“Innovative Item Types: Why Educators Care About the Future of Technology Enhanced Assessments”
Kathleen Scalise, CERA Nov. 2012
Why not same items?

- Some strengths for formats:
  New media, real time, feedback, expectations

- Some challenges:
  Scoring, measurement models, psychometrics, accessibility and more

Poll Everywhere: Text 575507 and your message to 37607; OR login PollEv.com/kscalise
Definition of Assessment

- “Collecting Evidence Designed To Make An Inference” (Scalise & Gifford, 2008)

- Along with growing possibility of TEIs, there are growing expectations for greatly expanded use in assessment.

- Are some “rules” of assessment “breaking” because of these expectations? Is this good, bad, indifferent...unknown? And why should educators care?

Poll Everywhere: Text 575507 and your message to 37607; OR login PollEv.com/kscalise


Tuesday, December 4, 2012
• Truth Statements: Some believe there is a major trade-off between capturing the rich characteristics of activities and tasks that technology now affords, and obtaining high quality measurement evidence.

• Is this necessarily the case?
Jared is testing how much weight a bag can hold. He plans to put juice bottles into three bags. He wants each bag to have a total weight within the given range.

- Drag juice bottles into each bag so that the weight is within the given range.
- Leave the bag empty if the given range is not possible using juice bottles.

Poll Everywhere: Text 575507 and your message to 37607; OR login PollEv.com/kscalise
J. Bransford: 2nd Life Scenarios That Adapt

Poll Everywhere: Text 575507 and your message to 37607; OR login PollEv.com/kscalise
Cisco “Aspire: IM Adaptive Serious Game

Read research on Aspire at https://research.netacad.net/mod/data/view.php?id=1&rid=30

Tuesday, December 4, 2012
Personalized Assessments: SAS Award Winning Free Writing Reviser Tool

The Award Winning Writing Reviser

“My students love the personalized feedback. They often list the Writing Reviser as one of the tools they learned the most from in the course!”

– Kristin Kipp, SREB/INACOL National Online Teacher of the Year

( Teachers can email me if they want access, UO hosting a community for teacher access, kscalise@uoregon.edu)
My Poem Graphic Organizer

Can you think of some ideas about this poem’s Mood and Meaning?
Type into BLUE BUBBLES, and connect with the PENCIL TOOL.

Your Pasted Poem:

Paste Poem Text Here.
by Zawgee

Bobbing on the breeze blown waves
Bowling to the tide
Hyacinth rises and falls

Falling but not felled
By flotsam, twigs, leaves
She ducks, bobs and weaves.

Ducks, ducks by the score
Jolting, quacking and more
She spins through?

Spinning, swamped, slimed, sunk
She rises, resolute
Still crowned by petals.

Poll Everywhere: Text 575507 and your message to 37607; OR login PollEv.com/kscalise
Are you ready to solve what caused the six legged frog?

**Be sure, because you can't go back!**

Yes, continue  No, Go Back
# Harvard Virtual Performance Assessments

Aligned to Next Gen and College Board Standards for College Success

Table 1. *Items and Dimensions for Exploration 1 A2 Data Set*

<table>
<thead>
<tr>
<th>Items or Observables</th>
<th>Dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control 1-3 (3 dichotomous)</td>
<td>1</td>
</tr>
<tr>
<td>Research 1-6 (6 dichotomous)</td>
<td>1</td>
</tr>
<tr>
<td>Experimentation 1-3 (3 dichotomous)</td>
<td>1</td>
</tr>
<tr>
<td>Sample 1-3 (3 dichotomous)</td>
<td>1</td>
</tr>
<tr>
<td>Claim (1 polytomous, rescored to dichotomous for the model runs due to similar p-values across the lower three levels)</td>
<td>1 &amp; 2</td>
</tr>
<tr>
<td>Farm (1 polytomous)</td>
<td>2</td>
</tr>
<tr>
<td>Backpack Contents (1 polytomous)</td>
<td>2</td>
</tr>
<tr>
<td>Q6-10 (Evidence use in explanation given student is provided with the inquiry data, 5 polytomous items, in order from Q6-10 as Evidence Tadpole 6, Frog 7, Water 8, DNA 9, Blood Test 10)</td>
<td>2</td>
</tr>
</tbody>
</table>

Poll Everywhere: Text 575507 and your message to 37607; OR login PollEv.com/kscalise
Harvard Virtual Performance Assessments

Research Questions

- Can mIRT models be successfully used to calibrate and model VPA serious gamification data aligned to standards-based constructs? If so, what results are obtained for New Frog middle school students, for the sample data sets? How can the information be used?
- Is the calibration above robust and stable enough for use on subsequent New Frog VPA data sets, within reasonable standards of tolerance for model fit, reliability and other characteristics of the instruments?
- Can mIRT models be extended to better handle the VPA data by inserting or embedding Bayes net analysis into the mIRT data aggregation, via a two-stage process? If so, can this amplify the information available, better align instruction with assessment, improve the ability to reliably report at more multiple grain sizes, reduce time required for assessments, and still sustain the quality of the measures?

Poll Everywhere: Text 575507 and your message to 37607; OR login PollEv.com/kscali
### mIRT-Bayes New Measurement Model

Table 3. *Some analysis results for New Frog VPA Exploration 1*

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>INQUIRY/EXPLANATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cronbach's alpha overall</td>
<td>.88</td>
</tr>
<tr>
<td>MLE person separation reliability</td>
<td></td>
</tr>
<tr>
<td>Dimension 1: INQUIRY</td>
<td>.82</td>
</tr>
<tr>
<td>Dimension 2: EXPLANATION</td>
<td>.84</td>
</tr>
<tr>
<td>Estimated aprior/person variance reliability (EAP/PV)</td>
<td></td>
</tr>
<tr>
<td>Dimension 1: INQUIRY</td>
<td>.89</td>
</tr>
<tr>
<td>Dimension 2: EXPLANATION</td>
<td>.86</td>
</tr>
<tr>
<td>Correlation between dimensions (adjusted by ConQuest for attenuation)</td>
<td>.48</td>
</tr>
</tbody>
</table>

Poll Everywhere: Text 575507 and your message to 37607; OR login PollEv.com/kscalise
mIRT-Bayes New Measurement Model

Poll Everywhere: Text 575507 and your message to 37607; OR login PollEv.com/kscalise
mIRT-Bayes New Measurement Model

Standard Errors: INQUIRY
SEM substantially reduced under mIRT-bayes

Poll Everywhere: Text 575507 and your message to 37607; OR login PollEv.com/kscalise
Why Educators Care

• I said I would discuss why educators care.

• **Juan D’Brot:** “Tend to value what we measure but haven’t always been able to measure what you value.”

• **Eric Crane:** “Sometimes when I am most puzzled by our decisions about evidence, I think you shouldn’t be allowed to be in the testing industry if you haven’t been in the classroom.”

• **Christensen, Horn & Johnson:** Predict by 2019 that 50% of all high school courses will be delivered online.
Going Viral Among Teachers

- “I Choose C”

Source: “I choose C” web address http://www.youtube.com/watch?v=dY2mRM4i6tY

Poll Everywhere: Text 575507 and your message to 37607; OR login PollEv.com/kscalise
Lost Opportunities

- QC practices: Savagely attacking individual items when as a measurement community we know the individual observable is fragile. What we care about is inference from a set of evidences.

- Fit statistics: Often discarded when we know teachers highly value instructional information about students performing differently, who may need intervention.

- Scale: New approaches to what it means to bring an item bank to scale. Can a smaller number of sufficiently complex tasks or activities be made “transparent” to teachers?

- Usage: A requirement that even summative have effective instructional classroom uses.

Poll Everywhere: Text 575507 and your message to 37607; OR login PollEv.com/kscalise
Students:

Trends in Routine and Nonroutine Task Input in U.S. Occupations: 1960 to 2002

http://dx.doi.org/10.1787/5kmhds6czgzq-en
## Teacher Eval & RTT Funds

### Evaluation Best Practices (Administrator/Teacher Colleague)

<table>
<thead>
<tr>
<th>High EBP/High SL</th>
<th>High EBP/Low SL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low EBP/High SL</td>
<td>Low EBP/Low SL</td>
</tr>
</tbody>
</table>

Poll Everywhere: Text 575507 and your message to 37607; OR login PollEv.com/kscalise
Even at the Olympics... 

Education could be your wisest investment.
Acknowledgements: This material is based upon work supported by the U.S. National Science Foundation under Grant Nos. 0737056 (CCLI) and 0722014 (REESE), and by Cisco, Intel and Microsoft through ATC21S, atc21s.org.
What is an Innovative Item?

Modified Bennett Framework:
Intermediate Constraint Taxonomy


<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1A. True/False</td>
<td>2A. Multiple True/False</td>
<td>3A. Matching</td>
<td>4A. Interlinear</td>
<td>5A. Single Numerical Constructed</td>
<td>6A. Open-Ended Multiple Choice</td>
<td>7A. Project</td>
</tr>
<tr>
<td>1B. Alternate Choice</td>
<td>2B. Yes/No with Explanation</td>
<td>3B. Categorizing</td>
<td>4B. Sore-Finger</td>
<td>5B. Short-Answer and Sentence Completion</td>
<td>6B. Figural Constructed Response</td>
<td>7B. Demonstration, Experiment, Performance</td>
</tr>
<tr>
<td>1C. Conventional Multiple Choice</td>
<td>2C. Multiple Answer</td>
<td>3C. Ranking and Sequencing</td>
<td>4C. Limited Figural Drawing</td>
<td>5C. Cloze-Procedure</td>
<td>6C. Concept Map</td>
<td>7C. Discussion, Interview</td>
</tr>
<tr>
<td>1D. Multiple Choice with New Media Distractors</td>
<td>2D. Complex Multiple Choice</td>
<td>3D. Assembling Proof</td>
<td>4D. Bug/Fault Correction</td>
<td>5D. Matrix Completion</td>
<td>6D. Essay and Automated Editing</td>
<td>7D. Diagnosis, Teaching</td>
</tr>
</tbody>
</table>

Intermediate Constraint Taxonomy: [http://pages.uoregon.edu/kscalise/taxonomy/taxonomy.html](http://pages.uoregon.edu/kscalise/taxonomy/taxonomy.html)