester (2018)
  - Focus on Language of reporting
  - Discourse analysis of SR language, to ID patterns

• Idea of activity theory
  - “analysis using this framework provides a window into how different individuals (i.e., subjects) who are all working toward the same overall goal within a system may have different priorities and motives, perhaps acting on different objects within the system.”
CHANGING THINGS UP, #2

- Displays of Information, Knowledge Claims, and Doing Accountability
CHANGING THINGS UP, #2

• Maybe activity theory is ‘right’, maybe it’s not

• But it’s a different approach to developing and evaluating reports

• New ideas, different perspectives on how we talk about test results might be helpful to move us past what we’re good at to what we need to work on as a field
  - especially with new & emerging tests, contexts, constructs
Evolution in how people interact with information
- Explanatory and Instructive
- Exploratory and Constructive

As the agents of reporting, these orientations inform the choices we make in developing and evaluating reporting systems

Not just a job for psychometricians

Please: Assemble a team (Slater, Livingston, Lester, 2019)
CHANGING THINGS UP, #4

• Figure out our angles for different stakeholder groups
  - Adopt a question-based approach (Van Winkle, Vezzu, & Zapata-Rivera, 2011)

• Build interactive report around specific question(s)
  - Graphical representation of answer
  - Results description
  - Highlights of results
  - Limitations
  - Statistical caveats
SO, CONCLUDING QUESTIONS...
• 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:
  - Let’s take a look, and have some hard conversations about what we are reporting, to whom?
QUESTIONING OUR STATUS QUO

• Will our current reporting conventions and approaches work with next-gen assessments? How well?

• Can we begin to harness how people consume and use information now and in the coming years?

• How do we reconcile have-to’s and want-to’s to make information available and useful at different levels of stakes?
Conceptually return to roots of report development

- Reporting responsibility is more than just the document with performance data

In reporting systems:

- Acknowledge range of uses, develop flexibilities in our approaches
- Allocate real resources to develop interpretive resources
- Be creative about formats and access
Thanks!

azenisky@educ.umass.edu