Designing Dynamic and Interactive Assessments for English Learners and Others which Directly Measure Targeted Science Constructs

Rebecca Kopriva, University of Wisconsin Madison
David Gabel and Cathy Cameron, Center for Applied Linguistics

The ONPAR projects are building and researching a series of dynamic and interactive prototype summative and formative type assessment tasks at the elementary, middle school and high school levels geared to near term use in large-scale assessment markets. The projects are investigating how effective meaning related to challenging coursework in science, mathematics, and receptive literacy might be conveyed using multi-semiotic representations where text is most often used in a supportive rather than primary role. They are also studying coaching in formative tasks, and salient data capturing and scoring schemes so that richer data can inform student learning processes and status relative to progression maps. The purpose of this work is to open up the task space using selected technological capabilities so that students with language and literacy challenges can access relatively sophisticated academic contexts and questions and be able to demonstrate their reasoning and conceptual abilities in ways that do not primarily rely on text.

In so doing, these projects have broken new ground in developing novel response approaches that are geared to communicating challenging cognitive skills and knowledge structures. These approaches include manipulating onscreen stimuli by building, modeling, assembling, classifying or drawing, or using relational templates to demonstrate complex causal inferences. As it turns out, these response environments are effective for both good and struggling readers, and, for a number of constructs, they seem to more directly measure latent cognitive abilities than reliance on indirect text-driven methods.

Assessment Design and the Conceptual Assessment Framework

In designing appropriate assessment forms for the ONPAR projects, the first step we consider is the intended inferential claims at the item aggregate level. To accomplish this, and depending on the purpose of the project, type of assessment (e.g. summative, formative) and subject and grade levels, relevant standards and learning progress maps are identified and analyzed in terms of pinpointing valued concepts, skills and cognitive processes the test maker is intending to target.

Next, preliminary student and evidence models are developed prior to task development. In our work the student model is two-fold. Relative to traditional item types (including tasks where items are embedded), our approach fundamentally alters the way item-level information is presented to students and the way response data are collected. Thus, we have found it is necessary to identify very specific and explicit targets that can be ultimately used at the task/item level, including the latent targeted skill or knowledge and the intended cognitive complexity. These targets become the ultimate reference anchor for the student models, which also informs and constrains our task development at almost every turn. These more detailed targets, along with our commitment to opening up the tasks to be broadly inclusive, form what we can call the first aspect or the substantive student model.

The second aspect of the student model is initially produced as we consider how targeted ideas might unfold. Here we identify the broad outlines of what items or tasks might need to consider that conform to the detailed targets and inclusive nature specified in the substantive aspect. This step involves unpacking valued processes, skills and concepts, and includes identifying what kinds of latent cognitive processes and structures, knowledge and skills might be involved to produce defensible observable evidence. Some of these latent schemas will be target driven; many will be necessary to build and explain the problem or provide structure to how the problem solving is funneled. These latter schemas will ultimately constrain the observed task/item in some way, by providing context, identifying needed prior knowledge or skill, allowing broad access, or imposing some kind of structure on the response environments students might use. In this aspect the relationships between the targeted latent processes or concepts and the other processes are initially specified with the objective of defining preliminary probability distributions of how the schemas should impact the accumulation of evidence and the development of tasks. Most often these relationships among latent processes begin with 'maximize this or minimize that'.

Concurrent with the preliminary task level designs are the aggregate student model specifications that involve both aspects named above. To date most of the aggregation work has been in building tasks with multiple items.

Initial evidence models flow from the student models and provide preliminary thinking about how the valued processes and concepts/skills might be demonstrated in observable items/tasks (and ultimately in assessments) in order to support the intended inferences. The statistical submodel links the latent relational intents of the various processes/concepts explained above to how the processes and concepts might play out in observed relationships in the items and tasks. At this point we have been focusing on how barrier-producing variables impact responses (e.g. reading or task irrelevant prior knowledge), how various support aspects of tasks and items and might be mitigating these relationships, and how to use these solutions and communicate usually text heavy complex concepts in alternative ways. The submodels are also designed to evaluate the impact of these irrelevant influences on the how well the ONPAR approach in the tasks and items might be facilitating students communicating their targeted latent understandings, skills and underlying schemas back to test makers.

After the preliminary work in the student and evidence models have been conceptualized, task development begins. Essentially, the unpacking continues and item writers identify salient item/task ideas and then design, screen by screen, how particular latent targets (and other irrelevant but necessary latent traits) might manifest. The first observable part of task development is the completion of an initial storyboard, with notes about how different pieces of each screen are being used to express or capture latent target information and the intended relationships between relevant and irrelevant screen elements. From here programmers and designers work with the item writers to build the tasks/items and several reviews are interspersed. In all, the ONPAR approach to building tasks/items is to convey meaning to or from the student over four environments. The environments each serve a different function and systematically meet the student, interactively unpack the problem with the student, introduce the target questions, and set up and accept student work within the response environments.

Research Results

A large number of cognitive labs have been undertaken to evaluate and adjust a wide range of ONPAR techniques. Results from a recently completed randomized experimental study in elementary and middle school science indicate that, when academic ability was controlled, low English proficient ELLs and non-ELLs performed similarly on tasks which utilized ONPAR techniques, while there were significant differences in how these students performed on more traditional approaches. A second study in grades 4 and 7 mathematics used a randomized experimental design to investigate how LDs and poor readers, as well as low English proficient ELLs, perform on ONPAR as compared to the traditional test measuring the same content. Preliminary results show that when adjusted for student math ability, the interaction between test type (ONPAR versus traditional) and reading level (poor vs. good) is significant with a combination of LDs and poor readers but not for native English speaking good readers, and that when controlling for math ability the LD/poor readers and good readers performed similarly on the ONPAR tasks. The latest study and several cognitive labs have also suggested that several of the ONPAR techniques are actually better at capturing data about complex skills and concepts than traditional methods for all students. A current project with high school students in biology and chemistry will test this out for different types of targeted constructs.