Designing research to investigate student learning

Diane Ebert-May[1], Janet M Batzli[2], and Everett P Weber[1]

The call for evidence-based research in education has accelerated in recent years, accompanied by the need to examine the nature of inquiry into student learning. Donovan and Pellegrino (2003) make a strong appeal to improve the quality of classroom research, stating that "education needs high-quality research if the results are to be reliable for the purposes of improving practice". The challenge for a researcher in the classroom is, how can one generate valid and reliable data to address a hypothesis when the research venue is variable and data are messy? In order to address this question, educators must establish standard methods and experimental designs for research that can be applied in many classroom settings.

Our goal in this article is to analyze several different experimental designs and evaluate the validity of each approach in the context of an example research question. We build on the series of recent *Pathways* articles, which provided standard methods to assess student understanding and practical advice for implementation. In March, we demonstrated how instructors can search for patterns in students' thinking by using validated rubrics. In April, we extended these ideas by providing examples of how to code students' work. In this article, we will explain how we would design a research study to analyze how students approach an ill-structured problem (one with multiple approaches and solutions). We propose the following research question: "How can guiding questions associated with an ill-structured problem help students scaffold their knowledge and promote critical thinking and analysis?"

Faculty research goal

- Use both observational and empirical approaches to answer a question about student learning.

Student goals

- Use effective and repeatable processes to address ill-structured problems.
- Demonstrate critical thinking (refer to March *Pathways* article for student learning objectives for critical thinking).

Research design

We designed an ill-structured problem (Panel 1) focusing on the topic of phytoremediation, as inspired by Pilon-Smits and Freeman (pp 203–10 in this issue). This problem integrates ecological and evolutionary topics, including plant–microbe interactions; herbivore interactions and competition; biomagnification and environmental characteristics of toxins, natural selection, and evolution of pollution tolerance; and (depending on the instructor's expectations) transgenic plants and gene flow. Central to the research question is the construction of guiding questions used to facilitate students' engagement with this problem, and to provide instructors with insight into students' current and prior knowledge.

Systematic observation

Students address the guiding questions in groups during a large class or in smaller discussion or laboratory sections. Ideally, a research colleague is present to monitor students as they attempt to approach the problem, while the instructor continues to provide guidance. The observer moves through the room, making notes on, for example, the types of questions students are asking (confirmatory/clarification *or* probing/analytical), and how they are answering guiding questions (connecting knowledge from other situations, courses, and their personal experience, *or* quickly moving to text, computers, and classroom resources).

Systematic observation allows instructors to identify elements of the problem that students find difficult, helps illuminate misconceptions, or identifies gaps in students' knowledge. Both students and instructor need to practice using the guiding questions with different problems throughout the course, in order to make this problem-solving approach transparent and automatic. Student responses are further evaluated using a critical thinking scoring rubric (Facione and Facione 1994) given to students at the same time as the assignment. Systematic evaluation of responses helps instructors refine their understanding of students' abilities to draw upon prior knowledge to formulate salient hypotheses, draw warranted, judicious conclusions from their expected results, and explain their assumptions and reasons. This information becomes the basis for the next phase of the study, which is focused on comparison.

Comparison studies

We propose two (of many possible) experimental frameworks for investigating the effectiveness of guiding questions on problem-solving approaches to address ill-structured problems: multiple-group comparison and split-group comparison. Without doubt, the classroom research we are modeling takes place in messy settings, is situation dependent, and includes multiple variables that are difficult to

[1]*Michigan State University*, [2]*University of Wisconsin*

Panel 1. Example problem and activity about phytoremediation

(note: species and biology are accurate, but situation is hypothetical)

The City of Grand Junction, CO, has received a grant from the Environmental Protection Agency (EPA) for an innovative effort to reduce selenium (Se) pollution in a 3500-ha area of the Colorado River Basin, specifically the Uncompahgre and Lower Gunnison River. Selenium is a naturally occurring element that often attains abnormally high concentrations in soils and water contaminated by the waste matter from agriculture, urban development, and other human activities. High levels of Se have been shown to cause reproductive failure and deformities in fish and aquatic birds. Current concentrations of Se within the area are 15–400 parts per billion (ppb) in water and 40–70 ppb in soil, levels far in excess of the EPA criterion of 5 ppb for protection of aquatic life.

Scientists plan to establish a phytoremediation program based on the planting of a genetically modified variety of Indian mustard (*Brassica juncea*) that absorbs high levels of Se throughout the affected floodplain area. This variety has been tested in the laboratory and in greenhouses, but not in the field. Although this seems to be a promising initiative, there are still many questions about the scale of the *B juncea* planting and the environmental/ecological impact of long-term use of *B juncea* as compared to conventional remediation methods. What is the efficacy of this phytoremediation plan and what are the potential effects for the local ecology?

Part 1 (Guiding questions – written response)

1. What things do you know or think you know about this problem?
2. What things do you not know?
3. What things are not known in the scientific community studying similar problems?
4. What things can you find out, given available review papers, primary scientific literature, and data?

Part 2

1. Develop a testable hypothesis focusing on one aspect of the ecological impact of phytoremediation of Se with *B juncea*.
2. Develop a well-supported rationale for your hypothesis.
3. Design an experiment to test your hypothesis.
4. Create a figure showing your expected results.
5. Describe your assumptions that are intrinsic to your expected results.
6. Discuss how your assumptions may lead to alternative results.

control (Suter and Frechtling 2000); in short, a type of situation that is similar to a field ecologist's research setting.

An important aspect of this research is determining the internal and external validity of the study design. Internal validity addresses the fundamental question, "Does the experimental design allow the researcher to truly test the hypothesis? Or will other variables confound results?" External validity asks the question of generalizability, that is, "To what other populations and classroom settings do the results of this study pertain?" (Shadish *et al*. 2002).

Multiple-group comparison is used when an instructional intervention is implemented in multiple sections of a course, or in a single course taught by the same instructor over multiple years (Knight and Wood 2005). Students are not selected randomly for class sections, which may jeopardize internal validity. For example, students who chose the 8 AM section might be different than students who chose the 2 PM section, and this difference cannot be separated from the treatment effect. In one section, students receive homework that includes guiding questions requiring a written response (Panel 1); in another section, students receive the same homework without the guiding questions. Student responses are coded (see March *Pathways* article).

Split-group comparison is designed to statistically remove differences caused by non-random sampling. All students receive both treatments; this design therefore has higher internal and external validity than the multiple-group comparison, and is an effective design in large enrollment courses (Ruiz-Primo *et al*. 2002; Shadish *et al*. 2002). In this case, half of the students experience the intervention being tested, while the other half completes a comparable activity concurrently. The intervention is done at least twice during the course, in different contexts; groups switch activities. Student responses are coded as in the multiple-group comparison above. The split-group design is commonly called a split-plot design in agricultural studies. The counterbalancing in this design allows the researcher to remove variation in student performance due to external factors, such as previous coursework, by comparing the same student's performance in both treatments.

The class is randomly divided into two groups and both complete the problem as homework. One group writes responses to the guiding questions, while the other group does not receive or address the guiding questions. Key to the split-group design is another comparable problem that is assigned later in the semester, in which the treatments are reversed. The applicability of experimental results is dependent upon the course type, size, students' experience, and classroom environment. Other factors that may be important include institution type (community college, liberal arts college, or research university), and student demographics.

Final note

In this article, we describe the use of systematic observations of students' thinking that guided our selection of effective research designs for classroom studies. In the next and final article of this series, we will address the analysis and reliability of data collected using these types of designs as well as human subjects research approval. We will identify key questions about teaching and learning in undergraduate science that merit investigation and challenge our community of ecologists to become change agents who elevate the standards for this research.

References

Donovan MS and Pellegrino JW. 2003. Learning and instruction: a SERP research agenda. Washington, DC: National Academies Press.

Facione PA and Facione NC. 1994. Holistic critical thinking scoring rubric. Millbrae, CA: California Academic Press. http://www.insightassessment.com/pdf_files/rubric.pdf. Viewed 1 April 2006.

Knight JK and Wood WB. 2005. Teaching more by lecturing less. *Cell Biol Educ* **4**: 298–310.

Ruiz-Primo MA, Shavelson RJ, Hamilton L, and Klein S. 2002. On the evaluation of systemic science education reform: Searching for instructional sensitivity. *J Res Sci Teach* **39**: 369–93.

Shadish WR, Cook TD, and Campbell DT. 2002. Experimental and quasi-experimental designs for generalized causal inference. New York, NY: Houghton Mifflin Company.

Suter L and Frechtling J. 2000. Guiding principles for mathematics and science education research methods: report of a workshop. Washington, DC: National Science Foundation. Report 00-113.

