Lessons Learned When Developing Evaluation Instruments for Inquiry-Based Aquatic Science Teacher Professional Development Modules

American Evaluation Association
Conference Symposium
October 26, 2012

Accessible Professional Development for Teaching Aquatic Science Inquiry, an IES Professional Development Project

- Preordinate design but responsive process
- Learned about details of instruments as we progressed
- CF CIPE issue and Evaltalk discussion
- Topics
  - PD and its evaluation
  - Lessons learned in instrument development
  - Process use

Overview of the Teaching Science as Inquiry-Aquatic Professional Development Project and Its Evaluation

Joanna Philippoff
Kanesa Seraphin
Paul R. Brandon

Teaching Science As Inquiry
Accessible Professional Development Structure

TSI Aquatic PD
Teaching Science As Inquiry
A Theoretical Framework

Phases of Inquiry
- Learning and instructional cycle
- Reflection of what happens when doing science
- Multi-directional
- Importance of community

Goals
Teachers will:
- Develop an understanding of the fundamentals of inquiry teaching and learning from the perspective of disciplinary science
- Learn how to integrate TSI teaching practices that facilitate scientific inquiry into their classroom
- Improve their self-efficacy in teaching science through the process of inquiry
The TSI-A Evaluation

Years 1–2

- Provided feedback to PD developers.
- Developed, pilot-tested, and validated instruments
  - Context, implementation, outcomes
- Tried instruments out with Cohorts I-III

Instrument Foci

- Professional development activities (observations & questionnaires)
- Teachers (assessments, questionnaires, log, interview)
- Students (assessment, questionnaires)

Pre- & Post-Project Tasks

Module 1
- Student questionnaire
- Teacher assessment and questionnaire
- Teacher interview
- Student questionnaire pretest the previous spring

Module 2

Module 3

Module 4

Pre

Post

Module Tasks

- Pre
- Post
- Teacher assessment
- Repeat each module

How Address Inquiry-Based Aquatic Science?

- Construct validity of the instruments
- Replaced concept maps, which did not work
- Observations too expensive
- Chose teacher vignettes
Overview

- We present our story, lessons learned, and the developments we made in:
  - Methods of collecting data to measure teachers’ knowledge of inquiry-based teaching
  - Methods of scoring teachers’ responses on these instruments

Inquiry-based science instruction

What do teachers know?
- Concept maps

What can teachers do?
- Observations

Vignette Development

- Vignette
  - To score
  - Inquiry Teaching Assessment (ITA)
  - Analytic rubric
  - Short analytic rubric
  - Holistic rubric

Development
Vignette Assessments

- A vignette describes a hypothetical scenario or situation, often provided in a written format.
- The scenario or situation described in a vignette requires the respondent to provide some judgments and reasoning often in the form of “what would you do?” (Barter & Renold, 1999; Ruiz-Primo & Li, 2003; Seguin & Ambrosio, 2002; Veal, 2002; Wason, Polonsky & Hyman, 2002)

Procedures for Development

- Developed vignette scenarios
- The scenario described must be realistic and plausible (Barter & Renold, 1999; Seguin & Ambrosio, 2002; Wason, Polonsky, & Hyman, 2002; Ruiz-Primo & Li, 2003)
- Drew from existing inquiry-based science curricula, and revisions were made in discussions with both teachers and science content experts
- Based on this feedback we decided to use a single vignette presenting a scenario about teaching Buoyancy and Density

Pilot Testing of the Vignette

- Administered online using survey software
- Three experts and the first cohort of teachers (n=17) in our PD
- Experts and teachers differed in the nature of their demonstrations of their breadth and depth of knowledge and practice of inquiry-based science in their responses.

Experts’ vs Teachers’ Responses

- Experts focus their writing demonstrating their understanding rather than practice of inquiry-based science teaching
- Teachers focus their writing demonstrating their practice rather than understanding of inquiry-based science teaching

Procedures for Development

- What do teachers know?
- What can teachers do?
- Information not relevant to our construct
- Single Vignette
- Experts
- Teachers
Examined Teachers’ Feedback

Three of the teachers were concerned about their lack of content knowledge.

Several comments indicated the vignette could be responded to with varying levels of specificity.

Two teachers recommended we use several shorter scenarios rather than a single vignette.

Examined Teachers’ Feedback

Three of the teachers were concerned about their lack of content knowledge.

Several comments indicated the vignette could be responded to with varying levels of specificity.

Too broad

Two teachers recommended we use several shorter scenarios rather than a single vignette.

Examined Teachers’ Feedback

Three of the teachers were concerned about their lack of content knowledge.

Several comments indicated the vignette could be responded to with varying levels of specificity.

Two teachers recommended we use several shorter scenarios rather than a single vignette.

Examined Teachers’ Feedback

Three of the teachers were concerned about their lack of content knowledge.

Several comments indicated the vignette could be responded to with varying levels of specificity.

Too broad

Revamp our instrument

Two teachers recommended we use several shorter scenarios rather than a single vignette.

Lessons Learned

• Using a single large vignette solicited a large amount of irrelevant information.
• Even though teachers were asked to describe the beginning, middle, and end of their lesson, the vignette did not capture teachers’ breadth of knowledge and practice.
• Both experts’ and teachers’ comments also suggested that the method felt burdensome, challenging to comprehend the task and time consuming.
• As a result, we concluded that our vignette instrument was unwieldy for the project.

Vignette Response Instructions

• You are the teacher, so you can alter the lab’s procedure in any way that you think is appropriate. You have unlimited supply of the materials listed in the lab.
• Describe what you would do from the beginning to the end of the lesson to ensure that it is inquiry-based, and state why you would do those things. Focus on how you would direct students along the way and on the questions (if any) that you would ask students at different steps of the process. Consider what you would do or say at the beginning of the lesson, in the middle, and at the end. What do you expect students to do and how would you deal with misconceptions or with students who get unexpected results?
Inquiry Teaching Assessment (ITA)

- Seven mini-vignettes, each comprising:
  - a classroom scenario, with a description of the context and the lesson objective,
  - a multiple-choice (MC) component asking which teaching approach would be best for making the lesson inquiry-based, and
  - a constructed-response component asking teachers to explain their answer choice.

Inquiry Teaching Assessment (ITA)

- Pedagogy of Science Inquiry Teaching Test (POSIIT) by Schuster & Coburn (2010).
- Adapting the POSITT, we developed the ITA as a viable alternative to the single vignette.

Lessons Learned

- The format and structure has been improved, to better fit the needs of the teachers
- ITA provides shorter scenarios with different content and different aspects of inquiry

Inquiry-based science instruction

- What do teachers know?
- What can teachers do?

Single vignette
Inquiry-based science instruction

What do teachers know?

What can teachers do?

What another teacher can do, rather than the teachers themselves

ITA

Lessons Learned

• The length and depth of teachers’ responses are still of concern with the ITA
• When to administer seems to be an important factor

Rubric Development

(Analytic rubric on vignettes)

Scoring vignette responses

To score

Analytic rubric

Short analytic rubric

Holistic rubric

Inquiry Teaching Assessment (ITA) responses

To score

To score

To score

Inference

Provide an inference for measuring

Vignette scores

Lessons Learned

The length and depth of teachers’ responses are still of concern with the ITA.
When to administer seems to be an important factor.

Rubric Development

(Analytic rubric on vignettes)

Scoring vignette responses

To score

Analytic rubric

Short analytic rubric

Holistic rubric

Inquiry Teaching Assessment (ITA) responses

To score

To score

To score

Inference

Provide an inference for measuring

Vignette scores
Analytic Rubric Development

- Inquiry-based science instruction
  - Interpreting TSI theory documents
- List of inquiry-based instructional practices
  - Discussions with and feedback from TSI experts
  - Revising
- Rubric category descriptors
  - Four categories of degree of knowledge within each component of the construct

Analytic Rubric Development

- Experts' responses
  - Rubric category descriptors
- Disagreements in classifying themes and segments from responses into components on the rubric

Vignette Rubric Lessons Learned

- Construct underrepresentation
- Could not capture experts' expertise
- Difficulty classifying responses into fine-grain categories
- Construct-irrelevant variance
- Difficulty interpreting teachers' responses
- Decisions about what the teacher meant
- Rater disagreements
- Difficulty scoring responses
- Time and effort

What We Did

- Shortened the component descriptions
- Removed the rating category descriptors and their example responses
- Reduced number of categories from four to three

Table

<table>
<thead>
<tr>
<th>Rubric Category Descriptors</th>
<th>Scoring</th>
<th>Experts' Responses</th>
<th>Pilot-Testing Teachers' Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpreting and classifying segments of the responses into components required too much time and cognitive load</td>
<td>Experts' responses provided too little evidence for rating in each component</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Examples

<table>
<thead>
<tr>
<th>Experts' Responses</th>
<th>Pilot-Testing Teachers' Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experts' responses did not neatly match upper category descriptors</td>
<td></td>
</tr>
</tbody>
</table>

Interpreting and classifying segments of the responses into components required too much time and cognitive load.
Scoring the ITA Responses

- We assumed we could score:
  - The constructed responses
  - Using the existing analytic rubric
  - The multiple-choice (MC) responses
  - To key it, four project personnel took the test
    - But, was disagreement on the best choice
      - Inquiry is achievable in seemingly non-inquiry teaching activities
Scoring the ITA

Sample Responses

Sample Responses

What We Did

Analytic Rubric Lessons Learned

ITA Analytic Rubric Scores

• Construct underrepresentation
  • Could not capture misconceptions
  • Could only capture the phases of inquiry
  • Construct-irrelevant variance
  • Only moderate interrater reliability

What We Did

• Made the instructions clear that we were assessing teachers’ knowledge of inquiry-based science instruction
• Created a holistic rubric to capture not only the phases of inquiry, but the modes, the demeanors of scientists, the role of the teacher, and misconceptions
• This also allowed the rubric to evolve with the project

• Scored responses of 27 teachers in the pre-PD condition
• MC scoring yielded very low reliability
• Consistent with the principle that inquiry-based teaching is possible in teaching activities that appear to not be inquiry
• Treated the MC prompt as part of the stem
• Excluded teachers’ MC responses from the assessment
• We scored teachers’ responses using the shortened analytic rubric.
• Two raters with 27 teachers’ responses.

Scoring the ITA

• Constructed responses to the mini-vignette prompts
• Multiple-choice responses

Sample Responses

• Evidence of the degree of knowledge of inquiry-based science teaching was difficult to often sparsely.
• For example, our first prompt says “Please explain why you selected this teacher’s approach for the lesson. In your explanation, show what you know about teaching science as inquiry.” Here are four teachers’ responses to this prompt:
  1. “This approach allows the most student led inquiry process”
  2. “Open-ended: Creating a self-directed class.”
  3. “This is student based and learning comes from them.”
  4. “To me approach A is guided inquiry. It works best with a young group of students. Approach B is so open ended that you don’t know what you will end up with. Students could make all sorts of things out of corks and bottles but not end up with a thermometer. Approach C and D are more along the lines of authoritative modes, not much exploration.”

Analytic Rubric Lessons Learned

• Insufficient evidence in responses
  • We edited the instructions
• Some responses indicated misconceptions
  • We decided to use negative scores
• Covered only some aspects of inquiry-based teaching
  • We considered a holistic rubric
  • Moderate interrater reliability ($r = .57$)

ITA Analytic Rubric Scores

• Scoring key
  • Inquiry CR score
  • Inquiry MC score
  • Total score

Covered only some aspects of inquiry-based teaching

Sample Responses

• Evidence of the degree of knowledge of inquiry-based science teaching was difficult to often sparsely.
• For example, our first prompt says “Please explain why you selected this teacher’s approach for the lesson. In your explanation, show what you know about teaching science as inquiry.” Here are four teachers’ responses to this prompt:
  1. “This approach allows the most student led inquiry process”
  2. “Open-ended: Creating a self-directed class.”
  3. “This is student based and learning comes from them.”
  4. “To me approach A is guided inquiry. It works best with a young group of students. Approach B is so open ended that you don’t know what you will end up with. Students could make all sorts of things out of corks and bottles but not end up with a thermometer. Approach C and D are more along the lines of authoritative modes, not much exploration.”

Analytic Rubric Lessons Learned

• Construct underrepresentation
  • Could not capture misconceptions
  • Could only capture the phases of inquiry
  • Construct-irrelevant variance
  • Only moderate interrater reliability

What We Did

• Made the instructions clear that we were assessing teachers’ knowledge of inquiry-based science instruction
• Created a holistic rubric to capture not only the phases of inquiry, but the modes, the demeanors of scientists, the role of the teacher, and misconceptions
• This also allowed the rubric to evolve with the project

ITA Analytic Rubric Scores

• Scoring key
  • Inquiry CR score
  • Inquiry MC score
  • Total score

Covered only some aspects of inquiry-based teaching

Sample Responses

• Evidence of the degree of knowledge of inquiry-based science teaching was difficult to often sparsely.
• For example, our first prompt says “Please explain why you selected this teacher’s approach for the lesson. In your explanation, show what you know about teaching science as inquiry.” Here are four teachers’ responses to this prompt:
  1. “This approach allows the most student led inquiry process”
  2. “Open-ended: Creating a self-directed class.”
  3. “This is student based and learning comes from them.”
  4. “To me approach A is guided inquiry. It works best with a young group of students. Approach B is so open ended that you don’t know what you will end up with. Students could make all sorts of things out of corks and bottles but not end up with a thermometer. Approach C and D are more along the lines of authoritative modes, not much exploration.”

Analytic Rubric Lessons Learned

• Construct underrepresentation
  • Could not capture misconceptions
  • Could only capture the phases of inquiry
  • Construct-irrelevant variance
  • Only moderate interrater reliability

What We Did

• Made the instructions clear that we were assessing teachers’ knowledge of inquiry-based science instruction
• Created a holistic rubric to capture not only the phases of inquiry, but the modes, the demeanors of scientists, the role of the teacher, and misconceptions
• This also allowed the rubric to evolve with the project

ITA Analytic Rubric Scores

• Scoring key
  • Inquiry CR score
  • Inquiry MC score
  • Total score

Covered only some aspects of inquiry-based teaching

Sample Responses

• Evidence of the degree of knowledge of inquiry-based science teaching was difficult to often sparsely.
• For example, our first prompt says “Please explain why you selected this teacher’s approach for the lesson. In your explanation, show what you know about teaching science as inquiry.” Here are four teachers’ responses to this prompt:
  1. “This approach allows the most student led inquiry process”
  2. “Open-ended: Creating a self-directed class.”
  3. “This is student based and learning comes from them.”
  4. “To me approach A is guided inquiry. It works best with a young group of students. Approach B is so open ended that you don’t know what you will end up with. Students could make all sorts of things out of corks and bottles but not end up with a thermometer. Approach C and D are more along the lines of authoritative modes, not much exploration.”

Analytic Rubric Lessons Learned

• Construct underrepresentation
  • Could not capture misconceptions
  • Could only capture the phases of inquiry
  • Construct-irrelevant variance
  • Only moderate interrater reliability

What We Did

• Made the instructions clear that we were assessing teachers’ knowledge of inquiry-based science instruction
• Created a holistic rubric to capture not only the phases of inquiry, but the modes, the demeanors of scientists, the role of the teacher, and misconceptions
• This also allowed the rubric to evolve with the project

ITA Analytic Rubric Scores

• Scoring key
  • Inquiry CR score
  • Inquiry MC score
  • Total score

Covered only some aspects of inquiry-based teaching

Sample Responses

• Evidence of the degree of knowledge of inquiry-based science teaching was difficult to often sparsely.
• For example, our first prompt says “Please explain why you selected this teacher’s approach for the lesson. In your explanation, show what you know about teaching science as inquiry.” Here are four teachers’ responses to this prompt:
  1. “This approach allows the most student led inquiry process”
  2. “Open-ended: Creating a self-directed class.”
  3. “This is student based and learning comes from them.”
  4. “To me approach A is guided inquiry. It works best with a young group of students. Approach B is so open ended that you don’t know what you will end up with. Students could make all sorts of things out of corks and bottles but not end up with a thermometer. Approach C and D are more along the lines of authoritative modes, not much exploration.”

Analytic Rubric Lessons Learned

• Construct underrepresentation
  • Could not capture misconceptions
  • Could only capture the phases of inquiry
  • Construct-irrelevant variance
  • Only moderate interrater reliability

What We Did

• Made the instructions clear that we were assessing teachers’ knowledge of inquiry-based science instruction
• Created a holistic rubric to capture not only the phases of inquiry, but the modes, the demeanors of scientists, the role of the teacher, and misconceptions
• This also allowed the rubric to evolve with the project

ITA Analytic Rubric Scores

• Scoring key
  • Inquiry CR score
  • Inquiry MC score
  • Total score

Covered only some aspects of inquiry-based teaching
Rubric Development (holistic rubric on ITA)

Scoring vignette responses
  To score
Analytic rubric
  Short analytic rubric
  Holistic rubric

Inquiry Teaching Assessment (ITA) responses
  To score

Holistic rubric
  Analytic rubric

Evidence of low to high degree of knowledge
  No evidence of knowledge
  Evidence of misconceptions

Analytic rubric
  Holistic rubric

Results with Holistic Rubric Rating
- Higher interrater reliability than with the analytic rubric. With two raters, rating 27 teachers on 7 mini-vignette items:

<table>
<thead>
<tr>
<th>Score Type</th>
<th>g-coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analytic</td>
<td>.57</td>
</tr>
<tr>
<td>Holistic</td>
<td>.74</td>
</tr>
</tbody>
</table>

Results with Holistic Rubric Rating
- With the entire set of pre- and post-PD responses, the g-coefficient was .82
- We also discovered that if we collapsed our nine rating categories into four categories, we could:
  - reduce the rating burden even further, and
  - arrive at the comparable reliability.
- With the four categories, reliability would still be adequate (g-coefficient = .80).
Several smaller constructed-response tasks were more appropriate than a single essay-type response. More samples of teachers’ knowledge across content-knowledge and classroom scenarios.

To measure this complex, hard-to-define construct, a holistic rubric was more appropriate than an analytic rubric.

Improved construct representation

The process of developing the analytic rubric led us to recognize what was missing. This helped us develop the holistic rubric.
EXAMINING EVALUATION PROCESS USE AS A RESULT OF PROJECT DEVELOPERS’ PARTICIPATION IN AN INQUIRY-BASED SCIENCE PROGRAM EVALUATION

Brian Lawton, Joanna Philippoff, & Kanesa Seraphin
Curriculum Research & Development Group
University of Hawai‘i at Mānoa

American Evaluation Association
Minneapolis, Minnesota
October 26, 2012
Session 616; Room 102A

Introduction

- Process use
- Method
  - Participants
- Results
  - The effects of the evaluation on the developers and vice versa
- Lessons learned

Process Use

- Introduction
  - Patton (1997)
- Three broad types (Amo & Cousins, 2007)
  - Learning
  - Changes in actions or behavior
  - Changes in affect or attitude

Methods

- Initial one-on-one interview
- Then focus group
- Interview/Focus group guiding questions
  1) Level of involvement/background
  2) Changes in thinking about the evaluation process as a result of involvement
  3) Changes in behavior as a result of involvement.
  4) Changes in attitude as a result of involvement

Participant Background and Level of Involvement

- Scientists
- Previous evaluation experience
  - General understanding of evaluation practice
- Ed Psych Ph.D. student
  - Some influence of classes
- Involvement from the beginning and how it has evolved

Results

Effects of process use on the evaluation

- Learning
- Changes in behavior
- Changes in attitudes
Results: Learning and the Effects on the TSI Evaluation

- Evaluation requirements
- Instrument use & results
- Formative feedback

- Setting criteria to teach inquiry; difficulty in measuring behavior
- More critical of items; item development
- How to self-critique their own process; better assess teacher practice

- Evaluation now has another critical eye; accountability
- Clear, measurable, & consistent components

Results: Changes in Behavior and the Effects on the TSI Evaluation

- Evaluation requirements
- Instrument use & results
- Formative feedback

- Continuous teacher exercises to ensure evaluation needs
- Use of own instruments
- More explicit about feedback provided to teachers

- Increased teacher expectations
- Additional results for project impact

Results: Changes in Attitudes and the Effects on the TSI Evaluation

- Review of instrument design, results & evaluation process

- Importance of evaluation process
- Patience
- Skeptical of findings
- Theory vs. application

- Increased interest and involvement in the evaluation process

Lessons Learned (Summary)

- Greater alignment between evaluators’ needs and developers’ practice
- Greater understanding about what goes into an evaluation and it is not always a nice, clean process
- Are we really going to make a difference, and if so, can we see it?
- We both learned from each other, now we need to see what’s missing...

References


Questions or Comments

Paul Brandon: brandon@hawaii.edu
Joanna Philippoff: philippo@hawaii.edu
George Harrison: georgeha@hawaii.edu
Lisa Vallin: vallin@hawaii.edu
Brian Lawton: blawton@hawaii.edu
Kanesa Seraphin: kanesa@hawaii.edu