Ed334x – Assignment (Nov 30, 2001)
Yunn Chyi Chao

Rich’s comments in red…
Decker’s comments in green
Q & A:

1. To decide whether you will focus on Bugs or Worms, do a quick design review of both assessments, report your findings and make a recommendation to Gray and Delaine as to which assessment you recommend (will focus on).

A. Design Review of the Performance assessment “Bugs”

- The conceptual Framework: This is a (science) content lean and process open assessment where the investigation could be conducted in a number of ways, some legitimate, some less so. And not much prior knowledge is needed except some common knowledge from daily experience. E.g. playing with bugs, how to create a dark environment, etc. Yep.

- Conceptual Analysis:
The “Bugs” concentrates on procedural and strategic knowledge

<table>
<thead>
<tr>
<th>Performance Task</th>
<th>Procedural (How to..)</th>
<th>Strategic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>design and conduct a scientific investigation</td>
<td>Recognize when she has reached the valid/reliable result. OK… may be procedural but since this is probably novel, I go with your analysis.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Response format</th>
<th>Write steps in procedure Draw pictures depict the setup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scoring System</td>
<td>Method of determining outcome Setup investigation Procedure of investigation</td>
</tr>
</tbody>
</table>

- Instructional Goals/ the science standards:
Apart from the standard set forth by the California Science education standards, the instructional goals of the bugs assessment are:
  - To Develop reasoning and inquiry skills
  - To Design and conduct sound scientific experiment
  - To Transfer their prior knowledge
  - To Develop skills and appreciation for careful observation and measurement.
• **Reliability:** Inter-rater reliability was 0.65 for a single rater (Baxter & Shavelson, 1994). (Inter-task is irrelevant in this case. Why? Suppose 2 or 3 tasks aren’t enough to learn what student can do because of partial knowledge?)

• **Validity:**
  
  **Content validity:** Validity studies showed that the content of the assessment fit the standards, at least the first two investigations. However, statistical evidence showed that the third investigation was beyond the capacity of the students and, consequently, the recommendation was to use the first two tasks (Shavelson, Gao & Baxter, 1995). In addition, the instructional goals of the lesson parallel with what “Bugs” is measuring. **OK**

  **Cognitive Process:** “Bugs” assess students ability to construct a plan to initiate the strategy, devise the strategies to solve the problem, monitor their thinking and reasoning while performing the tasks. **Well, supposed to... any empirical evidence does this?**

  **Performance level:** the assessment correlated with other achievement (multiple choice) measures positively, around 0.50, as expected. **Good.**

  **Exchangeability for other methods:** Performance assessment is exchangeable for direct observation. “Bugs” is a content lean PA, no evidence indicates that it is exchangeable for MC and MR. **OK**

B. Design Review of the Performance assessment “Worms”

• **The conceptual Framework:** This is also a (science) content lean and process rich assessment. **Well, a bit complex because the experiments students are to design and do are process open...**

• **Conceptual Analysis:** The PA concentrates on procedural and strategic knowledge **See comments on Bugs Table**

<table>
<thead>
<tr>
<th>Performance Task</th>
<th>Procedural (How to..)</th>
<th>Strategic</th>
</tr>
</thead>
<tbody>
<tr>
<td>design and conduct a scientific investigation</td>
<td>Recognize when she has reached the valid/reliable result.</td>
<td></td>
</tr>
<tr>
<td>Response format</td>
<td>Write steps in procedure</td>
<td></td>
</tr>
<tr>
<td>Scoring System</td>
<td>Planning and designing, making inference, observing</td>
<td>Care in procedure</td>
</tr>
</tbody>
</table>

To date, the researchers are in the process of collecting and analyzing the reliability and the validity of the assessment.
**Recommendation:** I recommend we focus on studying how to convert the “Bugs” assessment into a computer simulation. The main reason is research has been carried out for “bugs”. It is a well studied performance assessment and it has a relatively high inter-rater reliability and content validity. OK

2. **Build a storyboard for the assessment you’ve chosen (either Bugs or Worms, not both) that could be used for converting the hands-on assessment into a computer simulation.**

The Storyboard for performing “Bugs” experiment (only for part (a) Dark or Light) OK—I like what follows!

<table>
<thead>
<tr>
<th>Hands-on experiment</th>
<th>Computer Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. preparation</strong></td>
<td><strong>1. preparation</strong></td>
</tr>
<tr>
<td>students are provided five sow bugs and some equipment including a dish in which to conduct the investigation, a spray bottle with water, filter paper to create a damp condition, black strips of paper to create a dark condition, and a lamp to create a light condition.</td>
<td>On the screen, it emulates the real laboratory environment. It has a (3D) workbench and a shelf to display the equipment. The equipment displayed including a dish, a spray bottle with water, filter paper, and black strips of paper. The lamp is sitting on the workbench (with a on/off button on it). All elements has roll-over text to indicate what it is.</td>
</tr>
<tr>
<td><strong>2. To investigate “Dark or Light” (A scenario)</strong></td>
<td><strong>2. To investigate “Dark or Light” (A scenario)</strong></td>
</tr>
<tr>
<td>(1) The student get a <strong>dry</strong> dish. Cover half the dish by dark strip and have the lamp shine on the other half of the dish. (i.e. half dark and half bright)</td>
<td>(note: To begin, the lamp is “off” and sitting on the workbench)</td>
</tr>
<tr>
<td>(2) Place a sow bug in the dish</td>
<td>Click on the dish (on the shelf). The dish appears on the workbench.</td>
</tr>
<tr>
<td>(3) Observe where will the sow bug go.</td>
<td>Click on the black strip. It appears on the workbench.</td>
</tr>
<tr>
<td></td>
<td>If student drags the black strip close to the dish, a menu pops up with options: cover all, half, or 1/3?</td>
</tr>
<tr>
<td></td>
<td>Student clicks on “half” and the screen displays</td>
</tr>
</tbody>
</table>
the black strip covers half of the dish.
- Turn the light on by clicking on the on/off button (toggle). Animate the light source, how it shines on the workbench and so does the dish.
- Student moves the half covered dish under the lamp – now it is half dark and half bright
- Student click on the sow bug. A menu pops up with the number of bugs to choose 1, 2, 3, 4 or 5.
- Click on 1, a bug appears on the screen and movable (not dragging) “with” the mouse – moving the bug to the dish by moving the mouse. Click or hit enter, the bug is released. (it might drop on the workbench if it is not placed in the dish)
- Animate the bug moving, programmed it in a way so that it matches with the actual behavior of a sow bug

Nice plan…

3. Set forth a formative evaluation plan for the prototype assessment… What questions would you ask? How would things be sequenced?

1. Investigate the background of the subject: Ask for her Age, Science score, interest in science, computer skills, whether she has performed any similar experiments before? **OK**

2. To study the screen interface design: Without giving the subject any instructions, asks her to try manipulating on the screen and think aloud. Observe how the subject selects the elements and interacts with the elements. Ask her to interpret the elements display, what would be her preference, what are the likes and dislikes about the interface, suggestions. **Good**

3. To investigate how the subject carries out “bugs” using computer simulation: Again, allows the subject to manipulate the elements and ask her to think aloud how she carries out the experiment. Observe how she plan and conduct the investigation using the computer simulation. Ask what are the plans she will carry out the experiment, why choose the method, how to ensure/confirm your results? **Right…**

**What analyses would you do?**
Summarize the results and categorize under user interface, functionality and items (checklist) to redesign/improve for the next iterations. **OK… work this iteratively until design perfected… Yes!**

**And what would you hope to accomplish?**
The purpose of the user testing is to test and improve on the design, to discover how users perceive the design and most importantly, whether it serves the objectives of the experiment, i.e to assess what the “bugs” Performance assessment is designed for. The
process of the design of computer version will be iterative. The ultimate goal is to design a computer-based version that is exchangeable for the hands-on experiment. Yes!

4. *Assuming you have developed a spiffy computer-based assessment, what summative study might you recommend before using the assessment in practice? Describe briefly questions, methods, and expected findings.*

To study the reliability and the validity of the computer-based version of the PA, as well as the exchangeability of the computer version for the hands-on version.

Research Question:
- What is the reliability and validity for the computer version?
- Does the computer simulation exchangeable with the hands-on version? (testing for correlation)

Methods
- Quantitative method:
  - Randomly assign a group to a hands-on version then 2 hours later (approximately) switch that group to do the computer version. Vice versa.
  - Check the correlation between the two assessments.

- Qualitative methods could be employed at the same time to verify the reliability of the assessment as a supplement to the exchangeability testing.

OK… we talked a bit more about this in last class…

Well done, Yunn Chyi! Excellent analysis of the two assessments and sound, sensible, well-supported recommendations.