ORIGINAL RESEARCH AND PEER REVIEW USING WEB-BASED COLLABORATIVE TOOLS BY COLLEGE STUDENTS

Mustafa Çakır[1], William S. Carlsen[2]
[1]Marmara Üniversitesi, Atatürk Eğitim Fakültesi, Orta Öğretim Fen ve Matematik Alanları Eğitimi, 34722, Istanbul
[2]Pennsylvania State University, College of Education, University Park, 16802, Pennsylvania, USA
mustafacakir@marmara.edu.tr

ABSTRACT
The Environmental Inquiry program supports inquiry based, student-centered science teaching on selected topics in the environmental sciences. Many teachers are unfamiliar with both the underlying science of toxicology, and the process and importance of peer review in scientific method. The protocol and peer review process was tested with college students at 11 universities around the United States. The overall goal was to promote science education by engaging students in a sociologically authentic scientific research including anonymous peer review. Students were provided with the methods and knowledge to conduct a toxicology experiment and the technology needed for communication. They conducted a bioassay experiment, posted their results on a web, and completed anonymous peer reviews. Data consisted of peer reviews, anonymous online questionnaire, and another questionnaire about students' experiences and their evaluation of the project. There were statistically significant differences among schools in scores received for the quality of the argument and quality of technical writing. However, the only statistically significant difference concerned the average score received was the quality of technical writing. The findings suggested that the research and peer review protocols could be adapted for use by introductory level college science students, including prospective science teachers.

Keywords: Peer Evaluation; Environmental Inquiry; Science Teacher Education; Computer Uses in Education

INTRODUCTION
The Environmental Inquiry program supports inquiry based, student-centered science teaching on selected topics in the environmental sciences. Texts to support high school student research are published by The National Science Teachers Association (NSTA) in the domains of environmental toxicology, watershed dynamics, biodegradation, and the ecology of invasive species. The first of these publications, What's the Risk?, was published in 2001 and includes bioassay protocols for assessing the toxicity of substances. Secondary school science students can post the results of their bioassays on a web server and participate in a process of anonymous peer review and "publication" of their research. Teachers and secondary students who have participated in the process reported finding it interesting and useful; however, we recognized that many teachers are unfamiliar with both the underlying science (toxicology) and the process and importance of peer review in scientific method. We tested the protocol and peer review process with prospective science teachers in a secondary science methods course at Penn State, using a companion website set up specifically for college-level students. The College Peer Review project is a multi-university project that has been implemented every academic semester since fall 2001 (Trautmann, Carlsen, Yalvac, Cakir, & Kohl, 2003). The results of that test suggested that research and peer review protocols could be adapted for use by introductory level college science students, including prospective science teachers. This paper reports the results of a multi-site expansion and test of that work.

Participants and Purpose of the Study
This research involved college students in science courses, pre-service science education courses, and science studies courses at 11 colleges and universities around the United States. The overall goal of the project was to promote science education by engaging students in a sociologically authentic scientific research project including anonymous peer review. The project was designed to enable students to experience science as a mode of inquiry rather than a static collection of facts.

The aim of quantitative analysis was to identify the aspects of the project that are working and the aspects that need to be improved or omitted. This paper presents some quantitative data from the 11-campus project. Data are included from 10 campuses (the eleventh yielded only one student's data and is omitted from the analysis). This research intended to be used as a resource for discussion of the project and the development of plans for "next steps" and to understand the participants' initial engagements and attitudes toward the project by answering the following questions:

- What do students perceive as the strengths and weaknesses of the model, rating the protocol specifications and written materials, the online systems, the quality of the reviews they received, and the extent to which they perceived that their experiences were scientifically "authentic?"
- How are the final drafts of students' research reports affected by peer reviews?

- Do reports improve significantly when authors receive detailed, consistent reviews?

METHODS AND PROCEDURES

In the project, students were engaged in open-ended scientific investigations (Trautmann, Carlsen, Krasny, & Cunningham, 2001). Participants were provided with the methods and knowledge of science to conduct a toxicology experiment and they used the necessary tools (e.g., the chemicals, the organisms, Petri dishes) and methods (e.g., counting the number of germinated seeds, measuring the root length in mm) to finish their investigations. All activities were organized to provide an opportunity for students to learn how to frame research questions, design and carry out experiments, critically analyze their results, write a report, and defend their conclusions to their peers. Participating students engaged in original research, computer-mediated collaboration, peer review, and online publishing. They conducted a bioassay experiment, posted their results on a web server, and completed anonymous peer reviews. Peer reviews were submitted using an online form. A questionnaire with both fixed-format and open-response questions was administered anonymously at the end of the semester. Participants were asked to help us evaluate the College Peer Review project by completing a questionnaire about their experiences. Evaluation of the questionnaires helped us to determine the value of the project and to guide the project's future development.

Students worked in pairs to conduct the bioassay experiment and tally their results, but posted individual reports and completed individual peer reviews. The reports followed a common, question-driven format, and quantitative data were entered using a table tool. After completing their own lab reports, students had about a week to complete online peer reviews of two other students' projects. Students composed their peer reviews using a structured data entry screen with two quantitative items and three essay items.

Peer reviews were anonymous; only report authors and instructors were given access to their contents. The matching of reports and reviewers was nonrandom but anonymous across institutions.[1] User data, reports, and peer reviews were stored in the database in related tables. The final common stage of the project was "publication" of reports after students made revisions using peer review feedback. Since many of the major activities of the project occurred online (report writing, peer review, publication) most of the data were collected automatically.

Data Analysis and Discussion

Analysis began by reorganizing data tables that had been collected by our server using Microsoft Access. The first task was data cleaning and the creation of one inclusive table by combining a user table, reports table, written reviews table, received reviews table, and final questionnaire table. Once a comprehensive clean data table was created in Access, it was exported to statistical software (SPSS) for quantitative analysis. There were 411 participants. 341 (83%) gave permission for us to use their responses in research. A number of checks of participant-response bias were done and no meaningful differences between permission-granters and others were detected. The following analyses are limited to the 341 individuals who gave consent. However, the peer review scores assigned to consenters by non-consenters are included, without any identifying information about the latter. In the following pages, data are presented as were gathered by the automated system. Discussion to address related issues and their relevance are provided where necessary.

Are you in a teacher education program?

Although there were teacher education students at most of the participating colleges, they were outnumbered by science majors. 44 participants' major could not be identified (this information was provided in the final questionnaire, which not all consenters completed); therefore out of 341 participants, 297 are reported. Out of 297 participants 94 (%31.6) were in teacher education program and 203 (%68.4) were not. The following table reports the number of students and whether they are in a teacher education program, by school.

[1] Students at the different universities completed the experiment at different times within an approximately two-month time frame. Instructions to students about how to select reports to review were left to the instructors' discretion. At Penn State, for example, we had our students complete the experiment first, then asked them to hold off on completing reviews until the results had been posted from two other institutions. At least one instructor encouraged his students to try to review another report that assessed the toxicity of the same chemical they had assessed. In most cases, however, students chose reports to review based only on the title of the report, which included the name of the chemical being assessed and an author-determined 5-digit code. Lab partners shared their 5-digit codes with each other so they could avoid reviewing their partner's report, which would have presented a conflict of interest.

Table 1: Number of Students and Whether They are in a Teacher Education Program, by School

University										
School code	1	3	4	5	6	7	8	10	11	12	Total
Number of students											
Not a teacher ed student		20			1	123		1	9	49	203
Teacher ed student	16		28	12	2	5	11		20		94
Total teacher ed status known	16	20	28	12	3	128	11	1	29	49	297
Teacher ed status unknown	0	0	0	0	3	0	0	9	31	0	43

Missing values = 44, 12.9% of the total N of 341 consenters. One non-consenting participant is omitted, the only student from an 11th university.

What are your gender and minority group affiliations?

74 participants (21.7%) were male. Analyses did not yield significant differences on any variables between male and female students. Differences among schools in gender distribution were not statistically significant. With the exception of one school, universities with more than six participants all had female participants outnumbering male participants by at least three to one. This was true among science courses as well as science education courses. 17.6% of the students who completed the final questionnaire identified themselves as members of underrepresented minority groups (African-American, Hispanic, and Native American). There were no statistically significant differences associated with this response on any measure.

Basic descriptive statistics for the final student questionnaire

Of the 341 students who submitted reports and gave consent for research, 192 (57% of consenters) completed the final questionnaire. Summary statistics from the questionnaire are reported below. We used Likert-scale items, where 1= “strongly disagree,” 2=”disagree somewhat,” 3 = “Neutral,” 4 = “Agree somewhat,” and 5 = “Strongly agree.”

Table 2: Items and Summary Statistics from the Questionnaire

	Descriptive Statistics	N	Mean for all students	Mean for teacher ed.	Mean for other students
1	I learned something by writing peer review comments	192	3.96	3.82	4.05
2	I felt qualified to provide meaningful peer review of other students' reports	192	3.73	3.65	3.78
3	I believe that the peer reviews I wrote should be helpful to the students that received them	192	3.98	3.97	3.99
4	Peer reviewing other students has helped me to think more critically	193	4.10	4.08	4.11
5	Peer reviewing other students has helped me to improve my own scientific writing	193	4.02	3.90	4.08
6	I received useful peer review comments about my own report	192	3.53	3.36	3.63
7	The quantitative scores I received from peer reviewers were fair	192	3.60	3.51	3.66
8	I changed my mind about something in my report because of comments I received through peer review	192	2.99	2.94	3.02
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	192	3.71	3.69	3.72
10	I think that meaningful peer review is a reasonable expectation for college students	190	4.23	4.21	4.24

11	I think that meaningful peer review would be a reasonable expectation for high school students	190	3.88	3.96	3.84

None of the above differences is statistically significant at $p < .05$.
Although teacher education student means were lower for all items except item 11, these differences are not statistically significantly (ANOVA with correction for multiple t-tests). However, it is worth noting that item 11 evaluates high school students' ability to provide sound feedback to each other. Table 3 and Table 4 provide brief individual descriptive statistics for each final questionnaire item below.

Table 3: Frequencies and percentages for item 1 through item 5

	Item 1		Item 2		Item 3		Item 4		Item 5	
	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
Strongly disagree	2	1.0	7	3.6	3	1.6	6	3.1	2	1.0
Disagree	7	3.6	18	9.4	2	1.0	1	.5	8	4.1
Neutral	32	16.7	34	17.7	29	15.1	29	15.0	37	19.2
Agree	106	55.2	94	49.0	119	62.0	89	46.1	84	43.5
Strongly agree	45	23.4	39	20.3	39	20.3	68	35.2	62	32.1
Total	192	100	192	100	192	100	193	100	193	100

A majority of the respondents (79%) agreed that they learned something by writing peer review comments. 79% of the students reported that they felt qualified to provide meaningful reviews of other students' reports. 82% of the students thought they provided helpful reviews, and less than 3% anticipated that their review would not be helpful. 82% of the students agreed that peer reviewing enabled them to reflect and think about their own and others' research more critically. Providing feedback on other students' research reports was perceived beneficial by students. 75% of the respondents agreed that their technical writing improved because of the peer reviewing process.

Table 4: Frequencies and percentages for item 6 through item 11

	Item 6		Item 7		Item 8		Item 9		Item 10		Item 11	
	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%	Freq	%
Strongly disagree	10	5.2	7	3.6	30	15.6	13	6.8	4	2.1	5	3
Disagree	22	11.5	14	7.3	38	19.8	20	10.4	2	1.1	9	5
Neutral	51	26.6	69	35.9	49	25.5	36	18.8	17	8.9	33	17
Agree	74	38.5	60	31.3	54	28.1	63	32.8	91	47.9	87	45
Strongly agree	35	18.2	42	21.9	21	10.9	60	31.3	76	40.0	58	30
Total	192	100	192	100	192	100	192	100	190	100	192	100

Although 82% of the students thought they provided helpful reviews, only 57% reported that they received helpful reviews. 18% of students reported that peer reviews did not help them to improve their reports. Most of the students thought their peers were fair when they rated the quality of the reports. Previous research has shown that marks given by students can be as reliable as those given by instructors (Orpen, 1982). 11% of the participants reported that their score were "unfair." 39% of the students agreed that they changed their minds about some aspect of their report because of feedback they received via peer review. This might be attributed in part to the implications of peer evaluation, which involve a different relationship that that between instructors and students. It may contribute to a collaborative role rather than an adversarial one (Billington, 1997).

A majority of students felt positive about the anonymity of peer review. This is consistent with what actually happens in scientific community. According to Arnold Relman, the chief editor of the New England Journal of Medicine, about 85% of their reviewers have preferred to remain anonymous, and report that they are more candid and rigorous when they are not required to sign their reviews. 87% thought college students could provide meaningful and helpful peer reviews. Previous research has suggested that students appreciate the opportunity to comment on each other's work in a constructive manner, and that peer review can instill a sense of community within a class (Hay & Miller, 1992). When students were asked if it was realistic to expect meaningful reviews from high school students, 75% responded positively. There is no significant difference between teacher education students and other students on this measure. However as noted earlier, this item was the sole item on which teacher education students felt more positive than other students.

School Differences in Quantitative Review Scores

In their peer reviews, students rated the quality of the argument and the quality of authors' technical writing by assigning a score to each. We found some statistically significant differences between schools. An ANOVA procedure was used to detect these differences and then post hoc analyses were done to identify pair wise differences between schools.

The first measure, which was QScore1, asked reviewers to answer the question, "Did the author address each question fully and provide good support for his or her conclusions?" Responses were reported on a five-point scale ranging from 5 = "Excellent. Exceptionally well done" to 1 = "Failure. Unacceptable responses; report should be restarted from scratch." This was called the "quality of argument" score. Students at School 6 received significantly higher scores on this measure than students at Schools 3, 10, and 12. Because School 6 had a small number of participants (n=6), this result should be carefully interpreted. There were no other pairwise differences. Table 5 gives the ANOVA results for the quality of argument.

Table 5: One-Way ANOVA results for QSCO1 by SCHOOL

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
SCHOOL	9	11.9	1.32	2.53	0.0082
Error	309	161.1	0.52		
Total	318	173			

Post hoc tests

Duncan Grouping	Mean	N	SCHOOL
A	3.80	5	6
B	2.69	49	12
B	2.58	19	3
B	2.50	9	10

Significant differences at p<.05, means with the same letter are not significantly different

There were significant differences among schools in scores received for quality of technical writing (QScore2received). One-way ANOVA was performed, followed up with Duncan grouping post hoc analysis for pairwise comparisons, Table 6. Three groups of schools were identified, as seen in the table below, with statistically different average received mean scores. Schools 6 and 5 comprised two discrete "groups," A and B. Table 6 presents one-way analysis of variance results for quality of technical writing across schools. Schools 1, 3, 7, and 12 comprise a third group with a significantly different mean score, when compared to Groups A & B. There were no other differences.

Table 6: One-Way ANOVA results for QSCO2 received by SCHOOL

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
SCHOOL	9	15	1.67	3.53	0.0003
Error	309	145	0.47		
Total	318	160			

Post hoc tests

Duncan Grouping	Mean	N	SCHOOL
A	3.90	5	6
B	3.20	12	5
C	2.59	49	12
C	2.57	126	7
C	2.53	16	1
C	2.49	19	3

Significant differences at $p<.05$, means with the same letter are not significantly different

Students in each participating college reviewed and scored other students' reports. Scores on technical quality of reviewed reports were labeled as variable QSCO2Written. ANOVA results in Table 7 show that students at School 6 awarded significantly higher scores to others concerning the technical quality of reviewed reports, an interesting phenomenon given that they also received the highest scores. Students School 5 awarded significantly lower scores; however, they received the second highest scores for their reports. (Please note that these are only preliminary analyses; we still need to look at issues like which schools tended to review which other schools. Again, the matching of reports to reviewers was anonymous but not random, and it is likely that students were most likely to review reports by other students from their own campus, because their reports were most likely to be available for review at the time each campus's reviews were required by the relevant instructor).

Table 7: One-Way ANOVA results for QSCO2 Written by SCHOOL

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
SCHOOL	10	13.85	1.39	2.91	0.0017
Error	298	141.86	0.48		
Total	308	155.70			

Post hoc tests

Duncan Grouping	Mean	N	SCHOOL
A	3.5	4	6
B	2.6	122	7

Duncan Grouping	Mean	N	SCHOOL
B	2.5	5	10
B	2.2	12	5

Significant differences at p<.05. Means with the same letter are not significantly different

Differences in Quantitative Review Scores for Teacher Education Students
Students in teacher education programs generally received and assigned higher mean scores than non-teacher education students. However, among the differences in mean scores for all four measures, the only statistically significant difference concerned the average score received for the quality of technical writing. Table 8 reports that teacher education students were able to articulate their research and communicate results in a more effective way than the students who are majored in sciences or science studies. Analysis of variance results for quality of technical writing received score by major is reported in Table 9.

Table 8: Written and Received Score Differences in Reviews for Teacher Education Majors

Teacher Education	QScore1 Received	QScore2 Received *****	QScore1 Written	QScore2 Written
No	2.7508	2.60017	2.7362	2.6503
Yes	2.9147	2.84425	2.8653	2.7991

*Only the received quality of technical writing received score (QScore2) is statistically significant at p<.05.

Table 9: One-Way ANOVA results for QSCO2Received by Teacher Ed.

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Teacher Educ.	1	3.51	3.51	7.08	0.0082
Error	280	138.9	0.5		
Total	281	142.41			

CONCLUSION
Findings suggested that participants found the peer review and the original research aspects of the project engaging, unique and interesting. They enjoyed their experiences with the project activities, working in groups and the online collaboration. Through its original research, peer review, and online collaboration aspects, College Peer Review project led students to appreciate the social characteristics of science. As noted at the beginning of this paper, these are findings from a research study which is intended as background information to stimulate subsequent discussion and analysis by participating faculty and other interested researchers.

In looking for differences by school and other factors, our primary interest was in developing questions to guide formative evaluation of this project. For example, what are the advantages and disadvantages of restricting participation in a project like this to prospective science teachers? Do between-school differences lead to differences in review-related outcomes? Do positive experiences as a reviewer and as a review-receiver favorably incline pre-service teacher participants to consider using peer review with their own students some day?

REFERENCES
Billington, L. H. (1997). Poster presentations and peer assessment: Novel forms of evaluation and assessment. *Journal of Biological Education, 31*(3), 218-220.
Hay, I., & Miller, R. (1992). Application of a poster exercise in an advanced undergraduate geography course. *Journal of Geography in Higher Education, 16*, 199-215.

Orpen, C. (1982). Students' versus lecturer assessment of learning: a research note. *Higher Education, 11*, 567-572.

Trautmann, N. M., Carlsen, W. S., Krasny, M. E., & Cunningham, C. M. (2001). *Assessing toxic risk, student edition and teachers' manual*. Arlington, VA: NSTA Press.

Trautmann, N. M., Carlsen, W. S., Yalvac, B., Cakir, M., & Kohl, C. (2003, March). Learning nature of science concepts through on-line peer review of student research reports. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Philadelphia, PA.