Learning Nature of Science Concepts through Online Peer Review of Student Research Reports[1]

by Nancy Trautmann[2] (Cornell University) and William Carlsen[3], Bugrahan Yalvac, Mustafa Cakir, and Christina Kohl (The Pennsylvania State University)

Abstract

Through structured science experiments followed by online peer review of their research reports, undergraduates can experience science as a collaborative process of investigation rather than a static collection of facts. In 2001, over 400 students from 11 colleges and universities participated in a pilot study of double-blind online peer review, and the study has been continued on a smaller scale since then. Results to date indicate that multi-university peer review of research results provides potential for student learning that reaches beyond what is possible within individual classrooms. Students reported gaining analytical skills and key understandings about the nature of science. According to participating faculty, students gained critical thinking and analysis skills and important understandings about the nature of scientific research. The significance of careful writing, data analysis, statistics, presentation, and review all were cited as aspects of science that students had learned through this peer review experience. In addition to reporting project-wide findings from the pilot study, this paper reports some results of two smaller-N "sub-studies." The first used interviews and other data sources to examine the impacts of the experiment/peer review process on students' understandings of the sociological nature of science, and revealed that although conventionally defined "nature of science" views appear largely unaffected, the experience greatly influences students' views of how scientific knowledge is established. The second sub-study used content analysis to explore the effects of the "quality" of peer reviews on students' writing, and suggests which aspects of the peer review process stimulate substantive changes in student-authored research.

Much remains to be learned about how these types of experiences can best be extended to a broader audience of undergraduates, meeting the needs of students from heterogeneous backgrounds and ability levels while keeping the experimental design process flexible and open-ended.

Introduction

Peer assessment, an integral component of professional scientific communities, is gaining increasing attention in higher education because of its potential for yielding gains in cognitive, social, affective, transferable skill, and systemic domains (Topping 1998). Reviewing each other's presentations and reports encourages students to think critically, questioning their own and each other's assumptions and interpretations and improving

---

[1] Paper presented at the annual meeting of the National Association for Research in Science Teaching, Philadelphia, PA, March 22-26, 2003.
[2] Center for the Environment, Rice Hall, Cornell University, Ithaca, NY 14853 (nmt2@cornell.edu)
[3] Pennsylvania State University, 173A Chambers Building, University Park, PA 16802 (wcarlsen@psu.edu). Correspondence may be addressed to either of the first two authors.

their writing (Gratz 1990; Liu et al. 2002; Topping 1998; Towns et al. 2000). Students may be motivated to do a more thorough job when they know that their work will be critiqued by fellow students (Hiltz and Wellman 1997; Sullivan et al. 1998; Towns et al. 2000). In addition, engaging in a process of peer review allows classroom science to become decentralized and more analogous to true scientific processes; as students help each other, they may recognize that valid information can come from sources other than just the teacher or the textbook (Billington 1997; Towns et al. 2000). And finally, peer review gives students the opportunity to experience first-hand one of the ways in which scientists work together rather than in isolation, creating a professional community (Koprowski 1997; Liu et al. 2002).

Successful use of student peer review requires overcoming several challenges. For example, students are more accustomed to competition than collaboration and consequently tend to worry about plagiarism when asked to share their work (Gratz 1990; Rushton et al. 1993). However, participating in peer assessment provides students an opportunity to learn the value of free exchange of ideas in scientific inquiry (Gratz 1990). Another reported challenge is that students' subject matter understanding may be insufficient to enable them to provide insightful comments on substance rather than just style (Bos et al. 1997). This can perhaps be mitigated by having students critique projects similar to ones they have carried out (Towns et al. 2000), or by assigning reviewers based on their individual expertise (Liu et al. 2002). In face-to-face peer review, it is difficult for students to base their critiques on the merits of the work rather than on personal relationships. This can be overcome through double-blind peer review, which reduces student hesitation to offer critical comments to classmates (Lightfoot 1998; MacLeod 1999).

In recent years, computer-mediated communication has begun to be used for multi-university collaborative efforts engaging undergraduates in web-based peer review of science essays. These projects have demonstrated the merits of online peer review for enhancing critical reading and writing skills of participating students (Henderson and Buising 2000; Robinson 2001; Towns et al. 2000). In the study reported here, we expanded the online peer review idea to encompass double-blind peer review of reports about students' own scientific experiments. In this project, peer assessment encompassed not only the quality of the scientific writing, but also the experimental design, data analysis, and validity of conclusions.

Objectives

Our overall goal is to enhance student understanding of the nature of science by engaging students in sociologically authentic scientific research and review of their research reports. Our hope is that participating students will learn to view science as an ongoing process of discovery in which scientists work individually and collaboratively to design and conduct research and to negotiate the meanings of their results. This goal is in keeping with current science education reform initiatives that focus on the desirability of students learning science as a mode of inquiry rather than a static collection of facts (American Association for Advancement of Science 1993; National Research Council 1996; Project Kaleidoscope 1999; Siebert and McIntosh 2001).

Our objectives are to address the following research questions:

1) What understandings about the nature of science do students develop through engaging in online peer review of their research reports?
2) What other educational outcomes result from their participation in this multi-university collaboration?

Project Description

In 2001 we used the NARST listserv to recruit faculty from 11 universities who were interested in engaging their students in simple toxicology experiments followed by online peer review of their research reports. Participating classes were evenly split between education and science departments: five were science education courses and six were introductory courses in life sciences or chemistry.

The 411 participating students designed and carried out experiments to test the toxicities of chemical compounds using lettuce seed bioassays. Protocols were provided for carrying out serial dilutions and lettuce seed bioassays (Trautmann et al. 2001), but participants were given leeway to choose what chemicals to test and in what concentrations. After completion of their experiments, students wrote and posted summary research reports on the project website (http://ei.ed.psu.edu). Over a 2-week period that had been scheduled in advance, each student anonymously provided numerical scores and guided critiques of two reports that had been posted by others. After receiving reviews, students had the opportunity to revise and then "publish" their reports on the project website. Future students will be able to use this library of published reports for reference when designing their own toxicology experiments.

In its current form, our web-based peer review system for students is double blind. Although anyone using the website can read the research reports, access to review comments is restricted to the authors of each report and to the faculty teaching the participating classes. This differs from many of the reported computer-mediated peer review systems for students, in which all reviews are visible to anyone who wishes to read them (MacLeod 1999; Towns et al. 2000), but it is analogous to peer review among professional scientists.

Ten out of eleven participating faculty used the peer reviews in grading. Two graded their student's critiques of other students' reports. Another included peer review scores and students' use of the reviews they had received as factors in calculating each student's grade. Other faculty gave credit or extra credit for completing peer reviews but did not grade the quality of the critiques per se.

At the conclusion of the project, all 11 faculty responded to an email questionnaire about their perceptions of the peer review experience. The 341 students who had given permission for their results to be used in research were asked to fill out a web-based questionnaire. This yielded 192 responses, a response rate of 55% (based on the number who gave informed consent at the start of the project, or 47% of all participants). The open-ended questionnaire items were summarized using constant comparative analysis and emergent coding (Glaser 1969). The codes were created during an initial reading of student responses and were designed to summarize all of the themes raised. Fifteen codes subsequently were dropped because their frequency of use was less than 2%. After development of the coding scheme, the lead author systematically coded the text. A second investigator independently coded all of the student responses, yielding an overall inter-rater agreement of 0.71.

The methodologies and findings of the two smaller-N sub-studies are reported later in this paper. The following sections describe findings based upon the entire corpus of 11 campuses and 192 student respondents.

I. General Findings

Implementation

When asked why they had chosen to have their classes participate in this project, faculty cited the opportunity for students to learn about the nature of science and to gain critical thinking, analytical, and science processing skills. "Nature of science" (NOS) for the purpose of the faculty questionnaire referred to the values, beliefs, and assumptions that underlie the creation of scientific knowledge, in contrast to other ways of knowing about the natural world (McComas et al. 1998). Scientists create knowledge by using observation and inference to obtain empirical evidence; that knowledge is then interpreted and negotiated through peer review using both logical reasoning and imagination. Seen through this lens, science is a creative human endeavor that is influenced by society and culture, resulting in knowledge that is both tentative and subjective. This lies in stark contrast to the traditional school portrayal of science as a set of facts and concepts to be memorized for the exam.

Some participating faculty used the online peer review project as a vehicle for explicit instruction about the nature of science, accompanied by related readings and classroom discussions. For example, in one course designed to help freshman marine biology majors understand the nature of science, class discussions focused on the online peer review experience in relation to the nature of science assertions discussed above. In a biology course for science teacher candidates, attention was given to the conventions of scientific evidence and explanation in relation to the students' own bioassay results. In an introductory-level chemistry lab course, class discussions focused on the roles and responsibilities of researcher and reviewer, and the importance of anonymous, non-personal feedback in establishing credibility for published scientific reports.

Other faculty used the online peer review project as an opportunity for students to participate in collaborative inquiry but did not explicitly address nature of science issues with their students. One professor of a science methods course (the second author) intentionally did not cover nature of science issues during the period in which his students were engaged in the peer review project. This was done in order to test to what degree the students would implicitly gain understandings through their peer review experiences. Pre- and post- interviews of eight students revealed few if any changes in students' general epistemological orientations about science (for example, they generally viewed science as a tentative, empirical, somewhat value-laden human enterprise both before and after the project). However, students showed substantial changes in their understandings about the sociological nature of science, especially concerning the role and nature of scientific argument, persuasion, and peer review in the establishment of scientific claims, knowledge, and facts (Yalvac and Carlsen 2002).

Outcomes

Faculty Perspectives

Participating faculty reported that students took seriously their roles in providing professional feedback to peers and learned the need for rigor in scientific writing, data analysis, presentation, and review. When asked about learning gains that might not have occurred through a more traditional science activity, most faculty mentioned aspects of professional science such as the collaborative process of construction and refinement of scientific knowledge, the subjective nature of evaluation and peer review, and the role of creativity in scientific research:

- "I was amazed to learn that students understood that creativity in science occurs in all processes; asking questions, methods, results and conclusions."
- "It was a powerful reinforcement for them of the rather subjective nature of evaluation and peer review."
- "It allowed the students to experience how knowledge is constructed and refined through experimentation and dialogue with peers."

The faculty rated the project highly in terms of its impact on students' skills in writing, critical thinking, and data analysis and presentation (Figure 1). The lowest score was given to the question of whether students had gained motivation or interest in science through participation in this project. This reflects the faculty's perception that many of the students already were highly motivated in science before the project began.

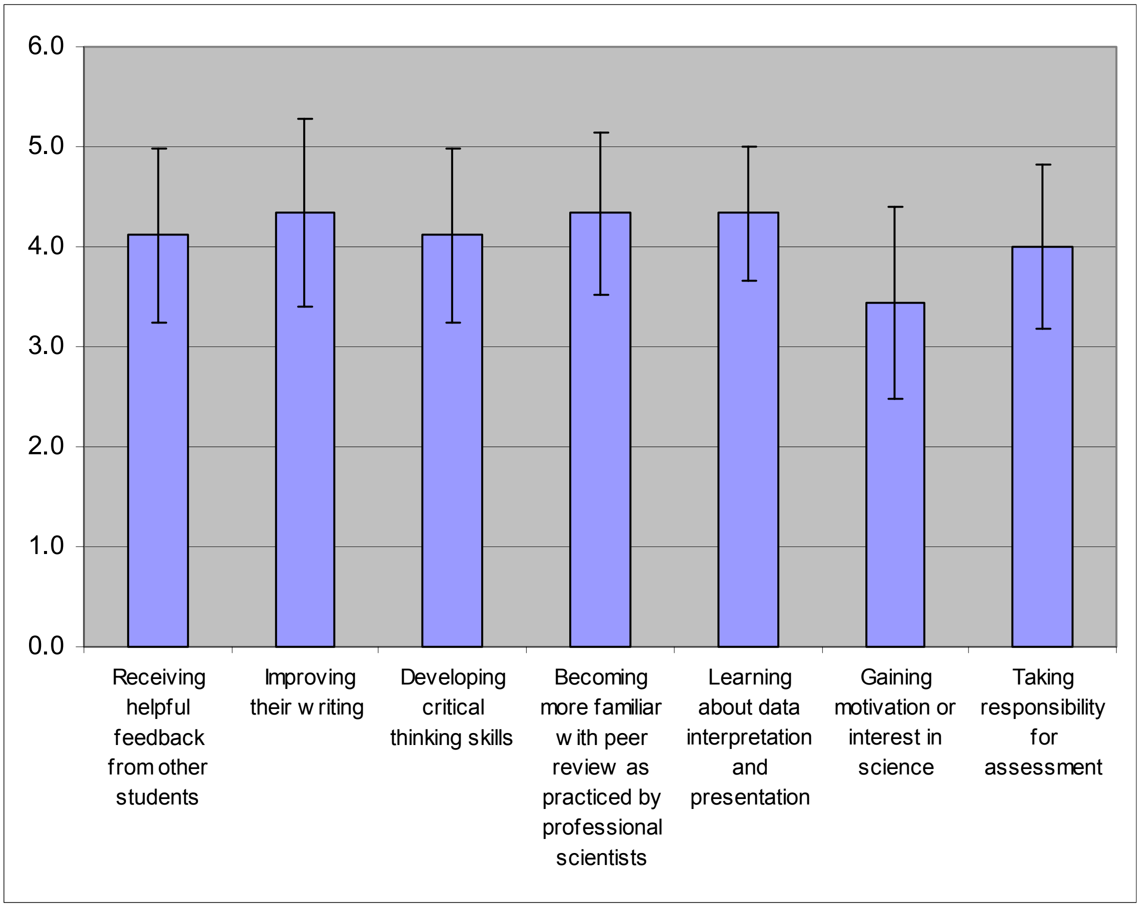

Figure 1. Faculty ratings of usefulness of online peer review to students

(5-pt. scale: 5=very useful, 1=not useful)

Student Perspectives

Student assessments of online peer review corresponded well with those of the faculty. The students rated the experience highly, both in terms of the reviews they received and the ones that they wrote (Table 1). In general, students perceived that they had gained more by writing critiques than by receiving them (items 1-5 compared with 6-8). They agreed that meaningful peer review is a reasonable expectation for students, and they indcated interest in incorporating it into their own practice if they become teachers (items 10-12).

Table 1. Student responses to web-based peer review experiences
(5-pt. scale: 5=strongly agree, 1=strongly disagree)

Questionnaire Item	Mean	S.D.
1. I felt qualified to provide meaningful peer review of other students' reports.	3.8	1.0
2. I believe that the peer reviews I wrote should be helpful to the students that received them.	4.0	0.7
3. I learned something by writing peer review comments.	3.9	0.8
4. Peer reviewing other students has helped me to think more critically.	4.1	0.9
5. Peer reviewing other students has helped me to improve my own scientific writing.	4.0	0.9
6. I received useful peer review comments about my own report.	3.5	1.1
7. The quantitative scores I received from peer reviewers were fair.	3.6	1.0
8. I changed my mind about something in my report because of comments I received through peer review.	3.0	1.3
9. It is easier to say what I really think when I don't have to sign my name or meet in person with the students who wrote the research reports.	3.7	1.2
10. I think that meaningful peer review is a reasonable expectation for college students.	4.2	0.8
11. I think that meaningful peer review would be a reasonable expectation for high school students.	3.9	0.9
12. If I taught science, I would like to use some type of formal student peer review.	4.0	0.9

The student questionnaire also included the following open-response question: “What do you believe students can learn about the nature of science by participating in projects like this?” Forty percent of the 174 responses discussed learning about the world of professional science, including the importance of communication among scientists and publication of results (Table 2). Not surprisingly, students reported that conducting peer reviews gave them insights into their own work and helped them to improve their writing, critical thinking, and critiquing skills. Perhaps most importantly, 11% reported discovering that science can be relevant or exciting, or that it can lead to unexpected results, and 8% percent mentioned discovering that science involves creativity or subjectivity (outcomes not readily obtained through traditional laboratory activities). Although the percentage of students who made such comments is small, it represents the number of students who independently brought up these topics, not the percentage who agreed with these outcomes on a standardized scale.

Table 2. Student responses to the question: What can students learn about the nature of science by participating in projects of this sort?

Topic	% of Students*
Learning about professional science, such as the need for communication and publication	40
Improving writing, critical thinking, or critiquing skills, gaining insights into own work	33
Conducting an experiment, learning scientific methods	32
Seeing the need for rigor, repetition, revision, or iteration in scientific research	16
Recognizing that collaboration and peer review are important aspects of science	12
Discovering that science can be relevant, exciting, or can lead to unexpected results	11
Realizing that science involves creativity or subjectivity	8

* *Percentage of students who mentioned one or more of the specified topics*

The final open-response question on the student questionnaire asked for any other comments or suggestions for improving the project. Of the 76 students who responded to this question, 45% stated that they had enjoyed the project and/or found it useful. Some made general comments such as:

- "I am thankful for the opportunity to participate."
- "I enjoyed this project very much. It helped me alot, and the peer review idea was a great idea."
- "I think it was really neat to be peer reviewed by many students across the world."
- "great!!!!!"

Other students were more specific about what they had found useful or enjoyable:

- "I thoroughly enjoyed participating in this project. I appreciated getting feedback from my peers. It was very helpful to me when the time came to revise and publish my report. The anonymity enabled me to give honest suggestions that I otherwise would not have made for fear of offending someone or hurting their feelings."
- "This was great to allow students to be part of research. Many people think that it is easy and that the answers come within minutes of testing. However, with exposing them to such experience, one can begin to imagine that research is never done and questions and variables are always possible."

Students also used this open-ended question to offer suggestions about how to improve the peer review system. These suggestions included revising the questions on the online form, and providing a tutorial to introduce students to peer review techniques before they begin the process. Several students mentioned being concerned about the numerical scores they had received from peer reviewers and indicated that it would be useful to have an online forum for discussion or clarification of comments or scores at the end of the peer review process.

Open-Ended Experimental Design

At the end of the project, a common suggestion made by both faculty and students was for the experiments to be more in-depth and open-ended rather than following well-defined research protocols that we had provided. Having everyone follow the same protocols simplified the logistics of coordinating multi-university peer review and ensured that the students had similar research experiences but left little room for individual creativity in experimental design. In future uses of the system, rather than suggesting that everyone perform dose/response experiments using lettuce seeds as the bioassay organisms, we plan to expand the range of choices and leave more design decisions up to the participating classes. For example, after learning the basic concepts underlying dose/response experiments, students will be able to select the type of bioassay organisms to use and the conditions to be tested. They might choose to use several types of organism or to use environmental samples in place of known chemical solutions. Or students could begin by reading bioassay reports that have already been published on our website, then develop their own research question based on the results of previous students' work.

Heterogeneous Grouping

In setting up this initial peer review project, we decided to include any classes that were interested in participating. This resulted in participation by a diverse mixture of science and science education students. For the most part, this diversity appears to have worked well, but one professor commented that most of his students did not find the peer review feedback to be as thorough or informative as they had hoped. Student comments on the final questionnaire also indicate a few instances in which mismatches in student ability level occurred:

> - "The peer reviews were very helpful and allowed me to step back and evaluate my own work through someone else's eyes. I did feel intimidated though, because when I went to evaluate my second report, I felt that I was not qualified to give the constructive critisism (sic) that was necessary for that report."
>
> - "One problem that I found was that the students participating in the project were at all different levels of knowledge of how to write scientific papers. Sometimes people did not give very helpful reviews if they reviewed someone who has a better understanding of the process because they did not know what they were doing themselves. If someone is better at writing papers, or at least on the same level, then it is extremely helpful."
>
> - "I think that peer review can be beneficial in many instances, present case included. However, when I actually read the reviews written on my report, two of the three reviewers questioned my conclusion that an improvement on the

> experiment would be a "larger sample size." They both commented, "I think the lettuce seed is a big enough organism to study," when clearly that is not what "larger sample size" means. It made me question the intelligence and qualifications of those reviewing me and subtracted from what I was able to gain from it. I suppose there really isn't any way around things like this happening. But I think it suggests a need to slightly "tweak" the reviewing system."

Although complaints such as these are serious, they also lead to opportunities for class discussions about peer review in professional scientific communities. Is peer review among professional scientists always rigorous, constructive, and accurate? Of course not, but that is the ideal, and problems that do occur usually are moderated by having multiple reviewers and journal editors or grant officers who mediate the peer review process. Should we modify the online peer review process to better accommodate various student ability levels, or provide some sort of calibration for interpretation of peer review results? Although too cumbersome to be feasible in this initial project, these considerations will need to be addressed in further development of the system.

General Conclusions

Results to date suggest that multi-university peer review of research results provides great potential for student learning, beyond what is possible within individual classrooms. Participating students embraced the opportunity to take responsibility for their own learning, and they gained motivation and self-respect when given the opportunity to present their findings and engage in double-blind peer review. At the conclusion of this collaborative process, students used words such as "creative," "exciting," and "dynamic" to describe their new perceptions of science. Although there is potential for abuse when students anonymously critique each other's work, we found that the students acted responsibly and benefited from their unmediated online interactions. Not every review was insightful, but none were offensive or otherwise inappropriate. Much remains to be learned about how this type of experience can best be extended to a broader audience of undergraduates, meeting the needs of students from heterogeneous backgrounds and ability levels while keeping the experimental design process flexible and open-ended.

II. Sub-Study One: Participant Perspectives on Knowledge Construction

A goal of the College Peer Review project has been to support studies of science-related educational processes that would be difficult to undertake in a one-campus project. The two "sub-studies" described here were both conducted at Penn State University, utilizing the multi-campus nature of the project in different ways.

Sub-Study One was a qualitative study to explore students' understandings of the Nature of Science (NOS) and the establishment of scientific knowledge, as well as their experiences with the Peer Review System. We collected most data through interviews, supplemented with other written tasks and observations in the laboratory when students were conducting their experiments. The primary population of that qualitative study was 21 Penn State (University Park) students enrolled in a science method course (SCIED 411) in the Fall semester, 2001.

All of the students enrolled in SCIED 411 completed an online essay-format Philosophy of Science questionnaire at the beginning of the semester. We categorized students' philosophies of science based on their responses. We observed 6 major categories of philosophies of science that students crafted in their responses to the questionnaire. These categories can be summarized as science (1) is the study of nature, (2) is structured, (3) is progressive, (4) is objective, (5) requires experimenting, and (6) is value free. Using these categories, we purposefully selected nine participants utilizing maximum variation sampling strategy (Patton 1990). The nine selected students' philosophies of science represented the spectrum of the philosophies we observed in our analyses. For example, one student stated that properly conducted science is objective and value-free, but that poorly done science deviates from these standards; another student mentioned that science could never be objective and is always value laden—these represent different philosophical stances. We used maximum variation sampling because we wondered how pre-conceptions of the NOS would be influenced by participation with peer review in the project. Eight of the nine selected students agreed to participate in the study.

The researchers designed two semi-structured interview protocols. First, a semi-structured interview protocol was designed to explore students' conceptions of the NOS and the establishment of scientific knowledge. Related literature (e.g. Bell, Lederman & Abd-El-Khalick, 2000; Schwartz, Lederman, and Crawford, 2000) was utilized in the designing this protocol, which we named the "NOS interview protocol." We considered the following characteristics central to this view of NOS: that science is tentative, empirical, value-laden, subjective, and a human endeavor. The other features we attempted to illuminate through interview items were students' understandings of how scientific knowledge is established and which steps are taken in generating scientific knowledge. The final version of this semi-structured NOS interview protocol is included in Appendix A. A second semi-structured interview protocol was intended to explore students' experiences with the Peer Review System. This protocol, the "CPR interview protocol," is included in Appendix B.

The Philosophy of Science Questionnaire was administrated twice, once at the beginning of the semester (Pre-Administration) and once at the end of the semester (Post-Administration). The third author (Yalvac) interviewed each of the eight participants before the activities of the CPR Project were implemented (Pre-Interview). In the pre-interviews only the semi-structured NOS interview protocol guided the conversations. After the CPR Project was completed, the participants were interviewed a second time (Post-Interview). In the post-interviews, in addition to the semi-structured NOS interview protocol, the semi-structured CPR interview protocol was used. The instrumentation sequence of the study is represented in Figure X.

Philosophy of Science Questionnaire (Pre-Administration)	College Peer Review Project	Philosophy of Science Questionnaire (Post-Administration)

Semi-structured Nature of Science (NOS) Interview Protocol (Pre-Interview)		Semi-structured Nature of Science (NOS) Interview Protocol & Semi-structured College Peer Review (CPR) Interview Protocol (Post-Interview)

Figure 2. The Instrumentation Sequence of Sub-Study One

Researchers met several times during Sub-Study One to discuss the coding processes and representations of the findings. After the verbatim data of the pre- and post-interviews were completed, researchers categorized the data according to the emerging codes. The data collected from the pre and post administrations of the Philosophy of Science questionnaire were also used to interpret and corroborate interview findings.

Sub-Study One Findings

In this section we discuss (1) differences observed in students' "before" and "after" conceptions of the NOS and the establishment of scientific knowledge, and (2) students' experiences with the Peer Review system.

Students' participation in the CPR Project did not dramatically alter their conceptions of the aforementioned features of the NOS. For example, before the project, all the participants described science primarily as "the study of nature." After the project, seven of the eight students still described science in this fashion. One of them responded differently, discussing science as "a way of living." All eight of the interviewees implied that science was "not completely objective" before the project, and all held very similar beliefs after the project ended. At the beginning of the project, seven of the participants mentioned that scientific facts were "subject to change;" all eight reported this in the final interview. Before the project, all eight participants responded that science involves "human values" and they reported similarly at the end. It should be noted, however, that all of the students had, prior to this investigation, completed at least one survey course in the Department of Science, Technology, and Society (a program requirement); consequently their initial perspectives on NOS may differ from those of students in other teacher education programs.

In contrast, students' understandings of the sociological nature of science (how scientific knowledge is established and accepted by the scientific community) changed dramatically during the study. Before the project, interviewees reported only "experimenting" and "inductive reasoning" as parameters that play important roles in the establishment of scientific knowledge. Even though we specifically asked them (through indirect emerging questions) to talk us about the scientific community and the establishment of scientific knowledge, students consistently responded that the process of generating scientific knowledge was to experiment a scientific claim "again and again." After the project, when we asked them to talk about the process of generation of scientific knowledge, they described in detail the role of peer review in the establishment of scientific knowledge, in addition to "experimenting" and "inductive reasoning." Participants explicitly stated that scientific experiments were needed for scientific knowledge to be established, but not solely adequate for the scientific community to accept them completely. Scientists' work must succeed through a peer review system involving other scientists as a means of verifying conclusions.[4]

When we asked about their experiences with the CPR project, participants responded that they enjoyed the project because of its peer review, original research, and large scale characteristics. Most of the participants indicated that they liked the idea they were investigating something where conclusion was unknown. The following interview excerpt illustrates that enthusiasm.

> *"I liked that we were doing an investigation that we didn't necessarily have the book that says this is the answer that you should get, so we were sort of figuring out by ourselves the toxicity of whatever the substance we were studying is…" (Carla)*

Participants mentioned that they didn't like the delays in posting their reports. Grammar errors in others' reports and reviews were another concern. Some suggestions involved using different organisms:

> *"… if they had little more freedom to make up their own concentrations to experiment a little or more a couple different things maybe an animal group in it, plant group in it, try bacteria or those types of things…" (Tim)*

The participants found the peer review and original research aspects of the project to be unique in their educational experience. The large scale of the review system was one of the original components of this project students mentioned.

> *"It [the College Peer Review project] was unique in the sense that a large number of people were doing that, it wasn't just a twenty five people in our class, there was other schools, you know… it actually made you think about people in other schools, you know, forty one thousand students at Penn State and it's hard to think that there are other colleges outside of the Penn State, that was neat, that way it was different…using the internet, most science experiments you don't really use the internet you don't publish your report, you follow the procedure they gave you, you do it, and you write up the report and hand in a teacher to grade." (Michael)*

Participants responded that their conceptions of the NOS did not change after the College Peer Review System. Six of the participants were optimistic about using a similar system in their future teaching. Their main concern was in grading:

> *"I probably consider it [using peer review system in her future teaching] I will definitely look for the experiments where it would be possible and fit into nicely I don't know how fair it would be to make it part of the grade you know if they do coming up with answers to questions about the people's paper like I can tell I read someone else's paper and they wouldn't get the points on that but I don't think that I would grade each student depending on how other students thought about their report because I don't think other*

[4] More information and excerpted student responses are provided in Yalvac and Carlsen (2002).

students will really have knowledge to say whether it is good or bad as far as you know making a great deal out of it..” (Carla)

Six of the participants agreed with the peer reviews they received and two of them didn’t find them useful. Four of them reported that they learned something new from the reviews and other four of them did not.

The primary conclusion we reached from Sub-Study One was that students’ understanding of the establishment of scientific knowledge was influenced concerning the sociological nature of science knowledge production. For other NOS conceptions, the CPR project did not appear to influence students. However, it should be noted that before the College Peer Review project was conducted, these students’ understandings of the tentative, empirical, subjective, value-laden, and human-endeavor characteristics of science were relatively consistent with the expectations of the science education documents (American Advancement of Association of Science 1989, 1993, 1997, 2001; National Research Council 1996) and researchers (e.g., Bell, Lederman & Abd-El-Khalick 2000; Lederman 1992; Matthews 1996; Smith & Scharmann 1999; Schwartz, Lederman & Crawford 2000).

Sub-Study One has evolved into a dissertation project for the third author, who is currently analyzing data from a more recent cohort of students who participated in a subsequent round of experimentation and peer review.

III. Sub-Study Two

Sub-Study Two, which is still in progress, is an exploratory analysis of the effects of peer review on students’ work, using content analysis and statistical analysis of several types of data. The general question addressed in Sub-Study Two is whether and how experimenters’ written reports are modified as a function of the peer reviews they receive.

Bear in mind that in this project student experimenters “submit” an initial laboratory report, later receive two anonymous peer reviews (these reviews include both quantitative evaluation scores and essay-type comments), make any revisions that they wish, then “publish” their revised report either anonymously or attributed with the author’s name. Only at the final stage is the report accessible to the public via the web, and only at the final stage is the author’s name made available.[5]

Within the general question of how peer reviews affect students “published” outcomes are a number of subsidiary questions:

1. What makes a <u>good quality</u> (that is, substantive and helpful, as opposed to positive and complimentary) review?
2. Do <u>good quality</u> reviews stimulate greater changes in students’ reports than <u>poor quality</u> reviews?

[5] The option to “publish” anonymously was a necessary condition for Human Subjects approval, although it does create a rather interesting nature-of-science anomaly.

3. Do negative reviews stimulate greater changes than positive reviews?
4. What exactly do we mean by changes in students' reports? How can this outcome be operationalized in a way that is useful?
5. How important is consistency between reviewers in stimulating changes in students' reports? How does quantitative consistency (that is, agreement on reviewer-assigned quantitative evaluation measures) compare to qualitative consistency (that is, agreement on other, researcher-assigned, measures of the content of reviews)?
6. What other features of initial reports, reviews, and subjects help to effect changes in students' writing?

Four issues prompted us to tackle these questions using a small-N exploratory study. First, our web interface and database systems were not designed to save multiple versions of students' reports: when a student revises his or her paper, the original version is overwritten.[6] Consequently, to get "before review" and "after review" versions of student reports, we had to work through tape backups of the databases: a laborious process that we undertook until we had approximately two dozen sets of student data. We judged these sufficiently varied for exploratory analysis, but their "representativeness" of the 411 students is unknown.

Second, once the data ("before" and "after" reports, peer reviews, outcome evaluation data, author and reviewer demographic data) are in hand, we knew that there would still be a number of complex, time-consuming decisions and analyses to undertake, which would require working back and forth between different databases and different data sources. Limiting the analysis to relatively small number of cases was judged essential, at least the first time through.

Third, there was by design a great deal of variability within the study of the meaning of the peer review task to the students. At a couple of campuses, instructors actually graded the peer reviews written by their students; at most, only the final reports were graded. Other potential sources of interaction and variance abound.

Finally, in the tradition of recent sociology of science, we wanted to avoid asserting a "privileged" view on the overall quality of students' manuscripts (and, for that matter, their peer reviews). After all, the purpose of peer review is to establish the merits of a scientific report; from that perspective, there is no independent, "correct" measure of the report's worth that can then be compared to student peer reviewers' assessments. This creates a methodological and conceptual conundrum: how do you address the questions enumerated above without some independent measures of the quality of initial and final manuscripts? Our general approach has been three-part: (a) to decompose the elusive idea of quality into some measurable (albeit incomplete) constituents, (b) to move back and forth between data sources, constructs, and analysis; and (c) to recognize that what we conclude will be at best suggestive. Should this exploratory approach be productive,

[6] The system was designed to allow students to revise their reports many times before "publishing" them, and we really intended that revision would eradicate the traces of earlier drafts.

we can always return to the larger dataset with a less flexible (and problematic) methodology.

Sub-Study Two Methods

All of the initial data for Sub-Study Two were collected electronically. Tape backups of the project database were pulled from several dates and a set of records were identified that met the following conditions: (1) the authors and reviewers had provided informed consent, (b) we had a copy of the database from a date on which a given report had been submitted but not reviewed, (c) we had a copy of the database in which that report had been peer reviewed twice and the original author had "published" his or her post-review version. Using this process, we selected 21 reports and 42 accompanying peer reviews for this initial analysis.

Three of the researchers met on a weekly basis to develop and carry out a content analysis approach for the reports and reviews. Most of the coding was initially done by the fifth author (Kohl), with subsequent review by others. Analysis of the "before" and "after" experimental reports utilized a number of measures; in this paper we have restricted our attention to fairly easy-to-measure changes, specifically, how many lines in a report were changed, and how many question responses were changed (the reports were written using a question-driven format). We recognize that these two measures are intercorrelated, as well as being fairly naïve ways of describing changes in content, but we felt that they would suffice as a starting point.

Each REPORT was assigned the following scores:

1. Did the report change from initial to final posting? (Boolean)
2. How many question-responses were changed? (Integer)
3. How many lines of text in the report were added or deleted? (Integer)
4. Did changes between the initial and final reports generally correspond to review comments? (Boolean)

Each report had two reviews. Because SPSS uses rectangular data matrices, we assigned most of the following measures to different variables, corresponding to the two reviews. In most cases, we also utilized a combined-review score. For example, measure #6 (list below) appears in our datafile in three columns: SubEla (the first reviewer), SubEla_2 (the second reviewer), and XSubEla (the sum of SubEla and SubEla_2):[7]

5. Was the overall assessment (positive or negative) of the review substantive and explicit? (Boolean)
6. Were reasons for the overall assessment elaborated? (Boolean)
7. Was evidence cited to support the assessment? (Boolean)

[7] Extensive analyses, not reported here, focused on the issue of interrater agreement and its effects, as well as other possible interactions. For example, is a substantive review plus a fairly insubstantial review more useful than two moderately substantive reviews? In general, we found that better stability of measures when we simply summed the two reviews. However, we did retain some constructs that look specifically at interrater agreement, and evaluate their utility in our regression models.

8. Were both strengths and weaknesses explicitly cited? (Boolean)
9. How technically proficient was the writing of the review? (three-point scale)
10. Did the quantitative scores assigned to the report agree with the written comments? (three-point scale)
11. How subject-matter knowledgeable did the reviewer appear to be? (three-point scale)
12. How long was the review? (three-point scale)

In addition, each review had two reviewer-assigned quantitative measures, which we labeled QScore1 and QScore2. QScore1 used a radio-button form to assign a score of 0-4 to the report, based on the reviewers' assessment of the report's overall substance and persuasiveness. QScore2 used a similar scale to evaluate the technical quality of the writing in the report.

The effects of reviews on students' writing was initially explored using Exploratory Data Analysis, including scatterplots and other graphics. One goal of this step was to better understand intercorrelations among the various measures and to assess the potential productivity of various constructs. Finally, we carried out a series of regression models to identify an economical set of predictors

Sub-Study Two: Preliminary Findings

A sequence of regressions on dependent measures showed little relationship between most review-related measures and the dependent measure. Furthermore, consistency between reviewers did not appear to be important. However, a very small number of predictors accounted for a significant amount of the variance in both dependent measures. Tables 3a –c show that 47% of the variance on the first measure is attributable to two review independent variables, both statistically significant. The first is QScore1, the quantitative reviewer-assigned score for the overall content of the report. Lower reviewer scores tended to yield greater changes between initial and final reports. (QScore2, which evaluated the quality of the authors' writing, was consistently non-significant).

The second significant predictor was a researcher-assigned score for the technical merits of the <u>review</u>. Poorly constructed reviews (e.g., reviews with serious grammatical or spelling errors) tended to stimulate less change by report authors than well-constructed reviews.

A third predictor—researcher-assigned assessment of the report author's subject-matter knowledge—approached significance but was omitted from the model. As you will see below, that predictor <u>was</u> significant in our second set of regressions.

Tables 3a, b, c. Regression on dependent measure "Number of lines changed"

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate

1	.682(a)	.466	.406	4.51259

a Predictors: (Constant), Technical quality of reviews (sum), QScore1 (substance) sum of reviewers

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	319.267	2	159.633	7.839	.004(a)
	Residual	366.543	18	20.363		
	Total	685.810	20			

a Predictors: (Constant), Technical quality of reviews (sum), QScore1 (substance) sum of reviewers
b Dependent Variable: Number lines changed in report

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	9.168	6.658		1.377	.185
	QScore1 (substance) sum of reviewers	-2.279	.769	-.514	-2.965	.008
	Technical quality of reviews (sum)	2.263	1.002	.392	2.259	.037

a Dependent Variable: Number lines changed in report

As noted previously, there is intercorrelation between our two dependent measures, so we should not be surprised to find similar predictors. The effects of QScore1 and the technical quality of the reviews is similar to that above. A third predictor, a researcher-assigned score for reviewer subject-matter knowledge, is also significant; but paradoxically, the sign of the beta coefficient is negative, suggesting that knowledgeable reviewers stimulate <u>less</u> change in students' reports than do reviewers with subject-matter misunderstandings. However, closer inspection revealed that, in fact, the knowledge construct was a three-part scale (1="Poor knowledge evident", 2="Difficult to evaluate", and 3="Good knowledge evident"), and all of the reviews were scored either 2 or 3. Consequently, the best way to interpret our third predictor is that reviews that commandingly demonstrated significant reviewer subject-matter knowledge stimulated less change than reviews in which the reviewer's competence was not as clear.

Tables 4a, b, c. Regression on "Number of questions changed"

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.684(a)	.468	.374	1.79805

a Predictors: (Constant), Knowledge of reviewers (sum), QScore1 (substance) sum of reviewers, Technical quality of reviews (sum)

ANOVA(b)

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	48.278	3	16.093	4.978	.012(a)
	Residual	54.960	17	3.233		
	Total	103.238	20			

a Predictors: (Constant), Knowledge of reviewers (sum), QScore1 (substance) sum of reviewers, Technical quality of reviews (sum)
b Dependent Variable: Number questions changed in report

Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	13.211	3.707		3.564	.002
	QScore1 (substance) sum of reviewers	-1.060	.311	-.617	-3.409	.003
	Technical quality of reviews (sum)	1.108	.592	.495	1.871	.079
	Knowledge of reviewers (sum)	-1.858	.823	-.605	-2.257	.037

a Dependent Variable: Number questions changed in report

Sub-Study One Findings

Exploratory data analysis, followed by a series of regression models, indicate that almost 50% of the variance in two (admittedly simple) dependent measures can be accounted for by a small number of independent measures. The independent measures include a reviewer-assigned "substantive content" score and the general technical competence of the review essay. Most other aspects of reviews appear to have little if any demonstrable effect on changes in students' scientific writing. These are intriguing findings, which we look forward to exploring further with a larger dataset.

Acknowledgements

This project was supported in part by NSF Award #9618142. The authors wish to thank Reizelie Barreto for her assistance with some of the analysis for Sub-Study Two.

Appendix A

Semi-structured Nature of Science (NOS) Interview Protocol (Pre and Post Interviews)

Research Question: What are students' understandings of the NOS and the establishment of scientific knowledge?

Question Bank
(some of the questions were not posed or asked differently according to the conversation)

- In your opinion, what is science?
- What makes science different from other disciplines of inquiry, such as religion or philosophy?
- What is a scientific problem? Can you give an example?
- What are some differences between a scientific problem and another kind of problem, such as a (NAME TWO: social, philosophical, historical, cultural, psychological, theological, economical, or political) problem?
- How do scientists decide that there is a problem they should solve?
- How can scientists be sure about their proposed solutions to those problems? (Predetermined coding for the responses of this question includes: "through experimenting" and "through peer review".)
- What is a scientific experiment?
- Does the development of scientific knowledge require experiments? (Follow-up: Please explain your response. Can you give an example?)
- Is science objective?
- What is a scientific fact?
- Is there a difference between a scientific fact and other kinds of facts?
- Do scientific facts ever change? Can you give me an example?
- How do scientists determine whether other scientists' factual claims are true?
- Is science value free?
- Are scientists influenced by societal, cultural, and personal beliefs and ways of viewing the world? Could you explain your response? Could you include any examples to justify your position?

Appendix B

Semi-structured College Peer Review Interview Protocol (Post Interview only)

Research Question: What are the students' experiences with and views toward the College Peer Review project?

<u>Question Bank (Almost all students were asked each question)</u>

- What did you like most about the toxicology project? What didn't you like? Why?
- Compared to other science classroom experiences you have had, how was the toxicology project unique?
- Do think your conceptions about the nature of science changed as a result of doing this project? How?
- Would you like to use a peer review system with your students when you begin teaching? Why/ Why not? How?
- Did you agree with your peers' evaluations of your work? Why/why not?
- As a student, did you feel uncomfortable with any aspect of the peer review system? Did you learn something from others work in peer review system? Can you give an example?
- Did you learn something new from others' evaluations of your own work? Could you explain?
- If you compare your online published reports before and after the peer reviews, do you think that you improved the quality of your report? How?
- If your online published reports hadn't been reviewed by someone else, but you had been asked to revise them later in the semester without that feedback, do you think that you would have made the revisions you did, anyway?

References

American Advancement of Association of Science. 1997. *Resources for Science Literacy: Professional Development*. New York: Oxford University Press.

American Association for Advancement of Science. 1993. *Benchmarks for Science Literacy*. New York: Oxford University Press.

American Advancement of Association of Science. 1989. *Science for all Americans*. New York: Oxford University Press.

Bell, L. R., Lederman, G. N. and Abd-El-Khalick, F. 2000. Developing and acting upon one's conceptions of the nature of science: A follow-up study. *Journal of Research in Science Teaching*, 37(6), 563-581.

Billington, H. 1997. Poster presentations and peer assessment: novel forms of evaluation and assessment. *Journal of Biological Education* 31:(2): 218-220.

Bos, N., J. Krajcik, and E. Soloway. 1997. Student publishing in a WWW digital library - goals and instructional support. Annual Meeting of the American Educational Research Association. Chicago, IL.

Glaser, B. G. 1969. The constant comparative method of qualitative analysis. pp. pp. 216-227 in G. L. McCall and J. L. Simmons, eds. Issues in Participant Observation: A Text and Reader. Reading, MA: Addison-Wesley.

Gratz, R. K. 1990. Improving lab report quality by model analysis, peer review, and revision. *Journal of College Science Teaching* 19: 292-295.

Henderson, L., and C. Buising. 2000. A peer-reviewed research assignment for large classes. *Journal of College Science Teaching* 30:(2): 109-113.

Hiltz, S., and B. Wellman. 1997. Asynchronous learning networks as a virtual classroom. *Communications of the ACM* 40:(9): 44-49.

Koprowski, J. L. 1997. Sharpening the craft of scientific writing. *Journal of College Science Teaching* 27: 133-135.

Lightfoot, J. T. 1998. A different method of teaching peer review systems. *American Journal of Physiology* 19:(1): S57-S61.

Liu, J., D. T. Pysarchik, and W. W. Taylor. 2002. Peer review in the classroom. *BioScience* 52:(9): 824-830.

MacLeod, L. 1999. Computer-aided peer review of writing. *Business Communication Quarterly* 62:(3): 87-94.

Matthews, M. R. 1996. The nature of science and science teaching, (981-999). In Fraser, B. J. & Tobin, K. G. (eds.). *International Handbook of Science Education*, 1996, Great Britain: Kluwer Academic Publishers.

McComas, W., M. Clough, and H. Almazora. 1998. The role and character of the nature of science in science education. pp. 3-39 in W. McComas, ed. The Nature of Science in Science Education: Rationales and Strategies. Dordrecht, The Netherlands: Kluwer Academic Publishers.

National Research Council. 1996. *National Science Education Standards*. Washington, D.C: National Academy Press.

Patton, Q. M. 1990. *Qualitative Evaluation and Research Methods, Second Edition*. Sage Publications: London.

Project Kaleidoscope. 1999. Steps Toward Reform. Washington, DC: Project

Kaleidoscope. Available online: http://www.pkal.org.
Robinson, R. 2001. Calibrated peer review: An application to increase student reading & writing skills. *The American Biology Teacher* 63:(7): 474-480.
Rushton, C., P. Ramsey, and R. Rada. 1993. Peer assessment in a collaborative hypermedia environment: a case study. *Journal of Computer-Based Instruction* 20:(3): 75-80.
Schwartz, R.S., & Lederman, N.G. 2002. "It's the nature of the beast:" The influence of knowledge and intentions on learning and teaching nature of science. *Journal of Research in Science Teaching*, 39, 205-236.
Siebert, E. D., and W. J. McIntosh. 2001. *College Pathways to the Science Education Standards*. Arlington, VA: NSTA Press.
Smith, M. U., & Scharmann, L. C. 1999. Defining versus describing the nature of science: A pragmatic analysis for classroom teachers and science educators. *Science Education*, 83, 493-509.
Sullivan, D., C. E. Brown, and N. L. Nielson. 1998. Cyberdimensions: computer-mediated peer review of student papers. *Journal of Education for Business* 74:(2): 117-121.
Topping, K. 1998. Peer assessment between students in colleges and universities. *Review of Educational Research* 68: 249-276.
Towns, M. H., K. Marden, D. Sauder, R. Stout, G. Long, M. Waxman, M. Kahlow, and T. Zielinski. 2000. Interinstitutional peer review on the Internet. *Journal of College Science Teaching* 30:(4): 256-260.
Trautmann, N., W. Carlsen, M. Krasny, and C. Cunningham. 2001. *Assessing Toxic Risk, Student Edition and Teachers' Manual*. Arlington, VA: NSTA Press.
Yalvac, B., and W. S. Carlsen. 2002. A qualitative study of college students' experiences with an online peer review system and effects on students' understanding of the nature of science. Paper presented at the annual meeting of the American Education Research Association. New Orleans, LA.