



What Happens When Large-Scale Items Actually Use Computer Capabilities?

Exploring Issues and Redefining Challenges

Rebecca Kopriva
University of Wisconsin, Madison
rkopriva@wisc.edu

David Gabel
Jim Bauman
Center for Applied Linguistics



Setting the Stage

This presentation will suggest several outstanding measurement issues that arise as educators contemplate using computer capabilities to tap the knowledge and skills of students.

Some recently developed interactive science and math items will illustrate some of the points.

- These items come from a pair of federally funded grant projects, called ONPAR, out of UW and Center for Applied Linguistics.
- The ONPAR projects focus on investigating interactive items that are effective for students with low English proficiency skills, but this presentation is aimed at identifying issues important for all students.
- The items are designed to take the same amount of time as simple or extended constructed response items.



Issue 1

Edys's work has shown us that using computer capabilities can mimic classroom or wildlife labs in formative assessment.

- a) As we think about using these in large-scale assessments, and especially as we think about using them in measurement situations that go beyond extended inquiry tasks, some questions are raised.....
- b) This is particularly the case when the interactive items allow students to *respond* in ways that are different from how data are now collected.



Issue 2

It is important to know about what interactive items are measuring relative to what we are doing now because it seems that we should be asking:

- Over time, what are the implications in the general assessment of changing item types or different approaches to measuring the same latent constructs?
- What are the implications of collecting different kinds of observations for different students?



Issues 1 and 2

Following are two examples:

Each are part of a pair that includes a computer interactive and a traditional item.

Within pairs, these items are not the same, but it could be argued that they measure the same science targets and that the science they are ‘invoking’ are at the same level of cognitive complexity:



Issues 1 and 2

- Mislevy's ECD suggests collecting different kinds of observations, over time or over students, is plausible.

The questions are:

- What kinds of evidence need to be collected to determine what is being measured?
and
- What kinds of evidence need to be collected to determine when is it 'enough' the same?



Issue 3

In thinking about the cognitive demands associated with traditional and interactive items, it appears that there are at least three aspects that need to be considered:

- *Directness* to the latent construct underlying the content target
- *Density* of the cognitive demands
 - Construct relevant
 - Construct irrelevant
- *How* target cognitive schemas are engaged

Examples



Issue 4

- Does probability of correct response make a difference?
- What are the implications when that probability changes dramatically?

Examples



Issue 5

- What are the tradeoffs within and across different item approaches in how communication devices are used?
 - Asking the target question
 - Test taker providing a response



Issue 6

- This is not a new issue, it's just one we haven't dealt with very much...
 - When is a cr item similar or different 'enough' from an mc?
 - When are performance demonstrations similar or different 'enough' from other forms of communication?
 - What about portfolio information?