A COMPARISION OF APPROACHES AND INSTRUMENTS FOR EVALUATING A GEOLOGICAL SCIENCES RESEARCH EXPERIENCES PROGRAM

Pamela C. Burnley Department of Geology, Georgia State University, Atlanta, GA 30303, Burnley@GSU.edu
William Evans Department of Communications, Georgia State University, Atlanta, GA 30303, jouwee@panther.GSU.edu
Olga S. Jarrett Department of Early Childhood Education, Georgia State University, Atlanta, GA 30303, ojarrett@mindspring.com

ABSTRACT

We are studying changes in knowledge of science and attitudes regarding science among participants in a summer Research Experiences for Undergraduates program run by the Atlanta Consortium for Research in the Earth Sciences. Existing survey instruments do not detect changes in our participants' attitudes over the course of our program and also fail to detect differences between our geoscience faculty and a group of college students with limited exposure to college level science. Therefore, we are developing a new survey instrument based on clusters of statements representing a variety of philosophical positions, from which respondents must pick one statement. We compare the distribution of the choices made by a group of respondents with the distribution of choices made by geoscience faculty. The first version of the instrument was able to differentiate between three different groups of students with different science backgrounds. Some of the statement clusters detected changes in our RUE program participants' attitudes over the course of the program. We believe that with further modification, an instrument can be developed that will detect changes induced by participation in a research experience. We have also studied the use of open-ended questions regarding the nature of science. Statistical analysis of responses to open-ended questions can also differentiate between college students with different science backgrounds and detect some changes over the course of our program.

Key words: Education - science

Introduction

Undergraduate research experiences are widely believed to be an important mechanism for recruiting undergraduates into science careers and for giving students an opportunity to test their interest in research (NSF, 1996; Mervis, 2001). Undergraduate research programs are also labor intensive (Manduca, 1997) and may not produce as many publishable results as graduate research programs (Spilich, 1997). Therefore, it has become increasingly important to evaluate whether undergraduate research programs are accomplishing what they set out to do. This study is part of a larger evaluation of a summer research experiences program for undergraduates and science teachers operated by the Atlanta Consortium for Research in the Earth Sciences (ACRES) at Georgia State University (GSU). ACRES was formed to extend and diversify the community of individuals participating in geoscience research. More specifically, ACRES works to: (1) promote undergraduate research; (2) introduce secondary science teachers to research in the geosciences; and (3) create strong collaborative ties between faculty at GSU and individuals and departments involved in teaching Earth Science at local colleges and universities that do not have a research emphasis.

The ACRES summer program is funded by the National Science Foundation (NSF) as a Research Experiences for Undergraduates (REU) site. The NSF's primary goal for REU programs is to attract talented undergraduates into careers in science and engineering. In its report, *Shaping the Future* (NSF, 1996), the NSF has also called for all undergraduates to have opportunities to learn science through inquiry. The NSF GEO Directorate has additional goals of introducing geoscience into the education mainstream and creating relationships between academic researchers and K-12 science educators (NSF, 1997). It is thought that this goal can also be accomplished by providing K-12 teachers opportunities for research experience and that teachers of secondary school science will share what they have learned with their students (NSF, 1996; NSF, 1997). As noted in *Shaping the Future*, teachers need to know "not just science facts, but just as important, the methods and process of research, what scientists and engineers do...." (NSF, 1996). Through its summer program, ACRES hopes to increase participants' understanding of scientific inquiry in general and of geoscience research in particular. Program evaluation is key to determining whether or not these goals are being met.

In order to evaluate their efficacy, undergraduate research programs primarily employ end-of-program satisfaction surveys, career tracking, compilation of publications and presentations, and retrospective comments from alumni (Kardash, 2000; for examples see Cole, 1995; Manduca, 1997; and Spilich, 1997). Due to the wide variations in the nature of student activities both between programs and within programs, much of what is known about the impact of undergraduate research programs is anecdotal (Kardash, 2000; Spilich et al., 1997; Mervis, 2001). However, because of both the large investment of faculty time required to create undergraduate research experiences and the need to be accountable to federal funding agencies, there is a move to develop additional means of assessment that more directly probe how students are changed by their involvement (Ahlm, 1997; Kardash, 2000). Blockus and others (1997) advocate measurements of student's perception of science as a career, personal epistemology, learning goals, and interest in science careers in order to disentangle the impact of the research experience from other influences on students career decisions. Kardash (2000) has developed a list of research skills for which changes in the student's confidence level can be measured. Ahlm (1997) suggests examining students' perceptions and attitudes regarding science.

We have used end-of-program satisfaction surveys, career tracking, compilation of publications and presentations, and pre- and post- questionnaires about interest in science and confidence in research skills in our attempt to evaluate the outcome of our research experiences program (Jarrett et al., 2000). We have also attempted to measure changes in attitude toward science and understanding of the nature of science. We choose these measures in part, because we explicitly deal with the nature of science in our program and because other studies suggest that this is an area in which we might see change. From our own early research experiences, we personally know that we learned much more than what was explicitly taught to us. A research experience involves an enculturation (Kardash, 2000) or socialization process (Hogan and Maglienti, 2001) where students pick up styles of speaking, the structure of explanation, and attitudes towards science from their mentors (Bleicher et al., 1996). Changes in understanding of the nature of science have been studied extensively in school children (e.g. Driver et al., 1996; Weinburgh, Hughes and Steele, 2000) and have also been used to evaluate science methods courses for school teachers (Palmquist and Finley, 1997; Bell et al., 1998; Akerson et al., 2000) and university content courses for science majors and nonmajors (Siebert & McIntosh, 2001). We decided to employ a variety of assessment instruments, including some we have developed ourselves, in hopes that we could systematically determine if, and how, the program affected the participants. We report here on these efforts.

SUMMER REU PROGRAM

Our program consists of an eight-week research experience. We report here on the first two summers (1999 and 2000) of the three year program. Students and science teachers come to GSU to work with eleven ACRES faculty members from GSU, Georgia Perimeter College, Columbus State University and Fort Valley State University. Participants and faculty work in teams of 2-3 faculty and 3-4 participants on four research projects that involve a variety of field- and lab-based activities. The participants study background material, collect and analyze data and present their results at the end of the summer. In addition to their research, the participants take part in a Philosophy of the Geosciences seminar, weekly group meetings, and science colloquia, as well as group social activities and recreational outings. The Philosophy of Geosciences seminar was designed to provide participants with a format for reflection on the nature of science and the process of doing research, both from the perspective of their own research experiences and in light of various claims about the nature of science made by philosophers of science.

PROGRAM EVALUATION

Program evaluation was conducted using a wide variety of techniques and had several goals. Some of the goals are related in a simple way to program implementation and the smooth functioning of the program. For this aspect of the program evaluation, we relied on interviews and questionnaires to query participants and faculty regarding their satisfaction both during and at the end of the summer. Our other evaluation goal was to determine how well we were meeting our broader program goals as well as the NSF's goals. Therefore, we have used both qualitative approaches and quantitative instruments to determine if participants changed their science interest level, attitude toward science, and understanding of the nature of science and career plans. The NSF strongly recommends quantitative evaluation techniques (Stevens et al., 1993). Accordingly, we focus much of our attention on quantitative instruments and analysis.

EVALUATION INSTRUMENTS

Existing Instruments - Knowledge of the nature of sciences has typically been treated, like knowledge of other subject matter, as something that can be measured by objective instruments (Hogan, 2000). Many tests and inventories have been developed that compare respondents understanding of the nature of science with the nature of science as it is understood by the instrument developers. We choose to use two of these existing instruments. We used the Realistic Understanding of the Nature of Scientific Knowledge (RUNSK) instrument (Weinburgh et al., 2000) and the Scientific Attitude Inventory II (SAI II) (Moore and Foy, 1997). The RUNSK, which was developed by Hughes (1998) for use with middle school children (Weinburgh et al., 2000) consists of 42 true/false statements that are based on eight propositions about scientific knowledge (Table 1). "Correct" responses are tabulated and respondents are assigned an overall score as well as subscores for each of the eight factors regarding scientific understanding that the RUNSK purports to assess. The SAI II is based on the SAI (Moore and Sutman, 1970) which was developed for use with middle through high school students, but also has been used at the university level (Siebert & McIntosh, 2001). The SAI was based on 6 pairs of positive and negative position statements (Table 2) from which a pool of attitude items was developed. The SAI was recently revised by Moore and Foy (1997) to reduce potential gender biases and to remove difficult language. The SAI II asks respondents to indicate (along a 5-point Likert scale) the extent to which they

RUNSK propositions	Faculty	ECE Grads	pre-REU	post-REU
	% correct			
Scientific knowledge is the product of human creativity.	86.7	86.2	75.4	89.2
Scientific knowledge is developmental.	94.4	98.7	97.4	96.2
Scientific knowledge is testable.	83.3	84.6	76.9	75.4
Scientific knowledge is unified.	97.2	94.9	98.7	84.6
Scientific knowledge is amoral	83.3	87.2	84.6	84.6
Scientific knowledge is parsimonious.	100.0	96.2	88.5	92.3
Scientific knowledge is relevant in many fields and/or endeavors.	95.8	97.1	98.1	94.2
Scientific knowledge reflects the contributions of many diverse individuals.	88.1	97.8	96.7	94.5
Total Score	91.1	92.8	89.5	88.9

Table 1. RUNSK propositions about science and scores for each group that filled in the RUSNK instruments. Scores are in terms of percent "correct" for each proposition.

agree or disagree with 40 statements regarding science. Subscores are calculated by summing ratings of related attitude items and subtracting summed ratings of opposing attitude items.

In the first year of the program we found that participant scores on the SAI II and RUNSK instruments were uniformly high and did not change over the course of the summer. There were several features of these instruments that appeared to be important in producing this result. The instrument developers worked hard to ensure that the items within their instruments would produce unanimous agreement among science educators and, for the RUNSK, among scientists as well (Moore and Sutman, 1970; Weinburgh et al., 2000). The result is that many of the items are very general, capturing science in its broadest form. They are too simple to address more subtle or controversial aspects of the nature of science. Another factor is that the instruments are scored by compiling various items together to generate subscores and a total score. In this process, responses regarding various ideas and values which are judged to be similar are added together to produce a single number. In deciding which ideas add or subtract from other ideas, the developers embed their ideological and philosophical views into the analysis. For example, Weinburgh et al. (2000) equate negative views about racial and gender diversity in science with the idea that one must be "really smart" to do science (intellectual diversity). The SAI II subtracts high ratings for recognition that science produces technology from the ratings of statements about science as an idea generating activity. As will become evident below, individuals with very different ideas can produce very similar scores on these instruments.

Statements About Science Instrument - Due to the limitations of the existing instruments, we undertook to develop our own instrument for measuring participants' understanding of the nature of science. Our goal was to include ideas about more controversial aspects of the nature of sciences and to create an assessment method that did not assume one correct view about the nature of science.

We designed the Statements About Science Instrument (SASI) to present a range of views without worrying about whether any of the choices accurately portrayed the "correct" answer, because there does not appear to be uniform agreement as to what the nature of science actually is. Philosophy of science presents us with widely divergent views regarding the nature of science. Although the extent of disagreement is debated (Elfin et al., 1999), modern philosophers hold somewhat different beliefs about of the nature of science than those held by science educators (Alters, 1997). The views of practicing scientists are different from those of both philosophers and science educators (Pomeroy, 1993). In addition, there is good evidence to suggest that geoscience is not identical in nature to other sciences such as physics and chemistry (Ault, 1998; Frodeman, 1995; Peters, 1996). Certainly, one can easily determine via a casual conversation with one's colleagues that geoscientists hold a range of opinions about the nature of geoscience. Therefore, the task of developing an instrument with a key of "correct" answers would be fraught with questions about both the validity of the questions and their answers.

Our approach to the "no right answer" problem is to stop thinking of the instrument as a test but instead think of it as an instrument like an oscilloscope that measures a signal. We do not expect an oscilloscope to render an exact replica of what it measures but rely instead on calibration with known input signals. In this case, our "signal" is the distribution of opinions about the nature of geoscience that a given population holds. We reason that since our goal is to make our program participants think and feel like geoscientists, we should compare the "signal" that we get from the participants to the "signal" we get from a population of geoscientists. When measuring a signal with an oscilloscope, it is critical to have the gain set correctly so that the signal is neither off scale nor its variations below the limits of resolution. To push the analogy further, the problem with the RUNSK and SAI II are that the "gain" is set too low; as we will discuss later, all the populations that we have surveyed saturate the instruments and we learn nothing of their differences.

The first version of the instrument consisted of 16 clusters of statements (Table 3) from each of which, respondents chose a single statement. Clusters 1-9 concerned the process of doing science, clusters 10-12 concerned the openness of the scientific community, and cluster 13 was about one's desire to work as a scientist. Three additional clusters dealt with images of scientists but included the option "other" with a blank to fill in; which we found made analysis difficult. These statement clusters are not analyzed here. We made an effort to write statement clusters that would not necessarily produce unanimous agreement among scientists. For sources of statements, we used our own understanding of modern philosophies of science including logical positivism and postmodernism, ideas from essays about science written by science education students and our personal thoughts about the nature of geoscience. For each statement, the number of respondents that choose that statement was converted into a percentage; which allowed us to compare the statement choices of different groups.

Open-ended Questions -To supplement data collected via closed-ended questionnaires (i.e., the SAI II and RUNSK instruments), participants were asked to respond to three open-ended questions:

1. What does it mean to study something scientifically?

2. What is a theory?

3. How can one distinguish good science from bad science?

Question 1 is borrowed from the National Science Board's Science and Engineering Indicators project, which has occasionally asked this question of a random sample of American adults (National Science Board 1993, 1998, 2000). Questions 2 and 3 were developed by us to further explore participants' understandings of science, especially in light of participants' experiences in the philosophy of science seminar conducted as part of the summer program.

The most recent National Science Board report on public understanding of science finds that only 21 percent of American adults can provide minimally adequate responses to the question "What does it mean to study something scientifically?" (National Science Board, 2000). Responses are considered minimally adequate if respondents touch on any one or more of the following three claims: (1) science involves theory building and theory testing, (2) science involves experimentation, or (3) science involves rigorous comparison. The National Science Board concedes that this would seem a rather generous standard for deeming a respondent to possess an adequate understanding of science. Again, however, most Americans do not offer any of these claims. Many Americans associate science with "precise measurement or with good or bad outcomes… but the work of scientists and the process of scientific inquiry are not understood" (National Science Board, 1998).

We asked participants to respond to the question "What is a theory?" to determine the extent to which participants believed that a theory is little more than a guess or an opinion. In everyday conversation, many people use the word "theory" in precisely this sense, even though most scientists would perhaps offer a more sophisticated and elaborate definition of theory, believing a theory to be a statement of general principles or laws thought to best account for a group of scientific facts or phenomena. A rather common rhetorical strategy of those who challenge mainstream science (e.g., creationists) is to suggest that scientific theories are little more than guesses or opinions. This strategy exploits the use of the word theory in everyday conversation and perhaps cultivates public misunderstanding of scientific processes.

We asked participants to tell us how they would distinguish good science from bad science because we wanted to determine the extent to which participants would refer to issues of scientific method and procedures as they told us how they would sort out good science from bad science. We also wanted to determine how frequently participants would refer to social and ethical factors when asked to provide criteria for distinguishing between good and bad science. Finally, we wanted to get a sense of whether or not participants could provide elaborate, sophisticated answers to this admittedly difficult question.

METHODS

ACRES REU participants in the first year consisted of nine undergraduates, finishing their freshmen to senior years, and three science teachers. In the second year, there were ten undergraduates, again finishing their freshmen to senior years, and three science teachers, one of which was returning for a second year.

On the first day of the program each year, participants filled out a questionnaire on their background experiences, their interest in science and research, and their confidence in field and lab techniques. They also responded to the questionnaires on the nature of science/attitude toward science; the SAI II, the RUNSK and, for summer 2000, the first version of the SASI. At the end of the program, participants again completed the above instru-

	Positive and Negative SAI II Position Statements	Faculty	ECE Grads	pre-REU	post-REU
		% correct			
1A 1B	The laws and/or theories of science are approximations of truth and are subject to change. The laws and/or theories of science represent unchangeable truths discovered through science.	82.8	78.3	81.0	83.6
2A 2B	Observation of natural phenomena and experimentation is the basis of scientific explanation. Science is limited in that it can only answer questions about natural phenomena and sometimes it is not able to do that. The basis of scientific explanation is in authority. Science deals with all problems and it can provide correct answers to all questions.	92.2	88.0	89.2	90.0
3A 3B	To operate in a scientific manner, one must display such traits as intellectual honesty, dependence upon objective observations of natural events, and willingness to alter one's position on the basis of sufficient evidence. To operate in a scientific manner one needs to know what other scientists think; one needs to know all the scientific truths and to be able to take the side of other scientists.	94.4	91.9	88.7	90.3
4A 4B	Science is an idea-generating activity. It is devoted to providing explanations of natural phenomena. Its value lies in its theoretical aspects. Science is a technology developing activity. It is devoted to serving mankind. Its value lies in its practical uses.	62.2	57.2	60.5	61.7
5A 5B	Progress in science required public support in this age of science; therefore, the public should be made aware of the nature of science and what it attempts to do. The public can understand science and ultimately benefits from scientific work. Public understanding of science would contribute nothing to the advancement of science or to human welfare; therefore, the public has no need to understand the nature of science. They cannot understand it and it does not affect them.	88.9	84.8	88.5	86.4
6A 6B	Being a scientist or working in a job requiring scientific knowledge and thinking would be a very interesting and rewarding life's work. I would like to do scientific work. Being a scientist or working in a job requiring scientific knowledge and thinking would be dull and uninteresting; it is only for highly intelligent people who are willing to spend most of their time at work. I would not like to do scientific work.	96.3	77.6	93.1	92.3
	Total	86.1	79.6	83.5	84.0

Table 2. Positive and negative position statements for the SAI II instrument showing subscores for each set of positions and the total score for each group that took the SAI II.

ments (with the exception of the background questionnaire). They were also asked open-ended questions on whether their career plans and, their perception of the nature of science had changed.

On the first and last day of the philosophy of geosciences seminar, the participants were asked to respond to our three open-ended questions: What does it mean to study something scientifically? What is a theory? How can one distinguish good science from bad science?

Comparative samples - In order to collect comparative data, we involved several other groups in our research. These additional groups had the following characteristics: (a) "Faculty" – nine geology faculty members (all but two with Ph.D. degrees) who teach geoscience and are involved in academic research. (b) "Freshmen" – 12 freshmen students taking an honors level, introduction to geoscience course. (c) "Science education (SE) graduate students" – 19 graduate students with a BS in science enrolled in a master's level teaching credential program taking an introductory course in geoscience for teachers; (d) "Early childhood education (ECE) graduate students" – 20 graduate students enrolled in a science methods course while working on their initial certification. (e) Undergraduate education majors – 5 undergraduate education majors taking an introductory course in geoscience for teachers. Both ECE graduate students and undergraduate education majors may have had little or no previous exposure to university level science courses. For analysis of the SASI, groups (d) and (e) were combined, and are referred to as "education majors" because of their similarity in background and the small number of undergraduate education majors.

We administered the SAI II and, RUNSK to 18 of the ECE graduate students at the start of the fall 2000 semester, and to the faculty during the summer. Twenty of the ECE graduate students were asked to respond to the open-ended questions at the start of the fall 2000 semester. In addition to the faculty, we administered the new instrument to the freshmen, SE graduate students, 18 of the ECE graduate students, and the undergraduate education majors, on the first week of classes of the fall 2000 semester.

Open-ended responses to questions regarding understandings of the nature of science were analyzed using WordStat, a software package for text analysis. Responses to the question "What does it mean to study something scientifically?" were coded according to the same criteria used by the National Science Board (1993, 1998, 2000) in their studies of the public understanding of science. A response was coded as adequate if it touched on the role of theory-building or testing, the use of experiments, or the application of rigorous comparison.

Responses to the question "What is a theory?" were coded in terms of whether or not the participant reported that theory was more than a guess or an opinion (e.g., "one person's ideas," " a guess that is not proven").

Responses to the question "How can one distinguish good science from bad science?" were coded in terms of whether or not the participant made reference to scientific method, the need for objectivity, or the application of peer review. Responses were also coded to assess whether or not the participant made any reference to social or ethical factors.

Two samples were analyzed: responses from 24 REU participants and 20 ECE graduate students. For the REU participants we combined samples from both summers, but used only the responses from the first year of the one participant that repeated the program. All responses were coded by one of the authors of this study (Evans) and one graduate student. Intercoder reliability, as determined via the application of Scott's pi, was greater than .89 for all items.

RESULTS

Existing Instruments - All of the sub scores as well as the total scores for the RUNSK were high for all groups that were tested in the second year (Table 1). There were no significant differences between the scores of faculty, ECE graduate students or REU participants. REU participants' total scores did not show statistically significant change from the beginning to the end of the program in either year.

The SAI II total scores were high for all groups tested (Table 2). With the exception of position statement 4, the subscores were also uniformly high and there were no significant differences between the subscores for each group with the exception of position statement 6 which revealed that ECE students were less interested in working as scientists (Table 2).

Statements about Science Instrument (SASI) - A difficulty we face in the analysis of the results from the SASI is the small number of subjects. The choices of statements within each cluster are categorical in nature, requiring nonparametric statistics. More subjects would be needed to meet the minimum expected value per cell required for chi-square tests of statistical significance. This difficulty can be overcome in the long term by combining data from different years, as we have done with the open-ended questions. We will also gather data from geoscience faculty at other institutions to develop better data for comparison.

For this pilot study, we have developed a simple method for comparing results from the different groups. For every cluster of statements we calculated the percentage of each group that chose each statement. We then compared these percentages to the percentage of faculty that chose the same statement. The percentages of the various groups which chose each statement is shown in Table 3. As was anticipated, many of the statement clusters in the SASI did not produce unanimous agreement among the faculty. The faculty was in agreement only on clusters 4, 6, 7, 9, and 13. There was disagreement, sometimes considerable, on statements in the other clusters (1, 2, 3, 5, 8, 10, 11 and 12). The other groups assessed with the SASI differed from the faculty to various degrees. The clusters showing the greatest differences were clusters 1, 5, 8-13

Only one cluster (7) produced almost complete agreement. An examination of the percentages for each group shows that the REU and Science Education graduate students answered more similarly to the faculty than did the education majors or freshmen. On some questions, the REU participants became more like the faculty as the summer progressed. For example, the percentage of REU participants who considered science a world view (cluster 1) and the percentage who considered science independent of culture (cluster 3) had become more like the faculty by the end of the summer. Over the course of the summer, REU participants also became more like the faculty on clusters 1,3,5,6, and 7. However they became slightly less like the faculty on clusters 2, 8, 11, and 12.

Open-ended Questions - At the beginning of the REU program, 62.5% of participants provided adequate answers to the question "What does it mean to study something scientifically?". At the end of the program, 88.9% percent of participants provided adequate answers, a statistically significant increase ($X^2 = 4.50$, df = 1, $p < .05$). No participant who provided an adequate answer at the beginning of the program failed to provide an adequate answer at the end of the program. At the beginning of the REU program, 12.5% of participants responded that a theory is no more than a guess or an opinion. At the end of the program, only one participant, or 5.6%, indicated that a theory was no more than a guess or an opinion. However, the decrease in the percentage of participants offering this answer at the end of the program is not statistically significant.

At the beginning of the REU program, 87.5 % percent of participants offered references to scientific method, objectivity, or peer review when asked "How can one distinguish between good science and bad science?" There were no changes evident at the end of the program. All participants who had mentioned scientific method, objectivity or peer review at the beginning of the program also mentioned one or more of these factors at the end of the program. And no participant who failed to mention one or more of these factors at the beginning of the program mentioned one or more of these factors at the end of the program

At the beginning of the REU program, 16.7% of participants mentioned at least one social or ethical factor when answering the question "How can one distinguish between good science and bad science?" For example, one participant wrote "good science… benefits the earth and its communities (people, animals, plants)." Another wrote, "bad science lacks…moral obligations (being harmful to others, etc.) in the way it is conducted." At the end of the program, 21.1% of participants mentioned at least one social or ethical factor, although this increase is not statistically significant.

As noted above, twenty ECE graduate students were also asked to respond to these three open-ended questions (these students completed the questionnaire only once, at the beginning of a science methods course, August, 2000). Responses from these students tend to differ markedly from the responses of the REU participants. Only 20% of the ECE graduate students provided adequate answers to the question "What does it mean to study something scientifically?" In comparison, 62.5% of REU participants provided adequate responses at the beginning of the REU program ($X^2 = 8.03$, df = 1, $p < .01$). When asked "What is a theory?", 42.1% of ECE graduate students indicated that a theory was little more than a guess or an opinion, whereas only 12.5% of REU participants shared this view at the beginning of the program ($X^2 = 4.88$, df =1 , $p <.05$).

When asked how they would distinguish good science from bad science, 42.1% of ECE graduate students mentioned scientific method, objectivity, or peer review. In contrast, 87.5% of REU participants mentioned one or more of these factors at the beginning of the program ($X^2 = 9.95$, df = 1, $p <.05$). No ECE graduate students mentioned social or ethical factors, while 16.7% of REU participants mentioned social or ethical factors at the beginning of the program (chi-square tests of significance would be suspect here since there were no respondents in one of the cells, a situation which violates the assumptions of chi-square testing).

It is worth noting that the question regarding how one can distinguish good and bad science seemed especially difficult for the ECE graduate students. Several students professed to be unable to answer the question, providing statements such as "[I'm] not sure what good science or bad science is." Others seemed to approach the question from a much different perspective than did the REU participants. For example, one student wrote: "Good science is hands on, fun, compelling and energizing." Another wrote: "I think good science is more inclusive than bad science."

DISCUSSION

Statements about Science Instrument (SASI) - Responses on the SASI suggest that it may be useful in distinguishing between groups having different levels and kinds of exposure to science. The Science Education graduate students with science degrees and, in some cases, professional experience were more like faculty in their choices than students with only high school level science course work. Education majors with little or no college level science coursework, were least like the faculty.

The small differences between the Science Education graduate students and the REU participants pre- and post- prompt us to continue to work on the sensitivity of the SASI by eliminating statements that drew unanimous responses and replacing them. We can keep clusters of statements where we did see changes in the pre- and post- test. For example, for clusters 1, 2 and 5, 25% of the REU participants changed their choice of statements in the same way (for cluster 2, an additional participant changed his/her mind in an opposing way), suggesting that these clusters more directly address changes in perspective brought about by the research experience. Statements that were not chosen by anyone can be replaced, and other clusters that don't exhibit disagreement between REU participants

	Table 3. Statementes are arranged in clusters. Percentages are listed in each column. Groups are: F - faculty, Ed - education majors, 1st - first year students, SE - SE grad students, pre - pre-REU, pos - post-REU	F	Ed	1st	SE	pre	pos
		%					
1	Science is a collection of true facts.	0.0	21.7	13.3	0.0	0.0	0.0
	Science is a procedure.	33.3	69.6	33.3	40.0	76.9	50.0
	Science is a world view.	66.7	8.7	53.3	60.0	23.1	50.0
2	Science assumes cause and effect.	70.0	52.4	53.3	60.0	69.2	50.0
	Science assumes nothing.	30.0	47.6	46.7	40.0	30.8	50.0
3	Science is independent of the culture.	37.5	47.8	46.7	40.0	46.2	33.3
	Science is affected by the language and culture it is conducted in.	62.5	52.2	46.7	55.0	53.8	66.7
	Scientific ideas are the stories that western scientists accept but have no special claim to describe reality over the stories of other cultures and people.	0.0	0.0	6.7	5.0	0.0	0.0
4	Scientific ideas are true.	0.0	13.6	13.3	5.0	0.0	0.0
	Scientific ideas are testable but can never be entirely proven.	100	86.4	86.7	95.0	100	100
	Scientific ideas are only relevant for western thinkers and only pertain to the western world view.	0.0	0.0	0.0	0.0	0.0	0.0
5	The scientific method is a loose procedure for making observations about the physical world.	33.3	4.3	20.0	23.8	15.4	41.7
	The scientific method requires that reproducible test be conducted.	66.7	95.7	80.0	76.2	84.6	58.3
6	Scientific thinking is logical and linear and does not involve intuition.	0.0	4.3	20.0	0.0	7.7	0.0
	Scientific thinking is a combination of logic and intuition.	100	95.7	80.0	95.0	92.3	100
	Scientific thinking is primarily intuitive.	0.0	0.0	0.0	5.0	0.0	0.0
7	Conclusions drawn by scientists are true.	0.0	0.0	0.0	0.0	0.0	0.0
	Conclusions drawn by most scientists are influenced by political pressures.	0.0	0.0	0.0	0.0	0.0	0.0
	Conclusions drawn by scientists are provisional and may be revised as more data become available.	100	100	100	100	92.3	100
	Conclusions drawn by scientists may not be relevant for other people.	0.0	0.0	0.0	0.0	7.7	0.0
8	Theories can be proven	44.4	50.0	53.3	20.0	30.8	16.7
	Theories can only be disproven.	55.6	50.0	46.7	80.0	69.2	83.3
9	Theories are effective models for describing reality.	100	75.0	80.0	85.0	100	100
	Theories are just stories.	0.0	5.0	6.7	5.0	0.0	0.0
	Theories are based only in facts.	0.0	20.0	13.3	10.0	0.0	0.0
10	The community of scientists is open and inclusive to anyone.	30.0	81.8	50.0	70.0	61.5	58.3
	The community of scientists requires successful completion of rigorous training to join	70.0	18.2	43.8	30.0	30.8	33.3
	The community of scientists is open to only specific demographic groups.	0.0	0.0	6.2	0.0	7.7	8.3
11	If I choose to, I could be come a scientists.	22.2	77.3	73.3	0.0	30.8	33.3
	If I tried to become a scientist, I might not succeed.	0.0	0.0	13.3	0.0	0.0	0.0
	I am a scientists.	77.8	22.7	13.3	100	69.2	66.7
12	Scientists would be completely accepting of me if I became a scientist.	14.3	54.5	20.0	25.0	38.5	50.0
	Scientists would mostly accept me as one of them, maybe a few would not.	85.7	40.9	66.7	65.0	61.5	50.0
	Scientists would never fully accept me into their ranks.	0.0	4.5	6.7	10.0	0.0	0.0
	The ranks of scientists will always be closed to me.	0.0	0.0	6.7	0.0	0.0	0.0
13	I think that science would be a good career path for me.	100	80.0	20.0	100	100	100
	I think science wouldn't be a good career path for me.	0.0	20.0	80.0	0.0	0.0	0.0

and faculty can be expanded. We have several sources from which to draw new statements and statement clusters. At the end of the summer we explicitly asked participants how they felt their understanding of science had changed. Virtually all of the participants responded that their ideas had not changed much. However, they also elaborated about the need be persistent in working through difficulties in the lab, the degree of integration between disciplines that they had previously perceived as separate, and the role of personal view point in making interpretations of data. We can also look at individual questions in the SAI II and RUNSK where there were shifts between the pre- and post- tests. For example, on the SAI II, six participants shifted toward agreement with the statements: "Scientists believe that nothing is known to be true for sure." and "Scientific questions are answered by observing things." On the RUNSK, three or more participants shifted toward agreement with the statements: "A good imagination is important when doing science.", "Using science to discover new things takes creativity." and "Scientific knowledge is a product of the human mind.". Therefore, we expect that clusters of statements discussing the role of imagination, creativity, and observation in doing science would be useful in capturing changes brought about by research experience

One of the weaknesses of our study is that we did not have access to a large number of faculty to serve as a comparison population. As we develop the next version, we will be looking for geoscience faculty that will be willing to fill out the instrument. We encourage you to visit our website [www.gpc.peachnet.edu/~acres/acres.html] to learn about how you can help in this development effort.

Open-ended Questions - Responses to our open-ended questions suggest that REU participants arrived on campus with an understanding of science that was more sophisticated than the understandings of the general public and of university students who are not science majors. REU participants were far more likely than the general public or our sample of education students to provide adequate answers to the question "What does it mean to study something scientifically?" Indeed, across the three questions we posed, REU respondents tended to provide more sophisticated responses than the education students who completed our questionnaire.

On one of our three questions -"What does it mean to study something scientifically?"- REU participants were more likely to offer adequate answers at the end of the program than at the beginning of the program. For all REU participants, those who provided adequate and sophisticated answers at the beginning of the program also provided adequate and sophisticated answers at the end of the program (that is, no participant seemed to become less knowledgeable or less sophisticated over the course of the program).

The use of open-ended questions to probe participants' understandings of scientific processes holds promise. REU participants are seemingly well-prepared to grapple with such questions, and their open-ended responses provide potentially rich data regarding their cognitive models of science. In the future, we intend to ask a few additional open-ended questions and to examine the resulting textual data for evidence of particular beliefs and the use of particular terms. In this way, we hope to more specifically link participants' responses to the content of the philosophy of science seminar (and the content of seminar readings). Hopefully, this will allow us to determine how and why the REU program may be cultivating specific views regarding science.

CONCLUSIONS

In summary, we have found that published attitude inventories developed to evaluate middle and high school students' understanding of the nature of science do not work well in assessing differences in populations of adults who have studied science at the college level. These instruments are similarly ill-equipped to help us understand if and how research experiences have affected REU program participants. In contrast, open-ended questions about the nature of science provide a potentially richer source of information. In addition, we have found evidence that a survey instrument designed to probe more subtle aspects of one's beliefs about science can be used to assess adults who have had a variety of different kinds of exposure to science. With further modification, the piloted instrument may be able to assess attitude and knowledge changes caused by participation in a scientific research experience. The new survey instrument that results from our development efforts should be helpful for other REU program administrators who wish to evaluate the impact of their program on participants' understandings of science.

ACKNOWLEDGEMENTS

This work was funded by a grant from the National Science Foundation, EAR98-20666. We wish to thank the participants in the ACRES REU program, the ACRES faculty as well as other faculty in the Geology Department at Georgia State University, and students in the authors' classes for being willing subjects of this research project. We would also like to thank Michelle Hall Wallace and an anonymous reviewer for many improvements to the manuscript, as well as Barbara Brocks, and Richard Bibbs for their assistance with data entry.

REFERENCES

Ahlm, K., 1997, A researcher's guide to key assessment principles and techniques: Council on Undergraduate Research Quarterly, v. 18, p. 70-72.

Alters, B.J., 1997, Whose nature of science?: Journal of Research in Science Teaching, v. 34, p. 39-55.

Akerson, V.L., Abd-El-Khalick, F., and Lederman, N.G., 2000, Influence of a reflective explicit activity-based approach on elementary teachers' conceptions of nature of science: Journal of Research in Science Teaching, v 37, p. 295-317.

Ault, C. R, Jr., 1998, Criteria of excellence for geological inquiry: The necessity of ambiguity: Journal of Research in Science Teaching, v. 35, p. 189-212.

Bell, R.L., Lederman, N.G., and Abd-El-Khalick, F., 1998, Implicit versus explicit nature of science instruction: an explicit response to Palquist and Finley: Journal of Research in Science Teaching, v. 35, p. 1057-1061.

Bleicher, R.E., 1996, High school students learning science in university research laboratories: Journal of Research in Science Teaching, v. 33, p. 1115-1133.

Blockus, L., Kardash, C., Blair, M., and Wallace, M., 1997, Undergraduate internship program evaluation: a comprehensive approach at a research university. Council on Undergraduate Research Quarterly, v. 18, p. 60-63.

Cole, F.L., 1995, Implementation and evaluation of an undergraduate research practicum. Journal of Professional Nursing, v. 11, p. 154-160.

Driver, R., Leach, J., Millar, R., & Scott, P., 1996, Young people's images of science: Philadelphia, Open University Press.

Elfin, J.T., Glennan, S. and Reisch, G., 1999, Nature of science: a perspective from the philosophy of science: Journal of Research in Science Teaching, v. 36, p. 107-116.

Frodeman, R., 1995, Geological Reasoning: Geology as an Interpretive and Historical Science: Geological Society of America Bulletin, v. 107, p. 960-968.

Hughes, M.A., 1998, Developing an instrument to measure the extent to which fifth-grade students' understanding of the nature of scientific knowledge matches that of a sample of scientists and science educators. Unpublished doctoral dissertation. Division of Education Studies, Emory University, Atlanta, GA.

Hogan, K., 2000, Exploring a process view about students' knowledge about the nature of science: Science Education, v. 84, p. 51-70.

Hogan, K. and Maglienti, M., 2001, Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions: Journal of Research in Science Teaching, v. 38, p. 663-687.

Jarrett, O. S., Burnley, P. C., Evans, W. 2000, Engagement in authentic geoscience research: Effects on undergraduates and secondary teachers. Paper presented at the national convention of NARST, New Orleans.

Kardash, C. M., 2000, Evaluation of and undergraduate research experience: perceptions of undergraduate interns and their faculty mentors: Journal of Educational Psychology, v. 92, p. 191-201.

Manduca, C. 1997, Broadly defined goals for undergraduate research projects: a basis for program evaluation: Council on Undergraduate Research Quarterly, v. 18, p. 64-69.

Mervis, J. 2001, Student Research: What is it good for?: Science, v. 293, p. 1614-1615.

Moore, R.W. and Foy, L.R.H., 1997, The scientific attitude inventory: a revision (SAI II): Journal of Research in Science Teaching, v. 34, p. 327-336.

National Science Board, 1993, Science & Engineer Indicators: Arlington, VA, National Science Foundation.

National Science Board, 1998, Science & Engineer Indicators: Arlington, VA: National Science Foundation.

National Science Board, 2000, Science & Engineer Indicators: Arlington, VA: National Science Foundation.

National Science Foundation, 1996, Shaping the Future: New expectations for Undergraduate Education in Science, Mathematics, Engineering and Technology. (NSF 96-139)

National Science Foundation, 1997, Geoscience Education: A Recommended Strategy. (NSF 97-171)

Palmquist, B.C. and Finley, F.N., 1997, Preservice teachers' views of the nature of science during a postbaccalaureate science teaching program: Journal of Research in Science Teaching, v. 34, p. 595-615.

Peters, E. K., 1996, No stone unturned: reasoning about rocks and fossils: New York , W.H. Freeman and Company.

Pomeroy, D.H., 1993, Implications of teachers' beliefs about the nature of science: Comparison of the beliefs of scientists, secondary science teachers, and elementary teachers: Science Education, v. 77, p. 261-278.

Siebert, E. D. & McIntosh, W. J. (Eds), 2001, College pathways to the science education standards: Arlington, VA, NSTA Press.

Spilich, G., 1997, Does undergraduate research pay off? Council on Undergraduate Research Quarterly, v. 18, p. 57-59.

Stevens, F., Lawrenz, F., Sharp, L., 1993, User-friendly handbook for project evaluation: science, mathematics, engineering and technology education, J. Frechtling ed. National Science Foundation (93-152).

Weinburgh, M., Hughes M. and Steele, D., 2000, Urban, African-American students' realistic understanding of the nature of scientific knowledge. Annual Conference of the National Association for Research in Science Teaching in New Orleans.