The Virtual Assessment Project

Jody Clarke-Midura
Harvard Graduate School of Education
vpa.gse.harvard.edu
vpa@gse.harvard.edu
Inquiry into authentic questions generated from student experiences is the central strategy for teaching and assessing in science (NRC, 1996).
NSES model of inquiry

• Identify questions that can be answered through scientific investigation (not independent of knowledge)
• Design and conduct a scientific investigation
• Use appropriate tools and techniques to gather, analyze, and interpret data
• Develop prescriptions, explanations, predictions, and models using evidence
• Think critically and logically to make the relationships between evidence and explanations
• Recognize and analyze alternative explanations and predictions
• Communicate scientific procedures and explanations
• Use mathematics in all aspects of scientific inquiry
the problem is

Science inquiry is difficult to measure with open-response and multiple-choice tests
call for new measures of inquiry

Paper-and-pencil tests:
• don’t measure inquiry well
• aren’t aligned with the standards

Policy-Level Initiatives:
• Multiple modes of assessment
• Innovative item types
• Computer-based assessments
historical perspectives

• problems with previous hands-on performance assessments:
  ▪ Not as reliable or cost-effective as multiple choice tests (Stecher & Klein, 1997)
  ▪ Task sampling variability (Shavelson, Baxtor, and Gao, 1993)
  ▪ Occasion-sampling variability (Cronbach, Linn, Brennan, & Haertel, 1997)
  ▪ Not enough time spent piloting tasks
we’ve been down this road before…. 
about the project

• Funded by the Institute of Education Sciences (IES)
• Supplement to existing state science assessments
• Summative performance assessments middle school science
• Proof of concept study
what is a virtual performance assessment (vpa)?

- Assessment in a 3-D immersive environment
- Simulation of an authentic, real-world setting
- Measurement of complex problem solving through inquiry
how the vpa works

- Students’ take on identity of a scientist
- Students’ gather evidence
- Summative assessment
- 55 Minutes
how the vpa is different

• Traditional assessments:
  • focus on individual test items
  • rely on student affirmation as a response that indicates knowledge.

• Our VPAs:
  • evaluation of student performance captured as in-world interactions.
  • students are required to make a series of choices as a part of an ongoing narrative.
• live demo
avatar selection
immersive environment
the problem
competing hypotheses

A six legged frog!

I knew his Pesticides would cause this!

No! It's all the Pollution at his farm!

It's obviously a Genetic Mutation from her UV lamps.

It could be a Virus or something like that!

Wait! Maybe it isn’t a frog at all. Maybe it’s an Alien!
You need to investigate what caused the six legged frog.

- Gather evidence from the farms
- Use the internet kiosk for research
- Do tests on your evidence at the lab
- Talk with the locals

Return to me once you think you know what caused the six legged frog and have enough evidence to support it!
visit the farms
built in feedback
collect data
collect data
be strategic about choices
conduct research
conduct research

Frog Mutations: Pesticides

Commonly used pesticides disrupt the development of frogs, weaken their immune systems, delay or stunt development, cause mutations including cancerous growths and limbs and otherwise contribute to declining frog populations.

Evidence

Water: A number of pesticides show up in pond water due to runoff from nearby fields. Common pesticides found include:

Blood: Pesticides concentrate in the blood and tissue cultures in frogs and can be seen in tests. At sufficient concentrations it will lower white blood cell count.

Tadpoles: Pesticides such as Atrazine stop the metamorphosis from tadpole to frog. Stunted growth will result in shorter tails in tadpoles when compared to healthy tadpoles.

Frogs: Pesticides can stunt the growth of frogs making them smaller than healthy frogs. They can grow cancerous growths and other mutations including limb growth. This is thought to be caused by a weakened immune system which makes the frogs susceptible to viruses and bacteria.
visit the lab

Welcome to the lab!
There are a number of experiments you can perform on the evidence you have collected. You may return at any time to complete more tests. All results will be available to use after the tests from buttons that appear on the right.

Choose a lab test below to obtain results on the items you have collected.

- Water Lab
  Tests the components of water samples

- Blood Lab
  Tests the components of the frogs blood

- DNA Lab
  Tests the components of the frogs DNA
conduct tests
conduct tests
conduct tests
final claim
final claim
select evidence

Claim: Parasites caused the six legged frog

Select the items in your backpack that support your claim.
select evidence from lab tests
select evidence from lab tests
select evidence from lab tests

Claim: Parasites caused the six legged frog

Select which data supports your claim from the blood lab results.

<table>
<thead>
<tr>
<th>Blood Test: Analysis of Frog Blood Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Components:</td>
</tr>
<tr>
<td>Plasma</td>
</tr>
<tr>
<td>Red Blood Cells</td>
</tr>
<tr>
<td>White Blood Cells</td>
</tr>
<tr>
<td>Other</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

| Yellow frog       | Normal          |
| Green frog        | Normal          |

None of the results support the claim
what if they didn’t collect data?
select evidence from all data
select evidence from all data
select evidence from all data
select evidence from all data
select evidence from all data

**Claim: Parasites caused the six legged frog**

Select which data supports your claim from the **blood lab results**.

<table>
<thead>
<tr>
<th>Component</th>
<th>Blue Frog</th>
<th>Red Frog</th>
<th>Yellow Frog</th>
<th>Green Frog</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Red Blood Cells</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>White Blood Cells</td>
<td>Abnormally Low</td>
<td>High</td>
<td>Abnormally Low</td>
<td>Abnormally Low</td>
</tr>
<tr>
<td>Other</td>
<td>Atrazine (Pesticide)</td>
<td>Benzene Sulfur dioxide</td>
<td>Atrazine (Pesticide)</td>
<td></td>
</tr>
</tbody>
</table>

None of the results support the claim
write up final claim
what we are measuring

• Inquiry centrally involves theorizing and investigating.

• Students collect data about phenomena and proposed evidence-based findings about causal factors.
cognitive model of inquiry
what we are measuring

Student develops a scientific explanation of what is happening that includes: 1) a claim about the phenomena, 2) the evidence (either empirical or observations), and 3) reasoning that links claims with evidence.

Student gathers data that help explain or provide evidence to justify the claim being made.
how do we design interactions?
From Behren’s et al. (in press). Evidence Centered Design for Learning and Assessment in the Digital World.
design templates

Harvard VPA Principles of Scientific Inquiry | Design Pattern 2572

Overview
The following is a general design pattern for the Virtual Performance Assessment project. It is intended to provide an overarching framework for how the VPA project is conceptualizing scientific inquiry. More specific VPA design patterns cover the details of the different phases of the assessment.

Focal Knowledge, Skills, and Abilities
PK1. Ability to formulate questions that initiate investigations and experiments.
PK2. Ability to develop hypotheses based on a testable question.
PK3. Ability to design systematic observations and controlled experiments (identify and control variables).
PK4. Ability to identify the type of data to be collected.
PK5. Ability to gather and organize data.
PK6. Ability to evaluate and interpret data.
PK7. Ability to draw conclusions from data to answer investigation questions.
PK8. Ability to review and select the most salient data to support conclusions.
PK9. Ability to critically analyze alternate hypotheses based on evidence.
PK10. Ability to identify cause and effect relationships based on evidence.
PK11. Ability to differentiate explanations from descriptions.

Rationale
R1. Performance assessments of scientific inquiry require that the student demonstrate the ability to carry out the various steps included in scientific investigations.

Additional Knowledge, Skills, and Abilities
AK1. Prior knowledge of scientific inquiry
AK2. Definition of what a conclusion is
AK3. Difference between observation and inference
AK4. Prior knowledge of ecological causal systems
interactions as basis for assessment

Logfiles Indicate with Timestamps

- Where students went
- With whom they communicated and what they said
- What artifacts they activated
- What databases they viewed
- What data they gathered using virtual scientific instruments
- What screenshots and notations they placed in team-based virtual notebooks
- What hints they accessed
Interactions as basis for assessment

**Database of Logdata** - Track students’ behaviors: where they went, what data they collected, path to solve problem
match in world interactions to rubrics

<table>
<thead>
<tr>
<th>Question</th>
<th>Skill</th>
<th>observable variable</th>
<th>Evidence</th>
<th>score</th>
</tr>
</thead>
<tbody>
<tr>
<td>question 1 final</td>
<td>Claim/Reasoning</td>
<td>20</td>
<td>55</td>
<td>claim pollution</td>
</tr>
<tr>
<td>question 2 final</td>
<td>Evidence</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>add item for 21</td>
<td>Evidence</td>
<td>31</td>
<td>1</td>
<td>dead bee</td>
</tr>
<tr>
<td>add item for 21</td>
<td>Evidence</td>
<td>31</td>
<td>4</td>
<td>green bee</td>
</tr>
<tr>
<td>add item for 21</td>
<td>Evidence</td>
<td>31</td>
<td>8</td>
<td>green larvae</td>
</tr>
<tr>
<td>add item for 21</td>
<td>Evidence</td>
<td>31</td>
<td>10</td>
<td>lab nectar</td>
</tr>
<tr>
<td>add item for 21</td>
<td>Evidence</td>
<td>31</td>
<td>13</td>
<td>green nectar</td>
</tr>
<tr>
<td>question 3 final</td>
<td>Experiment: Water</td>
<td>22</td>
<td>13</td>
<td>green nectar</td>
</tr>
<tr>
<td>question 3 final</td>
<td>Experiment: Water</td>
<td>22</td>
<td>10</td>
<td>lab nectar</td>
</tr>
<tr>
<td>question 4 final</td>
<td>Experiment: DNA</td>
<td>23</td>
<td>4</td>
<td>green bee</td>
</tr>
<tr>
<td>question 4 final</td>
<td>Experiment: DNA</td>
<td>23</td>
<td>1</td>
<td>six bee</td>
</tr>
<tr>
<td>question 5 final</td>
<td>Experiment: Blood</td>
<td>24</td>
<td>1</td>
<td>six bee</td>
</tr>
<tr>
<td>question 5 final</td>
<td>Experiment: Blood</td>
<td>24</td>
<td>4</td>
<td>green bee</td>
</tr>
<tr>
<td>question 6 final</td>
<td>All data: Evidence</td>
<td>25</td>
<td>6</td>
<td>green larvae</td>
</tr>
<tr>
<td>question 7 final</td>
<td>All data: Evidence</td>
<td>26</td>
<td>4</td>
<td>green bee</td>
</tr>
<tr>
<td>question 8 final</td>
<td>All Data: Experiment</td>
<td>27</td>
<td>13</td>
<td>green nectar</td>
</tr>
<tr>
<td>question 9 final</td>
<td>All Data: Experiment</td>
<td>28</td>
<td>60</td>
<td>no DNA results</td>
</tr>
<tr>
<td>question 9 final</td>
<td>All Data: Experiment</td>
<td>28</td>
<td>4</td>
<td>green bee</td>
</tr>
<tr>
<td>question 9 final</td>
<td>All Data: Experiment</td>
<td>28</td>
<td>1</td>
<td>six bee</td>
</tr>
<tr>
<td>question 10 final</td>
<td>All Data: Experiment</td>
<td>29</td>
<td>1</td>
<td>six bee</td>
</tr>
<tr>
<td>question 10 final</td>
<td>All Data: Experiment</td>
<td>29</td>
<td>4</td>
<td>green bee</td>
</tr>
</tbody>
</table>
what we learn about students

• Observations of student learning not possible with a multiple choice test

• Real-time analysis of student paths

• Assessments that capture the learning process in addition to the product
student feedback...
did this feel like a test?

<table>
<thead>
<tr>
<th>yes</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>131</td>
<td>473</td>
</tr>
</tbody>
</table>
yes....

• it did because you had to work at it
• i felt like it was a test because we had to solve a really big problem & see why there were six leg frogs.
• I had to use my head to find the answer
• I loved it! it was a game and a test put together it's brilliant! I think we should use this as the test instead of paper and pencil!
no...

• It didn't feel like a test because of the way I tested things and it felt more like an experiment.
• No because tests aren't fun and this was fun to me. It also felt like a game.
• It didn't because this way of learning is a fun way to understand science.
did you feel like you were conducting an experiment?

<table>
<thead>
<tr>
<th>yes</th>
<th>no</th>
</tr>
</thead>
<tbody>
<tr>
<td>447</td>
<td>149</td>
</tr>
</tbody>
</table>
no....

• It was really easy to me and to me conducting an experiment is supposed to be more complicated.

• It's a game I was playing. I did have to gather data and find out the answer to the problem, but the "experiment" feeling wasn't there.
yes...

- i did because we were figuring out what caused the 6 legged frog and i could try differnt things.

- I felt like I was conducting an experiment because I really did feel like a scientist. Now I know that mean sound sort of corny, but it's true. Scientists can chose how they solve an experiment or the chose how they go about doing it. I really felt like I could chose how I solved it.
advantages of vpas over performance assessments

• Standardizing the administration of hands-on performance assessment is difficult, so extensive training is required. In contrast, VPAs ensure standardization by delivering instruction to students in an identical manner via the technology.

• VPAs alleviate the need for developing, shipping, and providing schools with materials and kits for hands-on tasks. All that is needed are a computer and an internet connection.
advantages of vpas over performance assessments

• Scoring will all be done by the technology, so no raters are necessary, reducing cost, training, and the possibility of human error.

• VPAs potentially have fewer problems with task and occasion variability.
advantages of vpas over item-based tests

- Multiple choice, short answer, and essay questions do not present a realistic context within which to elicit complex performances.
- VPAs mirror the types of processes to which teachers should orient instruction more accurately than do conventional measures.
- The use of test-taking strategies can distort the outcomes of conventional item-based measures, but prior studies suggest that this may not be the case with VPAs.
- VPAs can seamlessly incorporate features to minimize the importance of prior content knowledge and can track the extent to which a student utilizes these.
- VPAs provide a more detailed record of student actions than do conventional item-based tests.
what’s next?
reliability studies

- Cognitive task analyses
- Cognitive flow charts
- Interaction Analysis using Item Response Theory
- Differential Item Functioning
- Attribute Hierarchy Method (AHM)
- Bayes Nets
we need you!!!

- fall 2011
- A lot of students
  - 6th, 7th, 8th grade
- Pre-content survey (~15 mins)
- Pre-Affective survey (~15 mins)
- Technology use survey (~5 mins)
- Randomly assigned to assessment (~55 mins)
- Feedback survey (~5 mins)
generalizability study

• Two weeks later
• Take second assessment (~55 mins)
• Feedback survey (~5 mins)
contact info:

Jody Clarke-Midura
vpa.gse.harvard.edu
jody_clarke@mail.harvard.edu
Thank you.