Handout #1

Project Evaluation

Eileen Lewis & Bert Holmes
March 19, 2013
Sheryl Sorby & Ginger Rowell
April 10, 2013
Ginger Rowell & Bert Holmes
April 24, 2013

Important Notes

- Most of the information presented in this workshop represents the opinion of the IWBW project team and not an official NSF position.
- Participants may ask questions using the *QUESTION BOX* on the meeting screen.
- Responses will be collected from a few sites at the end of each Group Activity. At the start of the Group Activity, we will identify these sites in the *CHAT BOX* and then call on them one at a time to provide a few of the ideas their group discussed.

Preliminary Comments on Workshop

- More than a set of guidelines on evaluation
- Intended to change the way you think about evaluation.
  - Improve your understanding
  - Help you learn
- Engagement makes learning more effective
  - Good learners are not simply listeners.
- Active, collaborative process to improve learning

Active & Collaborative Learning

- Effective learning activities
  - Recall prior knowledge – actively, explicitly
  - Connect new concepts to existing ones
  - Challenge and alter misconceptions
  - Reflect on new knowledge
- Active & collaborative processes
  - *Think* individually
  - *Share* with partner
  - *Report* to local and virtual groups
  - *Learn* from presenter's response
  - *Learn* from the IWBW team's response

Participant Activities

Two types of activities

- Group Activity ~ 6 min
  - Think individually ~ 2 min
  - Share with a partner ~ 2 min
  - Report in local group ~ 2 min
  - Report to virtual group
    - A few institutions selected
    - Check Chat Box for your Institution's name
- Individual Activity ~ 2 min

Workshop Goals and Expected Outcomes

Goal: Enhance the participants' understanding of evaluation concepts and methods so that they can more effectively work with an evaluator in addressing this important component in preparing proposals or in implementing funded projects.

Expected Outcomes: Participants will be able to:

- Identify stakeholders who would be interested in formative and summative evaluation results
- Compare the advantages and disadvantages of surveys and interviews, as an example of a comparison between two instruments.
- Define factors to be considered regarding an evaluation instrument before it is selected for use.
- Discuss potential confounding factors in interpreting evaluation data.
- Identify the strengths and weaknesses of an evaluation plan and suggest improvements.

Two Purposes of Evaluation

- Formative Evaluation
  - Provides information for project improvement as the project is underway
- Summative Evaluation
  - Assess quality and impact at the end of the project.

Formative Evaluation: Two Components

- Implementation Evaluation
  - Premise: before you can evaluate the project outcomes/ impacts, you must examine how it is operating and whether some modification is needed
  - Assess whether project is being conducted as planned
    - Describes and documents project activities
    - Early check if essential elements are in place
    - Identifies strengths and weaknesses of different strategies
- Progress Evaluation
  - Assess progress in meeting project's ultimate goals
    - Collects Information to assess if benchmarks are met, determine impact of activities and strategies
    - Early indicator of achieving project goals.
    - Changes can be made if progress is not being made
    - Data collected can form basis for summative evaluation

Summative Evaluation

- Summative Evaluation
  - Assess quality and impact of fully implemented project
  - May have same questions as progress evaluation (but at end of project)
  - Examines project's potential to continue (sustainability)
  - Examines contributions to broader knowledge base

Individual Activity: Stakeholders

Think of an educational project that you are currently conducting or you are thinking about developing

- List potential stakeholders for this project that would be interested in/benefit from
  - Formative evaluation results
  - Summative evaluation results

Think individually ~ 2 min and write your responses

Handout #2

Response: Stakeholders

- Stakeholders:
  - Have an interest in the project outcomes
  - Could benefit from the project outcomes in a variety of ways
  - Could be interested in the project outcomes at various stages of the project
- Important to identify stakeholders in the design phase of the project

Response: Stakeholders

- Formative Evaluation
  - *Project PIs* should be very interested in Formative Evaluation results because it helps to improve the project
    - You would not conduct a complex bench experiment without monitoring certain aspects while under way
  - *Other Stakeholders* who might be interested in Formative Evaluation results
    - Funding Agencies
    - School Administrators
    - Educators

Response: Stakeholders

- *Stakeholders* who might be interested in Summative Evaluations results
  - Potential other users
  - Reviewers of journal articles and conference papers
  - Funding Agencies
  - School Administrators
  - State Legislatures
  - Accrediting Agencies
  - Broader STEM Education Community
  - Parents

Types of Instruments

- Evaluation is about measuring changes in student learning or student behavior
- Learning causes changes in a student's knowledge, skills or attitudes
  - Changes in cognitive and affective behavior
- Existing instruments measure these changes
  - Instruments can provide
    - Quantitative data – numerical
    - Qualitative date – text or narrative

Examples of Tools for Evaluating Learning Outcomes

- Surveys
  - Forced choice (multiple-choice) or open-ended responses
- Interviews
  - Structured (fixed questions) or in-depth (free flowing)
- Concept Inventories
  - Multiple-choice questions to measure conceptual understanding
- Tests
  - Multiple-choice or open-ended to judge student knowledge and skills or gain in knowledge and skills
- Focus groups
  - Like interviews but with group interaction
- Observations in the class setting
  - Actually monitor and evaluate behavior

NSF's The 2010 User-Friendly Handbook for Project Evaluation
http://www.westat.com/Westat/pdf/news/UFHB.pdf

Group Activity: Comparing Instruments

A project proposes to teach Organic Chemistry using a new collaborative learning approach. It is expected that this new approach will improve students' interest in and attitude about science. The project team is trying to decide whether to use a survey or an interview process to judge the impact of the new approach on student's affective response toward science.

- Describe the advantages and disadvantages of these two approaches.
  - Think individually ~ 2 min
  - Share with a partner ~ 2 min
  - Report in a local group ~ 2 min

Handout #3

Response: Surveys

- Advantages:
  - Good for gathering descriptive data
  - Can cover a wide range of topics
  - Are relatively inexpensive to use
  - Can be analyzed using a variety of existing software tools
- Disadvantages:
  - Self-report may lead to biased reporting
  - Data may provide a general picture but lack depth
  - May not provide adequate information on context
  - Design of unbiased, effective survey can be time consuming

NSF's The 2010 User-Friendly Handbook for Project Evaluation
http://www.westat.com/Westat/pdf/news/UFHB.pdf

Response: Interviews

- Advantages:
  - Provides rich data, details, new insights
  - Provide opportunity to explore topics in depth
  - Allows observation of affective as well as cognitive aspects
  - Allows clarification through follow-up questions
  - Allows flexibility
- Disadvantages:
  - Expensive and time consuming
  - Need qualified, trained interviewers
  - Interviewee may distort information through recall error, selective perceptions, desire to please interviewer
  - Flexibility can result in inconsistencies across interviews
  - Volume of information very large

NSF's The 2010 User-Friendly Handbook for Project Evaluation
http://www.westat.com/Westat/pdf/news/UFHB.pdf

Examples of Established Evaluation Tools

- Concept Inventory
  - Series of multiple choice questions
    - Questions involve single concept
      - Formulas, calculations or problem solving skills not required
    - Possible answers include distractors
      - Common errors – misconceptions
- Pittsburgh Freshman Engineering Survey
  - Questions about perception
    - Confidence in their skills in chemistry, communications, engineering, etc.
    - Impressions about engineering as a precise science, as a lucrative profession, etc.
    - Study habits, motivation, etc.

Examples of Established Evaluation Tools (cont.)

- CAT Instrument is
  - Designed to assess critical thinking and real-world problem solving skills
    - Extensive development, testing, and refinement with a broad range of institutions, faculty, and students across the country
- Levels of intellectual development
  - Students see knowledge, beliefs, and authority in different ways
    - "Knowledge is absolute" versus "Knowledge is contextual"
      - Measure of Intellectual Development (MID)
      - Measure of Epistemological Reflection (MER)
      - Learning Environment Preferences (LEP)

Group Activity: Instrument Quality

Assume that you have an educational development project that uses problem-based learning and one of the expected outcomes is an improvement in critical thinking skills. You have found three existing instruments that could be useful.

- List the questions you would ask about these instruments in making your decision.
  - Think individually ~ 2 min
  - Share with a partner ~ 2 min
  - Report in a local group ~ 2 min

Handout #4

Response: Questions that Need to be Answered Before Using an Instrument

- Does it actually asses what you want to measure?
- Does it provide qualitative or quantitative data?
- How difficult is it to use?
- Does it require special skills?
- How much does it cost?
- How much time does it take?
- Has it been compared to other tools?

Response: Questions that Need to be Answered Before Using an Instrument

- Is it sensitive? Does it discriminate between a novice and an expert?
- Does it produce the same results for the same or similar groups of students?
- Has the tool been used by others besides the developer? At other sites? With other populations?
- Is there normative data?

Many of these questions deal with the instrument's reliability and validity – terms that evaluators use.

General Factors to Consider in Selecting Evaluation Tools

General factors:

- Match with the scope and nature of the project
- Time and cost constraints
- Required skill level
- Credibility of findings

NSF's The 2010 User-Friendly Handbook for Project Evaluation
http://www.westat.com/Westat/pdf/news/UFHB.pdf

Standards of Instrument Quality

- Reliability – Are measurements Consistent
  For example, does it give same results with same or similar cohort of students
  - Factors affecting reliability
    - Are results comparable on two or more separate occasions?
    - Are results equivalent on different forms of a test that are based on the same content?
    - Are results consistent across questions?
- Validity – Do measurements reflect what is intended
  For example, a written test is not a valid assessment of driving skills
  - Factors affecting validity
    - Does the content of the test measure stated objectives?
    - Do scores correlate to an outside reference?
    - Does the assessment correspond to other significant variables?
    - Does the assessment make sense, and is it seemingly correct to the expert reader?
- Are results similar across racial/ethnic, gender differences

http://en.wikipedia.org/wiki/Educational_assessment

Confounding Factors

- Educational research and evaluation are often complex and results can be highly variable
  - Lack of control by "experimenters" over external factors
    - Individual variability (multiple individuals)
    - External factors (confounding factors)
- Confounding factors can interfere with the results in the experiment
- Important to minimize the impact of confounding factors by the protocol used in collecting data

Interpreting Evaluation Data

Question	No. of Students		Percent with Correct Answer	
	Pre	Post	Pre	Post
1	28	32	23%	29%
2	28	32	34%	85%

An intervention was implemented to improve student understanding of fundamental concepts as measured by Questions 1 & 2 on a Concept Inventory. The data is presented in this table.

Individual Activity: Alternate Explanation for Change

Data suggests that the understanding of the concept tested by Q2 improved. One interpretation is that the intervention caused the change.

- List some alternative explanations
  - Confounding factors
  - Other factors that could explain the change

Think individually ~ 2 min and write your responses

Question	No. of Students		Percent with Correct Answer	
	Pre	Post	Pre	Post
1	28	32	23%	29%
2	28	32	34%	85%

Handout #5

Response: Alternate Explanation for Change

- Students learned concept outside of class
  - E.g., in another course or in study groups with students not in the course
- Students answered with what the instructor wanted rather than what they believed or "knew"
- An external event (big test in previous period or a "bad day" distorted pretest data
- Instrument was unreliable
- Data is not necessarily paired data – different students could have taken the pre/ post

Individual Activity: Alternate Explanation for Lack of Change

Data suggests that the understanding of the concept tested by Q1 did not improve. One interpretation is that the intervention did cause a change that was masked by other factors

- List some alternative explanations
  - Are these different from the confounding factors described in the previous part of this exercise?

Think individually ~ 2 min and write your responses

Question	No. of Students		Percent with Correct Answer	
	Pre	Post	Pre	Post
1	25	32	23%	29%
2	25	32	34%	85%

Handout #6

Response: Alternate Explanations for Lack of Effect

- An external event (big test in previous period) distorted post-test data
- The instrument was unreliable
- Implementation of the intervention was poor
- Population too small
- Data is not necessarily paired data—different students could have taken the pre/post
- Formats were different on pre and post tests

Role of Project Learning Goals and Expected Outcomes in Evaluation

- Learning goals provide overarching statements of what you *hope to achieve* with your project.
- Expected outcomes identify *specific observable or measureable results* for each learning goal.
- Expected outcomes are the basis for *evaluation questions*.

Evaluation

- An effective evaluation relies upon *clearly defined* project *learning goals* closely linked to expected *outcomes*.
- Learning goals and expected outcomes should be designed to produce changes in students related to
  - *Attitude or perception* about the subject matter (affective changes)
  - *Knowledge or skills* about the subject matter (cognitive changes)

Evaluation plan

The *evaluation plan* consists of the

- Evaluation questions
- Evaluation methods or protocol
- Evaluation instruments
- Person(s) responsible for conducting the data collection
- Person(s) responsible for analysis and interpretation of the evaluation data

Group Activity: Evaluation Plan

Read the Evaluation plan provided

- What are the strengths and weaknesses of the plan?
- What are some suggestions for improvement?
  - Think individually ~ 2 min
  - Share with a partner ~ 2 min
  - Report in a local group ~ 2 min

Handout #7

Response: Strengths of the Evaluation Plan

- Includes formative assessment to improve content, trainers, integration, and delivery
- Looks for unintended consequences
- Measures student engagement
- Plans comparisons across demographics, course grade, and education background in chemistry

Response: Weaknesses of the Evaluation Plan

- There are numerous weaknesses in this evaluation plan
- A few of the weaknesses are:
  - No measure of impact of labs on content knowledge until end of semester
  - No measures of content knowledge other than course grades
  - Does not use a comparison group
  - Dissemination/transportability will be affected by lack of comparison groups
- It is good practice to think about the questions reviewers will have when preparing an evaluation plan

Questions to Consider in Preparing an Evaluation Plan

1. Are the project's goals and objectives clearly articulated and measurable? (Is there a vision of what success would look like?)
2. Are the instruments and methods of evaluation appropriate and clearly described?
   - Is there a good match between the evaluation design and the project's goals and objectives?
   - Are the evaluation instruments and methods well designed?
   - Is the evaluation design appropriate for the nature and the scope of the project?

Questions to Consider in Preparing an Evaluation Plan (cont.)

3. Will evaluation data be used to shape project activities (i.e. formative assessment)?
4. Will the evaluation data be sufficient to support claims that are made?
   - Will the data be robust?
   - Will the data be ambiguous or anecdotal?

Questions to Consider in Preparing an Evaluation Plan (cont.)

5. Will the evaluation results be triangulated?
   - Unambiguous, mutually supporting or consistent results will be obtained from multiple assessment instruments or methods.
   - Multiple assessment instruments or methods will yield contradictory or unrelated results
   - Only a single assessment instrument or method will be employed.
6. Is the project itself, and its evaluation, of broad interest?

Thanks for your participation!

- This concludes the virtual session. Thanks for your participation.
- There will be a concluding local session where participants will reflect on their experiences in the virtual session
- All participants will receive an email message with a link to the post-workshop evaluation survey. Please go to the site and complete the survey so that we can identify areas for improvement and have information to report to NSF

Acknowledgement

- This workshop has been offered through a partnership between the American Association for the Advancement of Science (AAAS), Louisiana State University, and Higher Education Services, Inc.
- Support of this workshop has been through NSF grants DUE-1224063 & DUE-1224240

Evaluation References

- NSF 02-057: The 2002 User-Friendly Handbook for Project Evaluation, a basic guide to quantitative and qualitative evaluation methods for educational projects http://www.nsf.gov/pubs/2002/nsf02057/start.htm
- NSF 97-153: User-Friendly Handbook for Mixed Method Evaluations, a monograph "initiated to provide more information on qualitative [evaluation] techniques and ... how they can be combined effectively with quantitative measures" http://www.nsf.gov/pubs/1997/nsf97153/start.htm
- Online Evaluation Resource Library (OERL) for NSF's Directorate for Education and Human Resources, a collection of evaluation plans, instruments, reports, glossaries of evaluation terminology, and best practices, with guidance for adapting and implementing evaluation resources http://oerl.sri.com/home.html
- Field-Tested Learning Assessment Guide (FLAG): This website is designed for Science, Math, Engineering, and Technology Instructors who are interested in new approaches to evaluating student learning, attitudes, and performance. It has a primer on assessment and evaluation, classroom assessment techniques, discipline-specific tools, and resources - all in a searchable, downloadable data base, http://www.flaguide.org/.
- Student Assessment of Learning Gains (SALG): An on-line survey that measures student perceptions of their learning gains due to any components within a course. Faculty can modify a template to match any and all features of their courses, have their students take the survey on-line, and have the data returned to them as either raw data or with simple statistical analysis, http://www.salgsite.org/
- Developing Metrics for Assessing Engineering Instruction: What Gets Measured Gets Improved, a 2007 National Academies Press publication, http://www.nap.edu/catalog.php?record_id=12636
- American Evaluation Association. Online Resources (http://www.eval.org/resources.asp)
- Bond, S.L., Boyd, S.E., Rapp, K.A. 1997. *Taking Stock: A Practical Guide to Evaluating Your Own Program.* Chapel Hill, NC: Horizon Research, Inc.
- The Carnegie Foundation for the Advancement of Teaching. August 2008. The Faculty Inquiry Toolkit. (http://specctoolkit.carnegiefoundation.org/)
- Cornell Office for Research on Evaluation. 2009. The Evaluation Facilitator's Guide to systems Evaluation Protocol. Ithaca, NY; Cornell Digital Print Services.
- My Environmental Education Evaluation Resource (MEERA). University of Michigan (http://meera.snre.umich.edu/)
- FOOTPRINTS: Strategies for Non-Traditional Program Evaluation (NSF 94-51) (http://nsf.gov/pubs/1995/nsf9541/index.jsp).
- Online Evaluation Resource Library (http://oerl.sri.com/).
- Evaluation Handbook, W.K. Kellogg Foundation (http://www.wkkf.org/Pubs/Tools/Evaluation/Pub770.pdf).
- The American Evaluation Association (AEA) (http://www.eval.org/).
- 2010 User Friendly Program Evaluation Guide (http://www.westat.com/westat/pdf/news/ufhb.pdf)

Repositories of Evaluation Instruments

- FLAG -- Field-tested Learning Assessment Guide: homepage:
  - http://www.flaguide.org/
- SALG – Student Assessment of their Learning Gains: homepage
  - http://www.salgsite.org/
- OERL-- Online Evaluation Resource Library: homepage
  - http://oerl.sri.com/

Pre-workshop handout

Below is an overview of a proposed project.

This proposal will create a unique set of virtual reality (VR) experiments that can be used in lower division chemistry laboratories. These 3-D and fully interactive experiments use methods and technology that allows use in a large variety of institutions and locations. The goal of the proposal is to improve retention and recruitment of science students from diverse populations. The use of VR experiments will improve student understanding by increasing their engagement and content knowledge and provide access at a variety of diver sites.

Evaluation Plan

Formative evaluation will focus on improving chemistry offerings in terms of VR lab content, instruction of trainers, integration into curriculums, and delivery to student populations. It will also investigate any unintended consequences resulting from the program activities. Summative evaluation will be closely aligned with the overarching program goals, e.g. do students have increased engagement? On-site evaluation will include a continuous improvement feedback loop between program developers, faculty, and student users. End-of- course exams will include content from the labs and will be analyzed for knowledge retention. Instructors will be surveyed to identify perceptions of curriculum fit, support of alternate learning styles, and utility of VR lab instruments. Descriptive statistics on student and instructor results will be completed, including group comparisons on the VR lab experiments and participant demographics such as the grade/course being taught and the participant's educational background in chemistry.