A LIVELY ELECTRONIC COMPENDIUM OF RESEARCH, NEWS, RESOURCES, AND OPINION

Astronomy Education Review

2010, AER, 9, 010104-1, 10.3847/AER2009047

Multidimensional Education Research: Managing Multiple Data Streams

Julia Kregenow
Physics Department, Ithaca College, Ithaca, New York, 14850
Michael Rogers
Physics Department, Ithaca College, Ithaca, New York, 14850
Mark Constas
Education Department, Cornell University, Ithaca, New York, 14853
Received: 07/28/09, Accepted: 12/4/09, Published: 02/18/10

Abstract

We discuss collecting data simultaneously from many different instruments, including both qualitative and quantitative sources. A list of instruments used is provided, and data collection methods are described. Many practical logistical concerns are discussed, warning of potential pitfalls to be avoided. This article is intended to be a practical guide for other research groups wishing to use data from multiple sources, especially if they have not collected qualitative data before.

1. INTRODUCTION

The Ithaca College Physics Department has implemented the SCALE-UP (Student-Centered Activities for Large Enrollment Undergraduate Programs) approach to teaching its large-enrollment introductory physics and astronomy courses. The SCALE-UP approach, developed at North Carolina State University (Beichner 2008), facilitates an interactive classroom by creating a student-centered and technology-enhanced learning environment for classes of more than 100 students. We are nearing the end of a multiyear project to test this approach compared to traditional lecture instruction. We describe the collection of several types of data over several years, including both qualitative and quantitative sources. Using different, overlapping sources of data allows comparison and validation of the individual sources. The combination of qualitative and quantitative also provides a rich, flexible data set. The following data were available for all students from before the semester began: SAT score (math and verbal), prior GPA, major, and class year. These are described in Section 2. We collected the following data from all students during the course: informed consent, diagnostic exams given pre- and post-instruction, Views on the Nature Of Science (VNOS) written questionnaire—given both pre- and post-instruction, seating location at various times throughout the semester, and seating preferences from a previous large lecture class. These are described in Section 3. We targeted a smaller sample of students in order to collect further data through interviews and written surveys. These data and the selection of this subsample are described in Section 4. Finally, we collected data about the classes in general, including class observations and instructor notes about instructional approaches and choices, as described in Section 5.

2. RESEARCH QUESTIONS, SELECTION OF INSTRUMENTS, AND DEMOGRAPHIC DATA

Our National Science Foundation funded work listed six research goals that form the bases of our research questions and design.

- Increase student conceptual knowledge of physics and astronomy for both majors and nonmajors.
- Develop student problem solving skills and ability.
- Increase student understanding and appreciation of the nature of science as a way of learning about the natural world and as the basis for our technological society.
- Create a technology-rich classroom that enhances active learning and experimentation simultaneously with conceptual understanding of physics and astronomy.
- Create curriculum materials that help us meet these goals and that other physics and astronomy educators can use.
- Understand the connections between curricular innovation and changes in students' conceptual knowledge of physics and astronomy.

Each of our goals have distinct aspects while overlapping with the other goals in significant ways. Our research design was constructed to evaluate the distinct aspects and understand the intersections between goals. To understand student conceptual understanding, problem solving ability, and understanding of the nature of science we selected existing instruments that are well used or tested for validity and reliability or both. Questionnaires, reflection logs, and interviews were selected to help us understand the impact a technology-rich classroom and curricular innovation have on student learning. Interviews helped us understand the intersections between goals.

At the core of our research design is running two sections each of our algebra-based introductory physics course and our general education astronomy courses. One section of these courses is held in a lecture hall and taught in a traditional manner, while the other section is held in our SCALE-UP (herafter referred to as PbP, for Performance-based Physics) classroom using innovative approaches. To facilitate comparison between the traditional sections with the innovation sections, it is desirable to have the demographic profiles of the students in each section be comparable. The college's Office of Institutional Research provided data on all students registered in the course as of the beginning of the semester, including prior GPA (not including the current course, not specific to major or nonmajor), SAT scores (math, verbal, and writing if available), declared major (if any), and expected graduation date. These data are available upon request to qualified research projects. Brogt *et al.* (2008) suggest a code of ethics for astronomy education research that proposes researchers only collect demographic data necessary to answer their research questions to protect the privacy of the participants. Our research questions require that the two study populations be comparable and if not the differences accounted for, which is why we asked for the demographic data that we received.

3. DATA COLLECTED FROM ALL STUDENTS

3.1. Informed Consent

The project research staff visited all classes during the first week of the semester and explained the purpose and scope of the study to the entire class. Federal regulation 45 CFR 46 (U.S. Department of Health and Human Services 2005) as administered by the college's Human Subjects Review Board (HSRB) office requires all prospective student participants in the study to be provided with an informed consent form detailing all aspects of the study and any risks and benefits of participation. Note that the HSRB office may have different names at different institutions. Institutional Review Board is common, but there are other variants such as Ethical Review Board, Human Subjects Protection Office, etc. HSRB requirements vary by institution as well and, at a given office, may even vary by semester! If you are even considering collecting data from students for research, be sure to check with your HSRB early and often to be sure you have the proper permissions and forms. This must be done ahead of time. If you are thinking about publishing the results of your pedagogical endeavors you will need HSRB approval. Talking with your HSRB sooner instead of later will make the process smoother and ensure that you are complying with federal regulations. It is important to note that failure to comply with federal regulations will result in denial or termination of federally awarded funds (U.S. Department of Health & Human Services 45 CFR 46.122/123 2005). Brogt, Dokter, and Antonellis (2007) and Brogt *et al.* (2007) are excellent sources to learn more details about HSRB requirements and ethical considerations when conducting education research.

As per Federal and HSRB requirements, students were informed that they could change their mind about participating at any time. Two copies of the consent form were distributed to all students, each section of the form described, and students were given the opportunity to ask questions about the form and the study. Project staff emphasized that participation in the study would have no effect on the students' course grades, and the instructor would not even know who was participating during the semester. It was important not to have the instructor explain the study to the students to emphasize that the study was separate from the class, the instructor, and their grades. The informed consent form asked students to elect one of the following options regarding their participation in the study: (1) choosing to participate in the study and over 18 years old, (2) choosing to not participate in the study, or (3) unable to participate in the study because under 18 years old. See sample form in Appendix A. Students were asked to sign one copy of the informed consent form and return it to project staff indicating their choice and to keep the second copy of the form for their own records. Project staff collected and archived the forms, keeping information about who was participating confidential from the course instructor during the semester. The fraction of students consenting from each class is detailed in Table 1. Of all students represented in the 12 classes reported overall, 87% granted consent for participating in the study, and only 3% explicitly declined to consent. Another 7% were absent on the day consent forms were given, 1% signed the consent form but forgot to check the box indicating their consent, and 1% signed the consent form but were under 18 and so could not participate. (Note that our HSRB *does* allow informed consent from students under 18, but it would have also required parental consent. Since this applied to such a small fraction of our sample, we chose to forgo this complication and simply exclude this population.)

Table 1 summarizes the number of participating students from the last three semesters of data collection, when data were collected from multiple sources. (Previous semesters are not listed because the only source of data was the pre- and post-instruction diagnostic test.) The participation rates are also given as percentages and indicate that the majority of enrolled students consented to participate in the study, and completed both the pre- and post-instruction diagnostics. The diagnostic exams, though given to all students, do not show 100% participation rates because some students were absent on the day of the exams, and make-up exams were generally not offered. Completion rates for the interviews and reflection logs (not listed in the table) ranged from 86% to 100%.

Table 1. Numbers of participating students in semesters with multiple data streams

Class		Enrollment	Consent[a]	Diagnostic (preinstr.)	Diagnostic (postinstr.)	VNOS (pre)	VNOS (post)	Interviews and refl. logs
SEM-6 Phys 102	Traditional	49	43 (88%)	45	48	⋯	⋯	⋯
	PbP	57	55 (96%)	54	57	⋯	⋯	⋯
SEM-6 Astro	Traditional	51	46 (90%)	48	36	48	36	⋯
	PbP	93	89 (96%)	93	64	92	64	⋯
SEM-7 Phys 101	Traditional	62	59 (95%)	60	55	⋯	⋯	5
	PbP	68	66 (97%)	68	66	⋯	⋯	4
SEM-7 Astro	Traditional	56	44 (79%)	52	47	52	48	6
	PbP	85	70 (82%)	81	8	81	77	6
SEM-8 Phys 102	Traditional	51	41 (80%)	48	46	⋯	⋯	⋯
	PbP	68	61 (90%)	61	61	⋯	⋯	⋯
SEM-8 Astro	Traditional	96	69 (72%)	79	83	78	84	6
	PbP	92	81 (88%)	82	84	83	83	6

[a]Counts explicit "yes" consent forms only. Students who turned in a consent form but neglected to check either the "yes" or "no" box, and students who were absent on the day consent forms were distributed are automatically excluded.

3.2. Managing Privacy, Confidentiality, and Consent

A challenge doing education research is managing the instructor's need to evaluate students and the research team's need to have access to student performance on course assignments and examinations. The course instructor needs to know the student by name, and the research team may find it easier to work without knowing student names. There is no explicit federal requirement that student names must be held confidential from the research team (45 CFR 46.111 item (7) does state "When appropriate, there are adequate provisions to protect the privacy of subjects and to maintain the confidentiality of data."), but confidentiality from the research team may make some students feel more comfortable in participating in the study. It is also advantageous when possible to have the course instructor not know which students have elected to participate or not participate, and it is essential that the researchers conducting the analysis not see information from students who have elected to not participate unless seeing that information has a legitimate educational interest under Family Educational Rights and Privacy Act (FERPA). One method for managing these identity issues is to assign each student an anonymous (ideally random) project identification (ID) number. The project ID is needed to facilitate matching a particular student's performance on one instrument to other instruments (i.e., comparing pre- and post-test scores). A "neutral" party is identified to be the keeper of the master list that has the student names matched to the randomly assigned project ID number. An administrative assistant or one of the research team can be specifically assigned this role. They receive course performance updates from the course instructor, add the new information to a master spreadsheet or database, save these data, remove the student names, remove students who chose to not participate, save the spreadsheet as a deidentified version of the master spreadsheet, and send this version to the research team. By using a neutral party, the course instructor does not know who consented or not, and the research team only sees student performance using a project identifier from students who have consented to participate in the research.

3.3. Concept Diagnostic Exams

The pre-instruction diagnostic was given to all students, regardless of whether they elected to let their scores be used in the study or not. Their performance on the diagnostic was still pedagogically useful to the instructor, for assessing the incoming level of knowledge of the students. There was no course credit given for the score on this preinstruction diagnostic test, but rather students were informed that it was for diagnostic purposes only and they would not be penalized or rewarded in any way for their score. The diagnostic was given again during the last week of the semester. Depending on instructor preference, the diagnostic was sometimes given again for no course credit, sometimes for extra credit, and sometimes for credit often as part of the final exam. The rationale behind giving students credit for performance on this post-instruction diagnostic was to encourage them to take the test seriously during an otherwise stressful time of the semester, as students were sometimes observed to not put in a sincere effort when it was worth no credit. Hake (1998) compared student performance on the Force Concept Inventory given as post-test without credit and with credit and doubts that grade incentive has an impact on calculations of student learning gains. However, one drawback of giving credit for the post-instruction diagnostic is that this creates pressure for the students to do well, which in turn unfortunately provides an incentive for students to cheat. Indeed, this happened during one semester when students smuggled a copy of the test out at the beginning of the semester during the pre-instruction diagnostic and used it to study for the post-instruction diagnostic at the end. This unfortunate incident caused project staff to throw out all data related to the post-test and take more precautions to protect instrument integrity in subsequent semesters, such as numbering exam papers and counting them during administration of the exam. We cannot emphasize enough the need to protect the integrity of valid and reliable instruments used by many educators. We freely admit that we were caught off guard by how students breached what we thought were rigorous protocols for secure instrument implementation and now recognize how tightly these instruments must be controlled during use. This issue takes on such added importance due to how quickly a single breach can lead to a wide spread dissemination of the instrument on the internet thus compromising all current and future research projects using said instrument.

The instructor's involvement in the diagnostic test varied significantly in different semesters, from deliberately not seeing the test (to avoid the possibility of teaching to the test), to selecting the questions to be used on the diagnostic test from existing instruments, to writing questions and designing an instrument from scratch. The overall security of the instrument varied widely in different semesters as well, from numbering each exam paper and counting to be sure all were returned after administering the exam, to attempting to keep the exam secure but having students smuggle copies out of class to study, to giving the students questions from the exam to use as a study guide. Though instructor involvement and exam security varied between semesters, it was always held constant within a semester between the two sections of a class being compared. The role of the course instructor in a research project can range from allowing the research to be conducted without any other involvement to being the primary researcher. As the role of the course instructor in the research project increases, it is important to establish clear protocols for how the variety of assessment instruments will be used and analyzed. Using a valid and reliable instrument as intended by the designers allows for direct comparison to baseline data from other implementations of the instrument. If the instructor does not see the questions being asked, they will teach the course in a manner that is not biased by the instrument, but analysis of the results can be challenging without having some input from the instructor. It is important to identify the amount of overlap of course content and questions being asked by the researchers (Bardar *et al.* 2005). With a higher level of involvement in the research by the course instructor, the claims one makes about impacts on student learning must be carefully made.

The way in which implementation of a prescribed intervention varies from a planned course of instructional actions is a frequent problem in intervention studies. Such variation, which in methodological terms would be construed as a source of bias, is considered an inevitable feature of all except for the most mechanistic of interventions (e.g., class size, instructional time, and availability of an apparatus). As a methodological response to the problem of intervention variation, two strategies are recommended. The first strategy, following Penuel and Means (2004), is to develop and apply measures that assess implementation variation. The aim here is to measure departures from the prescribed intervention. The measurement strategy, which requires the collection of interview and/or observational data, involves the identification of critical sources of variation. Once those sources are identified, a standardized observational tool may be constructed and used across classrooms over time.

The second strategy is to incorporate findings on implementation variation into subsequent models of analysis that are used to compare the effects of one of set of instructional conditions against another. In a quantitative model, such as regression analysis, implementation variation typically is included as a moderator variable that reflects gradation in implementation fidelity. In those cases where departures from the planned implementation raise questions about the distinctions between an intervention and control condition, findings on implementation variation may be used to create entirely new, unplanned study conditions (e.g., intervention group one and intervention group two). In a study that compares a single intervention classroom against a single control classroom, the investigator should highlight the ways in which the intervention condition departed from the planned course of action. The important point here, from both a measurement perspective and an analytic perspective, is to recognize a distinction between the intervention as planned and the intervention as implemented.

The diagnostic exams given during the various semesters are listed in Table 2, and include the Force Concept Inventory (FCI) (Hestenes, Wells, and Swackhamer 1992), Conceptual Survey of Electricity and Magnetism (CSEM) (Maloney *et al.* 2001), Light and Spectroscopy Concept Inventory (LSCI) (Bardar *et al.* 2006), and Views on the Nature Of Science Questionnaire (VNOS) version B (Lederman *et al.* 2002). Sometimes two diagnostic exams were given to the same class to measure different concepts. Those instruments marked with an asterisk indicate that the exam was not given in its full published form, but rather a subset of the questions was used in order to make the exam shorter and appropriate for the course content. All diagnostics except the VNOS were multiple-choice and machine scored. The VNOS was a free written response to open-ended questions. The students' handwritten responses to VNOS questions were first transcribed, and then compared to a rubric to assess their demonstrated sophistication of understanding of the nature of science (NOS). This rubric was developed internally, but was based in part on the discussion of NOS understanding in Lederman *et al.* (2002). Modifying an instrument that was carefully created and tested by other researchers should not be done lightly. Any modification means that your student performance on the modified instrument cannot be directly compared to other student performance on the unmodified instrument. This is a real challenge facing education research in that many instruments exist that are created to assess specific student learning that may not match your course goals. Our decision to use only some of the questions on the VNOS version B was done to shorten the instrument in order to avoid overassessing our students. We also used a subset of the questions from the CSEM and the LSCI. Both of these instruments contained several questions that assessed material clearly not covered in our courses. These questions were removed to shorten the time spent completing these assessment instruments and to reduce student anxiety at being tested on topics never taught, but in hindsight it would have been very interesting to see how students performed on these omitted questions. Our working hypothesis is that pre-test and post-test gains on these particular questions would be zero, but we cannot confirm this due to removing these questions from our study. If you decide to modify existing instruments as we did, you will need to create your own baselines for those modified instruments against which to compare the results of your research efforts.

Two important methodologically related points should be made in connection with the alteration of instrumentation. The first point concerns the instrument validity and the second point concerns instrument sensitivity. For the first point, the alteration of instruments from their original form means that findings from prior validity studies are of marginal value. Instruments used in the present study will therefore be revalidated. We would recommend, however, that such revalidation work should be a standard practice even in those instances where instruments were not modified. Variations in study populations, instructional conditions, and test administration conditions all have the potential effect of weakening the inference validation findings from prior studies and the results of a newer study. The tactic we recommend here is grounded in the well-established approach referred to as multitrait-multimethod (MTMM) methodology of Campbell and Fiske (1959). What is in question in all validation studies is the degree to which measurements accurately reflect the set of abilities, skills, perceptions, or attitudes held by a particular study population. Within a MTMM approach, the investigator combines findings from a range of data sources (e.g., standardized measures, classroom tests, observations, and interviews). Areas in which strong corroboration or agreement is found are used to establish grounds for validity within a given study. Areas of divergence, or lack of agreement, reflect aspects where validity is weak or where the response patterns of subgroups (e.g., by gender by demographics, by major, and by years of study) within a study should be further examined. Such analysis might reveal that a given instrument is more valid for some study populations than it is for others.

The second point to make on the alteration of instruments concerns instrument sensitivity, meaning the degree to which a given instrument is able to accurately detect and the types of responses needed to assess the effectiveness of an intervention. While generic measures (e.g., NOS) provide a community of investigators with a common metric, the specifics of a given instructional situation may require modification of such measures. The modification of an instrument is well justified from both an instructional perspective (see Biggs and Tang 2007; Porter 2002) and from a measurement perspective (Lipsey 1990). It is important to recognize the importance of both the instruction and measurement because education studies are typically organized as field studies that are conducted in natural settings (in ordinary classroom settings) rather than laboratory studies.

Table 2. List of all courses and instruments in data set

Semester	Class	Diagnostic exam(s)	Other data
SEM-1	Physics 101: Newtonian Mechanics	FCI	
SEM-2	Physics 102: Electricity & Magnetism	CSEM[a]	
SEM-3	Physics 101: Newtonian Mechanics	FCI	
SEM-4	Physics 102: Electricity & Magnetism	CSEM[a]	
SEM-5	Physics 101: Newtonian Mechanics	FCI	
SEM-6	Physics 102: Electricity & Magnetism	CSEM[a]	
	Astronomy: Stars & Galaxies	LSCI[a]	
		VNOS[a]	
SEM-7			Interviews, refl. logs,
	Physics 101: Newtonian Mechanics	FCI	seating survey
	Astronomy: Solar System	Instructor's own	Interviews, refl. logs,
		VNOS[a]	seating survey
SEM-8	Physics 102: Electricity & Magnetism	CSEM[a]	
	Astronomy: Stars & Galaxies	LSCI[a]	Interviews, refl. logs,
		VNOS[a]	seating survey

[a]Used modified version of instrument.

3.4. Normal Course Data

For each semester, we also have access to information collected as part of the normal functioning of the course: final course grade, in-class exam performance, homework scores, and laboratory scores (if applicable). Although not collected explicitly for the study, these data offer insights into student effort and further probing of how student understanding of concepts developed over the semester, which is relevant to our research questions.

3.5. Seating Data

We also collected data about where students were sitting in the two classrooms in SEM-7 and SEM-8, in order to conduct an ancillary study about student seating preference and to investigate any possible correlation with course performance. To ascertain students' pre-existing seating preferences, a short seating survey was administered via clickers to inquire about their seat location in a previous large lecture class. Student seat location in the current class was also captured in both classrooms, either on the front of their exam papers on exam day, or as part of a short standalone survey which also asked students to list any possible reasons why they chose that particular seat location. The results of all these polls, plus an investigation of correlations with course grade, learning gains, and other student data for the two different types of classroom are described in Kregenow, Rogers, and Price (2010).

4. DATA COLLECTED FROM SMALL SAMPLE OF STUDENTS

4.1. Student Selection and Notification

A small sample of each class was chosen in both SEM-7 and SEM-8 for further study. Because the sample was relatively small (six students per study condition), the sampling strategy could not be regarded as representative sampling, in the strict technical sense. The method for selecting students was, however, designed to construct an interview sample to gain input from a range of students. Participants were therefore selected using the logic of stratified random sampling (Shadish, Cook, and Campbell 2002). This was done to minimize the biasing influences of volunteer effects and to assure that the sample was balanced according to gender and according to ability. Students were thus chosen from different quartiles (or terciles, or quintiles, etc.) of the class based on their pre-instruction diagnostic scores. In one case when a valid diagnostic instrument was not available, SAT scores were used to define the quartiles. Equal numbers of males and females were selected from each group.

The process of selecting these further participants is subject to rules dictated by the institution's HSRB and needs to be agreed upon and documented in advance. To comply with Ithaca College's HSRB, we were required to ask the whole class if they would like to volunteer for this further participation, which we did in two different ways. In both cases, the whole class was informed that a smaller sample of students was needed to collect further information, the extent of the further participation requested was described—in-person interviews and written reflection logs—and the class was told that compensation for the extra time commitment would be offered. Compensation was in the form of a gift certificate of up to $75 to the student bookstore, in an amount commensurate with the fraction of reflection logs and interviews completed. Students had an opportunity to ask questions about this further participation. The class was told that the selection would be random and selected students would be notified via e-mail. Project staff encouraged all students to please consider volunteering, as it would be helpful to the project.

In SEM-7, the entire classes were polled ahead of time via clicker about whether or not they would be interested in volunteering, and then the random selection was made *from the whole class*. If a selected student was found not to have either (a) consented to the study, (b) volunteered via clicker for further participation, or (c) taken the diagnostic pretest, their name was thrown out and a new random selection made. In this manner, six students were selected from each of the two Astronomy classes and four students from each Physics 101 class—plus one student who was handpicked as a case study for having a unique and near-perfect score on the diagnostic pre-test. Once randomly selected, students were asked to confirm via e-mail that they would like to participate. All students did.

A slight variant on this method was used in SEM-8. Instead of collecting student volunteer status via clicker response ahead of time, the random selection was made first. Again, if a selected student had not consented to the overall study or if they had not taken the diagnostic pre-test, their name was thrown out and a replacement selection made. Once randomly selected, students were again asked to confirm via e-mail if they would indeed like to participate as a volunteer. The majority of contacted students confirmed, though a small number (17%) declined or else failed to respond. In these cases, another student was randomly chosen in their place, maintaining gender balance in the sample.

We used different selection methods based on our desire to understand which method resulted in the maximum participation. In both cases, we achieved the desired student distribution, with the SEM-7 method requiring fewer contacts with students due to knowing which students were willing to participate before contacting them. This also required more work by the project staff by having to run the random selection process more frequently to find a student who consented to participate in the study and volunteered via clicker for further participation. We did notice that the SEM-8 method yielded a higher participation rate when students were contacted individually first. Early in our research design, we decided to divide students into top third, middle third, and bottom third and further divide the students by gender. We made these choices to best answer the specific questions posed in our National Science Foundation proposal. Your research may warrant using different divisions (e.g., if you are studying the impact your intervention has on women you may want to interview only women).

In all cases, before the first interview, students were asked to sign two additional consent forms for participating further in the study, one form for the interviews, and another for the reflection logs. The interview consent form included permission for being audio taped. Students were given extra copies of these forms for their records.

4.2. In-Person Interviews

Participants completed two 45 minute interviews, one each near the beginning and end of the semester. Interviews were schedule for 1 hour blocks to allow a little extra time for questions, signing consent forms, etc. Meeting times were arranged via e-mail, attempting to schedule all within a few day span to achieve some uniformity in sampling time. Interviewer stamina and sanity constraints limited the number of interviews per day to approximately nine or fewer. Due to student time conflicts, plus a few no shows, some creativity and flexibility in scheduling was helpful (e.g., evenings and weekends). Three of the project staff members had experience conducting student interviews, which was highly desirable. However, one of those was unavailable for the time commitment, and two were also the instructors of the courses, which disqualified them due to their influence over the students' grades. Therefore the interviews were conducted by the least intimidating of the remaining project staff (by consensus), who first completed an informal training on interviewing practices with the resident expert.

This "informal training" included running several example scenarios of possible vague subject responses and how to probe subjects to elaborate further. Four different general categories of probes were discussed and practiced: (1) semantic probes (e.g., "can you tell me what you mean by that?"), (2) illustration probes (e.g., "can you give me an example of that?"), (3) elaboration probes (e.g., "can you tell me more about that?"), and (4) justification probes (e.g., "can you tell me why that is?") (Constas 2009). The "trainee" and expert brainstormed examples of appropriate instances to use each and how to recognize them. The expert related numerous anecdotes from his own personal experience describing when he used each kind of probe and also related some pathological cases (e.g., hostile subjects, nervous subjects, persistent no shows) to be prepared for. After each round of interviewing, the trainee and expert met for debriefing, discussion of lessons learned, and how to be better prepared for the next round. This debriefing was invaluable, and often led to a refinement or clarification of the interview script questions. Rapidly accumulating experience over the course of the project caused the quality of the interviews (particularly the probes) to improve with each successive round.

To ensure confidentiality, all interviews were conducted in a small neutral conference room, located in a separate building far from the classrooms. Moreover, arbitrary identifiers were used on the recordings instead of student names: "Student A," "Student B," "Student C," and so on. Interviews were recorded (audio only) and later transcribed into simple text files. Recording was first attempted with an old analog recorder, but was later supplanted by the superior method of digital recording. Analog files were converted to digital *.MP3 format for archiving, using a free program called "Audacity" downloaded from the web. Digitally recorded files were captured in a proprietary *.WMA format, and were likewise converted to *.MP3 format for archiving. Digital transcription software was downloaded free from the internet ("Express Scribe," available at: http://www.nch.com.au/scribe/), which allowed playback control from the keyboard. Several levels of detail in voice transcription were considered, from recording only verbal utterances, to recording all audible sounds, to even encoding voice inflection and length of pauses. We opted for a medium level of transcription detail including all audible utterances, verbal or nonverbal, but ignoring inflection. Clear pauses were noted, but not their length. See an example in Appendix B. Our research uses a content analysis approach (Downe-Wamboldt 1992) where we created and defined categories of analysis and how these categories would be coded prior to conducting our interviews. We choose to focus our work on student understanding and perceptions and not on student confidence. If we had included confidence, it would have been important to transcribe the length of pauses and the inflections used by students when answering. When designing your research questions and methods, it is important to identify what data will be needed, how those data will be collected, and how the analysis will be conducted to answer your questions prior to the start of the research. It is helpful to do a "test run" to ensure that your choices are performing in the manner you expect.

Note that interviews are sometimes videotaped depending on the project needs, but HSRB approval is more rigorous. The additional level of detailed data provided by videotapes was not needed for this project. One further complication of using videotapes in our project would have been reduced participant privacy and confidentiality of the data. We used undergraduate student workers, in part, to help with the considerable task of transcribing more than 60 interviews over two semesters. Audio transcription was found to be rather anonymous, as the slow playback speed made voice identification unlikely. However, video files would not have preserved the anonymity of student participants as well. (If voice recognition by transcribers is a concern, voices can be made even more unrecognizable by voice distorting software. Our undergraduate transcriber helpers were instructed on the confidentiality of this project, however, and we found this additional level of security unnecessary due to the already present voice distortion.)

The topics covered in the interviews included student perceptions of the class—how it was conducted and how effective it was—as well as their views on the nature of science. The former helped us measure our progress toward project goals that were promised in our original funding proposal. The latter served as a complement to their written responses to the VNOS questionnaire that the whole class completed. NOS questions on the interviews were derived from a questionnaire developed by Liang *et al.* (2006). Questions about student perceptions of the class were developed internally to align with project goals. We sometimes included some supplemental questions on the interviews about the interview process itself, seeking student feedback about how the interview was conducted and how it could be improved. Response analysis provided additional feedback. This formative feedback in our project proved valuable, helping us simplify the wording of some unnecessarily complicated questions, and completely changing the approach to others. For example, when the interviewer asked the students about “problem solving,” we found that students were misinterpreting what we meant by the term. The interview script was revised the following semester based on this feedback to avoid such confusion.

The interviewer followed a set script, established ahead of time by the research team, to assure that the same basic questions were asked of all students. However, the interviewer was free to deviate from the script in order to ask probing questions—such as asking the student to give more detail, to define terms they used, or to illustrate with an example—whenever necessary to clarify a student’s answer. See an example excerpt from an interview script, followed by a corresponding transcript, in Appendix B.

4.3. Written Reflection Logs

Participants also filled out two or three written surveys or “reflection logs” during and shortly after the semester, answering questions about their feelings about the course and how it was conducted. These surveys were conducted entirely online at www.surveymonkey.com, and students were notified about each one via e-mail when their responses were requested. The reflection logs were designed to ask similar questions as the interviews and were timed to occur evenly spaced between the interviews, in order to achieve validation and time sampling on those topics, respectively. The last reflection log was given several weeks after the final exam, in order to give students a chance to look back and reflect on the course as a whole. See Appendix C for an example of an e-mail sent to students notifying them of an upcoming reflection log, as well as a sample screen capture of a typical reflection log web survey question.

4.4. Compensation

Compensation to students participating in these extra activities was given in the form of a gift certificate to the college-owned bookstore. This was preferable over giving cash and had the added bonus of the funds being directed back into the College due to the Ithaca College bookstore being college owned and run. (If your institution’s bookstore is not college-owned, other suitable college-owned venues might be available to allow you to support your school, such as gift certificates to a snack bar, tickets to sports games, etc.) The amount of the gift certificates given was commensurate with the fraction of reflection logs and interviews completed by each individual student, up to $60 if they completed everything asked, and up to $75 if they completed everything promptly. Be sure to check with your HSRB to see if there are restrictions on offering compensation to volunteers or in the amount you can offer. Many times the amount must be small enough that it does not coerce students into participating. If seeking external funding, be sure to budget for the cost of this volunteer compensation! In two semesters, we spent over $2300 on gift certificates to the 33 students interviewed. Also consider budgeting for transcription labor too, either by student workers or an external professional, as this is a very time-intensive process. The 61 interviews completed for this project took an estimated 288 hours to transcribe by project staff and undergraduate research assistants. Professional transcribers would be much faster, but also much more expensive to hire. It is important to note that all project staff and assistants were briefed on confidentiality and HSRB protocols. Training for HSRB protocols vary by institution so be sure to check with your HSRB guidelines. In general, as the privacy of the participant and the confidential nature of the data get harder to keep confidential the HSRB protocols must become more rigorous.

One last complication we encountered with delivery of the gift cards was that the last reflection log was due after the academic year was over, but several of the volunteers were graduating seniors who could not use the gift certificates after they left. We avoided this problem by awarding gift certificates to the seniors early, and extracting promises that they would complete all study requirements anyway. Three seniors followed through on their promises, but one did not, skipping out on the final reflection log.

5. DATA COLLECTED FROM INSTRUCTORS

We collected some data during the semester that was neither directly from nor about the students, but rather pertained to the instructors. First, the instructors documented their instructional approaches and materials used, particularly any differences between concurrent sections. Instructor implementation was also sampled using course observations. During three to four random, consecutive class days, one member of the project staff sat in on each class and recorded the type and duration of activities and methods used, using a fixed interval time sampling protocol. Every 5 minutes, the observer recorded the types of activities that took place in that time interval. This provides an objective measure of how class time was spent, supplemental to accounts (a) provided by instructors in their documented instructional approaches, and to those (b) recollected and described by students in interviews and reflection logs.

Another source of data available to the project is end-of-course questionnaires, routinely given at the end of every course at Ithaca College. Per College policy, these are always anonymous, and summaries are made available to instructors to see only after final grades are submitted. We obtained consent from instructors of participating classes to use these end-of-course questionnaires as an additional data stream for the project, whether the instructor was one of the researchers (in which case their consent was written into the project's original HSRB application) or unaffiliated with the project (in which case their consent was sought explicitly). Student responses on these questionnaires capture their perceptions of the class, the instructor, the class activities, the room, and even their understanding of the course material—all of which we queried explicitly as part of the study. Therefore this serves as an important calibration of study data, comparing the views of the whole class in a purely anonymous format with those of the few select individuals who were queried in the confidential, but not anonymous, interviews and reflection logs.

6. SUMMARY OF DATA STREAMS AND DATA MANAGEMENT

Overall, three primary lines of inquiry were pursued in this investigation: (1) student views of the course/instructor/room/class activities, (2) student understanding of the nature of science in general, and (3) student understanding of course concepts. Each avenue was probed by multiple instruments and using a combination of quantitative and qualitative data when possible in order to provide a rich data set. Several of the instruments were also given more than once, in order to be sensitive to change during the semester. Figure 1 summarizes the multiple data streams, what they were probing, how they overlapped, and when they were given. Bold text indicates they were given to all students. Plain text means they were given to subset of students only. Understanding of nature of science was probed by the VNOS questionnaire portion of the diagnostic exam (all students) and in the interviews (subset of students). Understanding of course material was probed by a pre- and post-instruction diagnostic assessment, course work, and performance on course exams (all students). Perceptions about class were queried through interviews and reflection logs (subset of students), and end-of-course questionnaires (all students).

As each data stream was collected, all were condensed and summarized into one large central database. The format of this database is a simple spreadsheet. The master spreadsheet contains all of the information about the students and the de-identified spreadsheet does not contain student names, college ID numbers, or information from students who did not consent to participate. Each student participating in the project has one entry, including every piece of information ever collected about them: ID number, (both student ID number, which is identifying information and therefore kept guarded, and a randomly generated six-digit project ID number, which is totally anonymous and therefore available to the whole team, including instructors), class, semester, whether they consent to participate in the study, prior GPA, SAT scores, major, class year, pre- and postinstruction diagnostic score, seat location, etc. This central organization of information is an excellent summary of our data set, and allows sorting the entire data set (to date, almost 1500 students!) by any field of interest to facilitate analysis centered around that aspect of the population. Quantitative data analysis is largely completed in Microsoft's Excel with some use of SPSS VERSION 7. Qualitative data analysis uses ATLAS.TI VERSION 6.0, a very powerful software package for qualitative data analysis of multiple data streams.

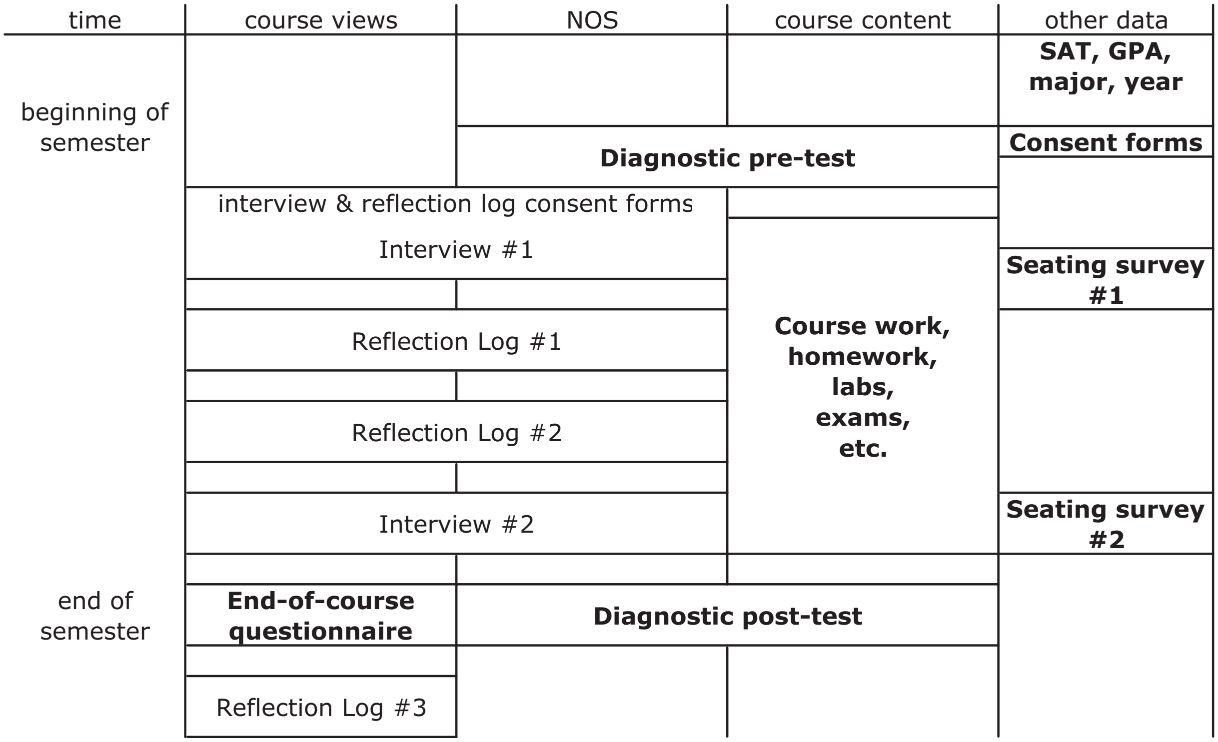


Figure 1. Timeline showing when and what data collection processes were done

7. FINAL THOUGHTS

Try to at least sketch out a data analysis plan before collecting any data in order to direct efforts most efficiently and to ensure you have all the data you need to answer your research questions. This is especially true in qualitative data collection, such as the interviews and reflection logs. An analysis plan can help determine what level of detail you need in student responses, and how to use follow-up interview questions to attain the desired level of detail. Better yet, do a trial run semester first if you can. Data analysis and results will be the topic of future papers from this group. Stay tuned!

Acknowledgments

This work was funded by NSF Grant No. 0536246 (PI: Michael Rogers) and Ithaca College, for which the authors are deeply grateful. We also thank an anonymous reviewer for providing detailed feedback that has led to a final paper that is more complete. We also thank the input and support of the other members of the Astronomy and Physics Education Research Group at Ithaca College: Sarah Burleson, Andrew Crouse, Rebecca Grollman, Luke Keller, Jill Neeley, and Matthew Price.

Appendix A

HSRB informed consent form given to all students. (Note: additional forms were required for interviews and reflection logs.) All students were given two copies: one to sign and return and one to keep.

APPENDIX B.3: INFORMED CONSENT FORM – Class Participation

Creating a Performance-based Physics Program for Introductory Physics and Astronomy Classes Using the SCALE-UP Model of Teaching Physics

1. Purpose of the Study: The primary goal of this project is to help out faculty deliver more effective physics instruction to Ithaca College science and non-science majors. Our objectives are to increase students' conceptual understanding of physics and astronomy, their problem solving abilities, and their understanding of the nature of physics and of science in general.

2. Benefits of the Study: Students will benefit through increased learning of the course material. The physics scientific community will benefit by having proven curriculum materials and methods available for implementation and adaptation.

3. What You Will Be Asked to Do: *Your participation will not involve any time beyond existing course requirements*. As a participant in this project you will complete the regularly assigned laboratory reports, activities, homework, quizzes, tutorials, and exams. You would complete these assignments even if this study were not being conducted. Your participation will make your performance available to the study, but not your name or any identifying data. Analysis of performance and conclusions reached by the study will be shared with other educators in the form of scientific journal articles.

4. Risks: Knowing that a study is being conducted in a class you are taking for credit can be stressful. We are aware of this and every effort has been made among instructors to ensure that all students are treated equally and fairly. If our experimental teaching methods result in significant grade differences among sections the lower averages will be curved to match the higher averages. This is standard practice in this multi-section course even when a study is not being conducted. **Your final grade in this course will not be affected by the study.**

5. If You Would Like More Information about the Study before, during, or after your participation you should contact the primary investigator, Ithaca College Physics professor Michael "Bodhi" Rogers either by stopping by his office (CNS265), calling him at 607-274-3963, or emailing him at mrogers@ithaca.edu. Bodhi is very accessible and will address all of your questions.

6. Withdraw from the Study: You are free to withdraw from this study—even if you previously expressed interest in participating—at any time with no impact on your final grade in this course. To withdraw from this study please contact Professor Rogers using the contact information in item 5. Your instructors will never know of your consent, non-consent, or withdrawal of consent from this study.

7. How the Data will be Maintained in Confidence: The data are your performance on normally scheduled assessments such as laboratory reports, homework, and exams. Your instructors keep a record of your grades that connect your name, student ID number, "clicker" number, and grade. Instructors use this information to track your performance in the course, assign your midterm, and assign your final grades. The primary investigator of this study (Professor Rogers) will only receive a list of "clicker" numbers and grades on assessment instruments. Your instructors will keep your name and performance confidential and Professor Rogers will never be able to connect your name to your performance. Professor Rogers will conduct all of the analysis involved in this study to determine what instructional approaches lead to enhanced student learning in physics.

I have read the above and I understand its contents.

[ ] I agree to participate in this study and I am 18 years of age or older

[ ] I do not wish to participate in this study and I am 18 years of age or older

[ ] I am younger than 18 years of age and cannot participate.

____

Print or Type Name

________________________________ ________________

Signature Date

A copy of this form should be given to each student. If more than one page will be used, each page before the signature page should have a line provided at the bottom for students to initial.

Appendix B—Interviews

Samples from interview script, used by the interviewer during each interview to assure uniformity in questions:

PART III. Student Understanding of Science and Scientific Inquiry Questionnaire

Please read EACH statement to the student (2x?), and ask the student to indicate the degree to which he/she agrees or disagrees with EACH from the options given: (SD= Strongly Disagree; D = Disagree More Than Agree; U = Uncertain or Not Sure; A = Agree More Than Disagree; SA = Strongly Agree).

26. Methodology of Scientific Investigation (introduce line of questioning)

A.	Scientists use different types of methods to conduct scientific investigations.
B.	Scientists follow the same step-by-step scientific method.
C.	When scientists use the scientific method correctly, their results are true and accurate.
D.	Experiments are not the only means used in the development of scientific knowledge.
With examples, explain whether scientists follow a single scientific method **OR** use different types of methods, whichever you think.	

SUMMARY QUESTIONS

27. So far, how does this class compare to other classes you've had in college?
    - Semantic probe, Illustration probe (if needed: Elaboration, Justification)

28. How does it compare to other SCIENCE classes you've had in college? (If haven't had any other science classes in college, compare to science classes in High School.)
    - Semantic probe, Illustration probe (if needed: Elaboration, Justification)

Excerpt from a representative interview transcript, as an example of the range of utterances captured. "J" is the interviewer, "S" the student:

J: When scientists use the scientific method correctly, their results are true and accurate.
S: Um, unsure.
J: Experiments are not the only means used in the development of scientific knowledge. [repeat]
S: Agree.
J: With examples, explain whether scientists follow a single scientific method, or use different types of methods, whichever you think.
S: Um, I think scientists definitely use different types of methods. Um, well there is like the standard method that we've been learning in school for you know however long, but I think um, you know in the real world scientists definitely use other, um, means of coming to conclusions and finding data. Um, so as an example… I don't know, I mean, I don't know if… I don't know, heh heh. Um, I guess like with the standard method, like you know you'd have your hypothesis, and then you'd come up with an experiment to determine that hypothesis, but um, I think it's definitely possible to just, um, make observations without a hypothesis or anything of that nature.
J: And, what is a hypothesis? To you?
S: Um, well, this probably isn't exactly right, but I've always thought of a hypothesis as is-you know, the if-then statement. You know, if something happens, then this must happen because, you know. It's just, um, a statement that you're gonna try to prove or disprove with your experiments.

Appendix C—Reflection Log: e-mail to students, followed by sample screen capture of web survey

Dear Student, this is the SECOND Teaching and Learning Reflections Log for Physics 175. The questions will be VERY SIMILAR to the last log—this is intentional. So yes, you may find it redundant. (I hope this is not too annoying!) The purpose is NOT to see how consistent you can be with your previous answers, but rather to see how your views may or may not have changed in the last few weeks. Please answer all of the questions thoughtfully, even if your views have not changed at all and even if you feel like you are saying the same thing as last time. That is totally fine. See the very bottom of this e-mail for a reminder of the purpose and confidentiality of these reflection logs (copied from last time).

DIRECTIONS: Record your thoughts for each of the questions on the website linked below. You do not need to complete the reflections log in one session: feel free to begin the reflections log and complete it later. You can modify any answers up until you submit the completed log. However, please note that if you leave the survey website before submitting your complete reflections log, your answers will NOT be saved. So if you want to complete the survey later, you must leave the survey window open. **This is not advised on a shared public computer, to maintain the confidentiality of our questions and your responses.** While you may work at your own pace, we request that you complete and submit the log no later than **April 2, 2009.** If you submit it on time, you will get a $5 bonus on your gift certificate at the end of the semester, as an extra thank-you for your punctuality. ☺ WHEN FINISHED, BE SURE TO CLICK "Done" TO SUBMIT!

We anticipate that answering all the questions should take roughly 20–30 min, though this will vary a little by student depending on how much you have to say.

Please complete the survey at this website:

http://www.surveymonkey.com/s.aspx?sm=lT8gbyJusu0OF2nY9zM7vA_3d_3d

Thank you! Do not hesitate to contact me if you have any questions, and thank you for your help.

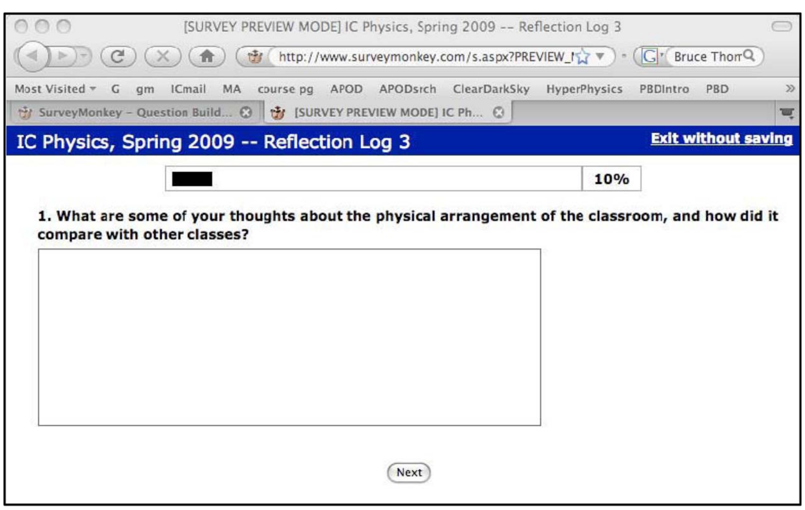
Best regards, Julia Kregenow

Read on for more info about study.

PURPOSE: In educational settings, reflections logs are commonly used as a tool to access and document the learning experiences of students. We appreciate the time and attention you dedicate to this activity as we seek to better understand the experiences of students in Physics 175.

CONFIDENTIALITY: You should know that we guarantee complete anonymity in connection with your reflection logs. None of your individual responses will be shared with the instructor in the class. The responses we receive from different students will be merged and considered as a group for each of the questions that we have posed.

We ask that you do not share this document or your responses with other students in the class. We make this request for two reasons: (1) to protect the promise of confidentiality, and (2) to prevent your reflections, as recorded here, from having systematic, undue influence on the perceptions and experiences of other students in the class. You are of course free to share your ideas about the class, but we ask that no documents associated with our investigation of the class be circulated.



An example question from the web survey.

References

Bardar, E. M., Prather, E. E., Brecher, K., and Slater, F. T. 2005, "The Need for a Light and Spectroscopy Concept Inventory for Assessing Innovations in Introductory Astronomy Survey Courses," *Astronomy Education Review*, 4, 20.

Bardar, E. M., Prather, E. E., Brecher, K., and Slater, F. T. 2006, "Development and Validation of the Light and Spectroscopy Concept Inventory," *Astronomy Education Review*, 5, 103.

Beichner, R. 2008, "The SCALE-UP Project: A Student-Centered, Active Learning Environment for Undergraduate Programs," an invited white paper for the National Academy of Sciences.

Biggs, J., and Tang, C. 2007, *Teaching for Quality Learning at University*, Maidenhead: McGraw Hill.

Brogt, E., Dokter, E., and Antonellis, J. 2007a, "Regulations and Ethical Considerations for Astronomy education Research," *Astronomy Education Review*, 6, 43.

Brogt, E., Dokter, E., Antonellis, J., and Buxner, S. 2007, "Regulations and Ethical Considerations for Astronomy education Research II: Resources and Worked Examples," *Astronomy Education Review*, 6, 99.

Brogt, E., Fostr, T., Dokter, E., Buxner, S., and Antonellis, J. 2008, "Regulations and Ethical Considerations for Astronomy education Research III: A Suggested Code of Ethics," *Astronomy Education Review*, 7, 57.

Campbell, D., and Fiske, D. 1959, "Convergent and Discriminant Validation by the Multi-Trait-Multi-Method Matrix," *Psychological Bulletin*, 56, 81.

Constas, M. 2009, *A Guide for the Standardization of Follow-Up Questions in Interview Studies*, Ithaca: Cornell University.

Downe-Wamboldt, B. 1992, "Content Analysis: Method, Applications, and Issues," *Health Care for Women International*, 13, 313.

Hake, R. 1998, "Interactive-Engagement Versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses," *American Journal of Physics*, 66, 64.

Hestenes, D., Wells, M., and Swackhamer, G. 1992, "Force Concept Inventory," *Physics Teacher*, 30, 141.

Kregenow, J., Rogers, M., and Price, M. 2010, "Is There a Back of the Room When the Teacher is in the Middle?." *Journal of College Science Teaching* (submitted).

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., and Schwartz, R. S. 2002, "Views of Nature of Science Questionnaire: Toward Valid and Meaningful Assessment of Learners' Conceptions of Nature of Science," *Journal of Research in Science Teaching*, 39, 497.

Liang, L. L., Chen, S., Chen, X., Kaya, O. N., Adams, A. D., Macklin, M., and Ebenezer, J. 2006, "Student Understanding of Science and Science Inquiry (SUSSI): Revision and Further Validation of an Assessment Instrument," Annual Conference of the National Association for Research in Science Teaching (NARST) (San Francisco).

Lipsey, M. 1990, *Design Sensitivity: Statistical Power For Experimental Research*, Thousand Oaks: Sage Publications.

Maloney, D. P., O'Kuma, T. L., Hieggelke, C. J., and Van Huevelen, A. 2001, "Surveying Students' Conceptual Knowledge of Electricity and Magnetism," *American Journal of Physics*, 69, S12.

Penuel, W., and Means, B. 2004, "Implementation Variation and Fidelity in An Inquiry Science Program: Analysis of GLOBE Data Reporting System," *Journal of Research in Science Teaching*, 41, 294.

Porter, A. 2002, "Measuring the Content Of Instruction: Uses in Research and Practice," *Educational Researcher*, 31, 3.

Shadish, W., Cook, T., and Campbell, D. 2002, *Experimental and Quasi-Experimental Designs for Generalized*

Causal Inference, New York: Houghton Mifflin.

U.S. Department of Health and Human Services 2005, Code of Federal Regulations, Title 45: Public Welfare, Part 46: Protection of Human Subjects, http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.htm (accessed December 3, 2009).

ÆR
010104-1–010104-16