

Pigeonhole Questions and Comments
Paper 3. O'Reilly and Sabatini

Thank you for inviting us to present at the roundtable at Teacher's College. Below are the set of Pigeonhole questions we received and our subsequent responses in *italics*.

Q: How are LOA different from traditional assessments?

A: Traditional assessments are designed to maximize the precision of measurement of the construct in question. While LOA's are also designed to produce high quality scores, they serve additional functions. These include modeling and supporting skill development as well as measuring partial understanding that can be used to inform instruction.

Q: Gottlieb and others talk about the need to design assessments with ELLs in mind from the beginning, rather than retrofitting them - can you speak to this?

A: We agree that quality assessment begin with careful preplanning. Assessments that are designed for ELLS should utilize the latest research and best practices from the start.

Q: How is the construct of interest defined in the GISA?

A: The construct measured in our assessments is described in:

Sabatini, J., O'Reilly, T., & Deane, P. (2013). *Preliminary Reading Literacy Assessment Framework: Foundation and Rationale for Assessment and System Design*. (Research Report No. RR-13-30). Princeton, NJ: Educational Testing Service.

As well as in:

Sabatini, J., & O'Reilly, T. (2013). Rationale for a New Generation of Reading Comprehension Assessments. In Miller, B., Cutting, L., & P. McCardle (Eds), *Unraveling Reading Comprehension: Behavioral, Neurobiological, and Genetic Components*, (pp. 100-111). Baltimore, MD: Brookes Publishing.

O'Reilly, T., & Sabatini, J. (2013). *Reading for Understanding: How Performance Moderators and Scenarios Impact Assessment Design* (Research Report No. RR-13-31). Princeton, NJ: Educational Testing Service.

Q: To what extent was teacher input utilized when creating the GISA? How about standards (e.g., Common Core)?

A: Prior to any framework development, we convened feedback from a panel of high school, community college and 4 year universities faculty to discuss the construct of reading and how it should be measured. After we designed the assessments we solicited feedback from teachers on the design and content of the forms. This feedback was helpful in revising our ideas and designs

to make them more appropriate and useful for both teachers and instruction. While the assessment is not directly designed to measure the skills in the common core, it is consistent with a broad scope and spirit of this progressive approach and we did consult Common Core in drafting our frameworks.

Q: What role does background knowledge have in topic-driven assessment?

A: Our GISA assessments are thematic and contain content area reading materials. To the extent that background knowledge plays a larger role in such circumstances, we are exploring ways to measure it in the assessment context. This may help clarify the interpretation of the scores.

Q: Impressive tasks. Some tasks can provide good modeling for reading instruction for teachers. Did you gather any teachers' feedback on your tasks?

A: Thank you. Yes, as stated above, we solicited feedback from teachers prior to the design of the assessments and after they were constructed. Their feedback was very valuable in improving our assessments. We continue to solicit feedback from teachers throughout the project to further enhance the utility of our future designs.

Q: Is it natural that these purpose-driven tasks are tested through MC questions? Does the purpose get lost in the MC structure, without production?

A: We understand the historical criticism surrounding the use of multiple choice questions to measure reading. While many of our items do use selected response format, they are carefully designed to address many of the limitations people have expressed in the past. We also include some constructed response items in the design to encourage the production and generation of answers. We attempt to strike a balance, while maintaining feasibility in large-scale applications. The balance would shift when used entirely in classroom settings.

Q: What does 'providing estimates' of student motivation or background knowledge entail?

A: We are experimenting with ways to potentially measure these moderator variables in the test design. Providing estimates is another way to say we are trying to measure them. That is, returning scores on sections designed to measure motivation and background knowledge specifically, so users can estimate their impact on comprehension scores.

Q: Scenario-based assessments like this are very impressive, but how can classroom teachers (w/ limited resources) adapt this for smaller-scale uses while retaining the best qualities?

A: The assessments we produced are in the research phase. It is not an operational test. However, overtime, we hope that assessments like these will be common place in the future as school-based technology (e.g., tablets) and computer assessments become more common. They are administered via web, so web access is a usual pre-requisite.

Q: How are you addressing what the CCSS calls "intensive" reading? What are the reading sub-skills involved in that kind of reading?

A: Interested readers should consult our frameworks listed above, however we view reading as a broad construct that encompasses a wide range of skills. One of which is to build a coherent mental model of text that does require intensive reading.

Q: Is bar raising the future trend?

A: Depends upon who you are talking to. Some people say the new standards are already raising the bar.

Q: Many of the tasks described elicit operations other than (complementary to) reading - excellent. How do you account for & measure constructs like written production, lexical control, etc.

A: As described above, while many of our assessment tasks are measured with the selected response format, we also include production tasks such as summary writing into the design. Again, the frameworks represent our most complete response to this at present, but we are open to applying new ideas and research in future versions.

Q: What is Scenario-Based Assessment?

A: Scenario-based assessment is a technique designed to structure the assessment tasks, items and texts in an assessment context. It serves many functions including the ability to incorporate a reading purpose into the design, the use of multiple texts, increasing the social context, to provide modeling of key skills, account for the role of performance moderators, and potentially identify what parts of a more complex task students can or cannot do. It is our hope this design is more useful for both teachers and students.

The research reported here was supported by the U.S. Department of Education Institute of Education Sciences, Award No. R305F100005, to Educational Testing Service as part of the Reading for Understanding Research Initiative. The opinions expressed are those of the authors and do not represent views of the U.S. Department of Education or Educational Testing Service.