Journal of Science Education and Technology, Vol. 9, No. 2, 2000

High School Students' Critical Evaluation of Scientific Resources on the World Wide Web

Nathan Bos[1]

This research explores a new web-based curriculum idea, that of having students write and publish critical web ''reviews'' of scientific resources as a means of both practicing critical evaluation of web resources, and of making an authentic value-added contribution to the web. This paper presents content analyses of selected sections of 63 web reviews published by eleventh grade students in a project-based science class. Two aspects of critical evaluation are focused upon: summarization of content and evaluation of credibility. Content analyses show that student summaries were usually accurate, but had room for improvement especially in areas of comprehensiveness and level of detail. An ideal model of a content review is developed from analysis of a second set of reviews. When asked to evaluate credibility, students struggled to identify scientific evidence of claims in web resources, but analysis of web documents shows that this is often because such evidence is missing. Students could accurately determine the publishing source of web documents, but challenges arose in identifying potential biases. Recommendations for future iterations of this curriculum idea are presented throughout. A companion paper that will appear in this journal will examine how student reviews serve the function of social filtering on the web.

KEY WORDS: World Wide Web; critical evaluation; critical thinking; project-based science; social filtering; digital library.

COMPONENTS OF CRITICAL EVALUATION

The World Wide Web is an exciting an challenging new information resource for K–12 science. It is exciting because of the incredible breadth and diversity of scientific resources now available from any networked computer. It is challenging because the diverse and uneven nature of the web demands that students develop new skills of critical evaluation. These critical evaluation skills are an important part of media literacy for students who will likely have access to the Internet and other distributed information sources throughout their lives and careers (Callison, 1993; Cunningham, 1997; Smith, 1997). Modern science education standards also call for a renewed emphasis on process skills of critical evaluation, including evaluation of materials accessed from scientific data bases (National Research Council, 1995). Our model of critical evaluation has four components: summarization of content, evaluation of credibility, evaluation of organization, and evaluation of use of media. These four categories are grounded in previous research in critical evaluation, educational psychology, and science education, as will be described.

Because the focus of this paper is on critical evaluation in the content domain of high school science, the first two of these components, summarization of content and evaluation of credibility, are examined here. A companion paper (Bos, in press) will focus on more general characteristics of useful reviews, and examine how students evaluated docu-

[1]University of Michigan, 701 Tappan C2420, Ann Arbor, Michigan 48109-1234. e-mail: serp@umich.edu

ment organization and use of graphics. What aspects of a scientific resource should students be able to critically evaluate?

Content Summarization

The first category of critical evaluation is content summarization. To evaluate and effectively review a web resource, students must be able to identify and describe the main topics of those resources. The ability to extract the "gist" from a text is a well-studied reading skill (Brown and Day, 1983; Kintsch and van Dijk, 1978; Pressley and Afflerbach, 1995), and previous research has examined expert strategies for summarizing. By the high school level, students are often capable of creating topic sentences not taken directly from the text, and can produce superordinate terms that encompass content (Brown and Day, 1983). The higher-level skills of creating new topic sentences do not necessarily arise by themselves, are sometimes missing from post-secondary students (Brown and Day, 1983), but are teachable to students at a younger age, and once taught, seem to transfer between content domains (Palincsar and Brown, 1984). This paper will begin to identify the key challenges for summarizing content in the unique context of the web.

Evaluation of Credibility

Because of the variety of sources and purposes of web information, perhaps the key component of critical evaluation on the web is the ability to make judgments about the credibility of resources. Previous research in electronic environments has shown that students do not spontaneously evaluate credibility of the resources that they are accessing (Marchionini, 1995; Pitts, 1994).

In this research, two related but different approaches to establishing credibility of web resources were implemented. In project #1, a domain-specific method was used, derived from science education research and focusing on scientific evidence. In project #2, evaluation of credibility was more domain-independent, and was derived from methods recommended by general information specialists. More explanation of the differences between these two is warranted.

Domain-Specific Evaluation of Credibility

Experts in a discipline can evaluate information in a scientific resource according to the quality of its evidence and, in some cases, the methods of data collection and analysis. There is also a current push in science education to help students, become more proficient in these discipline-specific ways of thinking. *The National Science Education Standards* (National Research Council, 1995) advocates that high school students should learn to think about evidence for scientific information, should be able to evaluate their own and others' methods of investigations, and should use these skills in a variety of settings, including information extracted from electronic data bases. As a means of engaging in domain-expert–like evaluations, students can examine web-based resources to see whether scientific evidence is given to support scientific claims in the resource.

For this study, I adapt a definition of scientific evidence taken from Deanne Kuhn's work. Kuhn (1991, p. 45) lists two requirements for information to be considered evidence: evidence must bear on the claims it supports, and evidence must be distinguishable from the theory itself. To illustrate these requirements, consider the scientific claim (perhaps given on a web page) that air pollution is detrimental to health. One argument that would not be evidence for this claim is the argument that "Humans are polluting the oceans and lands at an alarming rate, and these pollutants have known health effects." Although conceptually related, this statement is not theoretically related to the relationship between air pollution and health effects, and thus would violate the first component of Kuhn's theory of evidence. An example of a violation of the second part of the definition would be an explanation of how air pollution is thought to affect health, e.g., "tiny pollution particles penetrate the membrane of the lungs, and cause damage there." Although this may be true, it is not independent corroborating data, but rather is non-evidential information that elaborates the theory. An example of true evidence for the claim that air pollution causes health problems might be statistical data showing that lung cancer mortality rates are higher in cities with higher particulate pollution counts. Note that to qualify for Kuhn's definition of evidence, information does not have to be conclusive—the above evidence is correlational rather than causal. But this evidence does fit Kuhn's two criteria, because measures of pollution are independent of measures of health problems, and because there is a plausible theoretical link whereby one might cause the other. In contrast, in her studies of informal reasoning, Kuhn often found that subjects would often give various forms of elaboration or other non-evidence to support claims.

Although Kuhn's definition is conceptually difficult to implement and teach, it provides a rigorous standard that will be useful for analyzing both web-based scientific resources and students' critical reviews of them.

Domain-Independent Evaluation of Credibility

The field of library and information science provides domain-independent methods for establishing credibility of different types of web source materials. Reference librarians, who are experts in evaluation but are not usually experts in scientific subdisciplines, have methods for evaluating sources that do not rely their own assessments of evidence or data analysis methods. This paper will draw on three on-line guides to evaluating web resources, published by professional library staff at three universities (Alexander and Tate, 1996; Engle, 1997; Grassian, 1995) (see also Smith, 1997). Comparing across these three guides, five categories recommended for establishing credibility of web resources are identified: the identity of the web author, identity of the publisher, detectable bias in the source, detectable bias in the text, and date of publication/revision. Three of these categories are used the design experiments described in this paper: identity of the sponsoring organization, bias related to the source, and bias detectable in the text.

Identity of the Sponsoring Organization. Alexander and Tate (1996) stress the sponsoring organization above the individual author. They advise looking for links that describe the sponsoring organization, and also looking for ways to independently verify the legitimacy of the page's sponsor. There is some disagreement among guides here, in that two of the guides stress identification of the individual author above identification of the sponsoring organization. I agree with Alexander and Tate that in a K–12 setting, identifying the sponsoring organizations is more important than identifying named individual authors. For example, it is more important for a high school student to know that certain information comes from the Environmental Protection Agency (EPA) than it is for them to name individual page authors.

Bias Related to Source. A particular source, e.g., a private corporation, may have a bias related to the information given in the web resource. Each guide treats this source of bias in a slightly different way. Grassian (1995) advises readers to look for third-party financial sponsorship. Alexander and Tate (1996) stress identifying the type of publication. Their analysis scheme is based on identifying whether a resource is primarily intended for advocacy, business, news, personal, or public information, and provide slightly different analysis schemes for each type. Engle (1997) similarly stresses identifying the purpose for publication of a particular resource.

Bias Detectable in Text. Besides bias related to the identity of the publisher, there may be detectable bias in the way a resource is written. Engle (1997) advises readers to look for "objective reasoning," although this guide does not give much specific advice on what the characteristics of objective reasoning are. Alexander and Tate (1996) ask whether advertising is evident on a page, and whether it is clearly separate from information content.

These three categories offer a starting point for students to evaluate the credibility of web resources. Content analyses in this paper will examine how students performed these analyses.

RESEARCH QUESTIONS

Analysis of data from two design experiments provided data to address the three research questions listed below.

1. Can students summarize the scientific resources that they find on the web?
2. Can students identify and evaluate evidence in the scientific resources that they find on the web?
3. Can students identify the source and potential biases of the scientific resources that they find on the web?

SUBJECTS AND SETTING

The setting for this study was two eleventh grade science sections at an alternative high school in a medium-sized midwestern college town. This high school of approximately 400 students accepts new students through a combination of lottery and first-come-first-served sign ups each school year. The lottery system is controlled to ensure a proportion of both minority students and special-needs students. Although students at this school are not a representative sample of any particular population, they are not believed to be overly represented by either high- or low-achieving students. Most graduates of this school do attend college.

The class involved in these studies was in the third year of the Foundations of Science (FOS) sequence, which is an integrated science curriculum that follows the principles of projectbased science (Blumenfeld *et al.,* 1991) and has a heavy emphasis on the use of educational technology. The science curriculum integrates the three traditional content areas of Earth science, biology, and chemistry into one 3-year sequence focused on investigative science.

Forty-four students took part in these projects, in two sections taught by different teachers. There were 27 girls and 17 boys in this group. In each review-writing project, students were allowed the choice of working in groups or individually, with the stipulation that they needed to produce the equivalent of one review each; e.g., a group of two students could work together to write two reviews, or could work individually and each write one review.

Students wrote web reviews as part of their normal background research at the beginning of two 8-week projects. At the end of each of these two projects, students worked in groups to produce a culminating artifact, which were also published on the web. The air pollution project (project #1) took place in September and October, and the final artifacts were reports of students' local testing of air pollutants. The infectious diseases project (project #2) took place in February and March of the same school year, with the final product being a Hyper Studio hypermedia report about a particular disease.

DATA SOURCES AND METHODS

The data for this paper are 63 reviews published by students in the two review-writing projects, and the 41 original web source documents that could be retrieved by researchers shortly after the projects' completion. This research will present a series of short content analyses of different sections of the students' reviews. Most of the content analyses will be focused on a single section of the reviews published in one project.

Forty-one original source documents are also examined. A number of source documents are missing because of one of several reasons. Some students also chose to review resources that were not on the web, such as library books or magazine articles. Although this was perfectly acceptable as part of their review-writing assignment, these reviews were excluded from analyses because they are outside of the focus of this paper. In some cases students appear to have give an incorrect URL and the researcher was unable to find the correct one, and in other cases it appears that the pages were either removed from the web or moved to different locations shortly after the students completed their reviews.

Software Environment

Students published reviews by filling in an HTML form, which solicited comments in text fields tailored to each project (Appendices B and C). After students submitted a review, a cgi script parsed the student reviews and published it, along with other reviews from the students' class section. Review-publishing was supported on the high school's own Macintosh server, although it could as easily have been supported on another Macintosh server at a remote location. We used Maxum's $136 *Netforms* software, which allows non-programmers to parse the output of HTML forms and write the contents to web pages or other text files. This review-publishing model represents an inexpensive, scalable model that could be implemented by most school districts.

Content Analyses

Results for this paper are based on a series of content analyses of student reviews and web source documents. Content analyses of unambiguous features of these documents were performed by the first author. When analyses were more ambiguous, a second rater was consulted to confirm and provide a reliability check on the first author's judgments. Unless otherwise stated, all content analyses are reported as frequency tables of reviews, where *n* is a number of reviews, not a number of unique documents (some web documents were reviewed multiple times) and not a number of authors (each review had between 1–4 student authors.)

RESULTS

Strengths and Shortcomings of Student-Written Reviews of Scientific Resources

In the remainder of this paper, I examine student-written reviews to determine strength and weaknesses of these first attempts at student-authored critical evaluations. Analyses are divided by project.

Analyses of the following sections of each project are presented here:

- Project #1 air pollution
  - Summarization of content through identifying claims
  - Evaluation of credibility through identification of scientific evidence
- Project #2 infectious diseases
  - Summarization of content through writing of an abstract
  - Evaluation of credibility through identification of source and bias

Analysis of Selected Review Fields in Project #1, Air Pollution

Summarization of Content Through Identifying Claims (Project #1)

In project 1, we asked students to identify what the scientific claims were that were being made in the web pages they reviewed. This is a disciplinary-specific method of summarizing content. Scientific writing such as is published in professional journals are often organized around a few central "claims," and a normal way of critiquing such writing is to make sure the claims are defined tightly and match the evidence given. The ratings scheme used to compare students evolved through several iterations by the first author. Students review summaries had some strengths and unique qualities that were not captured by the version of this scheme reported here, but which are reported in an earlier paper (Bos, 1997). The categories used were accuracy, centrality, level, and comprehensiveness.

The best way to understand the four categories used in this ratings scheme is by example. Therefore, Fig. 1 shows a hypothetical page, whose content is represented by the concept map in Fig. 1, column 1, and whose "claims" are illustrated by the exemplary review in Fig. 1, column 2. The four ratings categories of accuracy, level of detail, centrality, and comprehensiveness of claims are explained in Fig. 2, column 1, and example deficits in summaries of claims are given in Fig. 2, column 2.

Ratings Process. Because there is a degree of ambiguity in each of the last three ratings categories, a second rater was employed to provide a check on the first author's ratings. Both raters independently scored each of 24 reviews. Where there were disagreements between raters, the disagreements were reconciled using a negotiation process. Although some room for subjective judgment still exists within this scheme, I hope that these ratings are useful in painting an overall picture of the students' performance on the claims-identification task. This analysis has proved useful for guiding the summarization in project #2. Table I shows the agreed-upon ratings for level of detail, centrality of claims, and comprehensiveness of claims in student reviews.

Accuracy. Students mostly avoided outright inaccuracy in reporting claims, which is an encouraging finding considering the difficulty of much web source material. Only one student claim was clearly inaccurate. Five other articles received "qualified" ratings on accuracy. These were not outright mistakes, but claims that missed the point, describing claims that were orthogonal to the main point of the article. This problem is better captured in discussion of centrality and comprehensiveness.

Level. The second challenge in identifying claims is to capture the correct level of detail. In 9 of the 20 reviews, student reports of claims were rated as too broad. An example is, in a review of an EPA page on automobile emissions, the student describes the claim as "this resource claims that pollution from automobiles greatly contributes to the air pollution problem." Although accurate, this reporting of claims does not capture the level of scientific reporting present in the article, which mostly made claims about several specific chemical pollutants in car exhaust. Three students also made the opposite error, specifying claims that were too specific. Students would sometimes pick out very specific statistics or facts and present them as the article's claim. Both of these mistakes may have been exacerbated because secondary source materials often includes densely worded information without clearly identifying a few key claims.

Centrality. The third challenge for students is to identify central claims. Eight of 20 reviews were judged to have captured central claims, while 9 of 20 were only partially correct, and 3 of 20 identified claims that were entirely peripheral (or identified no claims). As before, two factors made identification of central claims more difficult for students: first, many of the secondary source documents were organized more as broad overviews of information rather than being focused on a few central claims, and second the technical level of some of the articles was quite high. Students also sometimes picked out only those claims that were relevant to their own research.

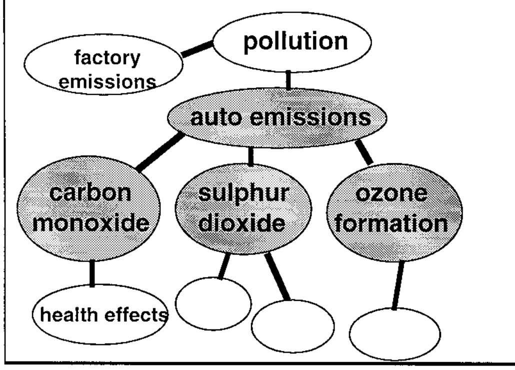
Concept map of a hypothetical source page	Example summarization of claims (hypothetical source page)
	"This page discusses air pollution, focusing on pollution produced by automobiles as a byproduct of combustion. Three dangerous pollutants produced by cars are carbon monoxide, sulfur dioxide, and chemicals that contribute to the formation of ozone. The page describes how each of these has dangerous health effects. "

Fig. 1. Concept map and exemplary identification of claims in a hypothetical Web document.

In one of the reviews that was rated as entirely peripheral, this seems to have been the case: the student identified as the claim a statement about ozone, which was the only statement about ozone in the entire document, and so was hardly central. This student "error" makes more sense, however, in light of the fact that the students group was studying ozone as part of the class project, and so from their own perspective, the claim was central. However, an effective review requires the reviewer to take an outside perspective on what is central about a resource, which these students failed to do.

Comprehensiveness. The fourth challenge in identifying claims is to give a comprehensive picture of the claims covered in the source. Students in this project did a very poor job of this: only 1 of 19 reviews was rated as comprehensively covering important claims, with 7 of 19 covering partially, and 12 of 19 covering poorly. It seems possible that the students in this project did not see this as part of the assignment, but assumed that if they picked out a few claims they would have satisfied the criteria. The only review that was rated as comprehensive identified claims was closely based on the resource page's headings. Observing this led us to suggest to students that they make more use of headings in the second project.

Evaluation of Credibility Through Examination of Scientific Evidence (Project #1)

After identifying scientific claims of web documents in project #1, students were then asked to identify what scientific evidence was given to support these claims. Two raters examined the web source documents for structured scientific evidence, and student identification of evidence is compared to rater-identified evidence in Table II.

Rating category and explanation	Examples of possible errors
Accuracy. Accuracy was judged in reference to the source page, e.g. the source page was considered authoritative and complete, in terms of this rating.	Inaccurate: "Carbon monoxide leads directly to ozone formation"
Level of detail. Each source page is assumed to have a central level of detail (often corresponding to the section headings). The correct level of detail should be comprehensively summarizable with 2-10 discrete topics.	Too broad: "This page is about pollution" Too specific: "This page is about the health effects of Carbon Monoxide"
Centrality of claims. Each source page is assumed to have a set of central topics, and possibly some peripheral topics.	Not central: "This page is about factory emissions and ozone formation"
Comprehensiveness of claims. Each source page is assumed to have a discrete number of topics (2-10) at the central level of detail.	Not comprehensive: "automobiles emit the dangerous pollutant carbon monoxide, which has the following health effects...."

Fig. 2. Description of ratings categories and examples of hypothetical errors related to each category.

Table I. Ratings of Accuracy, Level, Centrality, and Comprehensiveness of Students' Summarizations of Claims

Category	Highest rating	Medium rating	Lowest rating
Accuracy	Accurate (14)	Accurate with qualification (5)	Inaccurate (1)
Level of detail	Right level (8)	Too broad (9)	Too specific (3)
Centrality of claims	Identifies central claims (8)	Partially identifies central claims (9)	Poor identification of central claims (3)
Comprehensive claims	Comprehensive coverage of claims (1)	Partial coverage of claims (7)	Poor coverage of claims (12)

Identifying evidence turned out to be a difficult task, both for raters and for students. Raters rediscovered what previous researchers had also found (Ranney *et al.*, 1994) that what counts as evidence is sometimes highly dependent on the prior content knowledge of the reader. Experts often "see" evidence differently than novices, because they infer evidence structures that are not explicitly stated in the text. To resolved this ambiguity, our ratings scheme counts as evidence only information that is clearly structured as evidence within the text. All text documents were broken into sections roughly corresponding to paragraphs. (In the case when an evidence structure was found to span several paragraphs, paragraphs were combined into a single section.) Each document section, then, was rated on a scale of 0–3, with 0 being no evidence present, and 3 indicating that the section contained clear statements of claims followed closely by scientific data that fit Kuhn's definition of evidence. Levels 2 and 1 evidence fell short of this standard in some way. Raters were within one ratings category of each other on 92% of sections in 10 documents rated by both.

Table II shows the highest levels of evidence present in 23 reviewed documents (19 unique documents, with 4 being reviewed twice.) This table shows that 30% of reviewed documents contained at least one passage that built a well-structured level 3 argument with scientific evidence. Thirty-five percent of documents contained no argument structures that were coded as evidence. This overall dearth of well-structured evidence has implications for student reviews, as will be discussed.

As Table II shows, students seldom identified the same evidence structures that the raters identified in the source web pages. Only two of seven possible sources that raters judged as containing at least one instance of level 3 evidence were so identified by students. One of four pages with level 2 evidence were also identified by students. (This review is reprinted as an example review in Appendix B.) It should be noted, however, that even in documents with well-structured evidence, this evidence was not always easily picked out of the document, and was not always tied to the main claims that students identified in the pages. In other words, even at its best, identifying evidence was a difficult task, and students' failure to do so may be taken as more a reflection on the state of web documents than on these students' abilities. This question will be examined further in the Discussion section.

So, what did students write in the "evidence" field in lieu of reporting on well-structured scientific data? Six times, students correctly reported on external citations in documents. Although it was preferable for students to have made an analysis of evidence, citing other studies is a common proxy for directly cited evidence and is related to a document's credibility. Two other categories of what students wrote in the evidence field were identified and examined: reports of no evidence, and non-evidential information as evidence.

Seven reviews stated that no evidence for claims was given. Three of these were in agreement with the raters, who also found no identifiable evidence in those source documents. In four others, raters did

Table II. Students' Identification of Evidence in Agreement with Raters' Identification of Evidence

	None	Level one, poor evidence	Level two evidence	Level three, well-structured evidence
Quality of evidence and evidence structure ($n = 23$)	8 (35%)	4 (17%)	4 (17%)	7 (30%)
Students' matching identification of scientific evidence	—	1	1	2

identify either level 2 or level 3 evidence in the pages, not identified by these students.

At least two of the reviews that claimed "no evidence" seem not to have looked for evidence, because they went on to argue that owing to the nature of the information or the source, no evidence was needed. One student review stated that the page had no evidence, "just the facts written down. They're the EPA, they can do that." Although the EPA is an authoritative source, this is not what we were hoping to see in the way of disciplinary-based evaluations.

Three student reviews also gave non-evidential information as evidence. These cases fit with the findings of Kuhn (1991) that laypersons, when pressed to give evidence for arguments, often expand the theory they are arguing rather than giving separate substantiating evidence for the theory. An example of this from a student review about ozone listed evidence as, "ozone attacks the tissues of the throat and lungs and irritates the eyes," presumably as evidence for the claim that ground-level ozone has negative health effects. While expanding and detailing of how ozone affects health does bear on the credibility of the claim that ozone has health effects, it is not properly considered evidence for the claim. Unlike in Kuhn's studies, however, only a small fraction (3 of 24) of student reviews substituted non-evidential information for evidence.

Overall, identifying evidence in web pages was a challenge for both raters and students. There are two reasons for this that should be mentioned. First, the web is not a scientific journal, and so does not follow conventions for identifying methods, data, results, etc. in the way a scientific report would. This is bound to make domain-specific analysis of evidence more difficult. Second, information given on the web often comes in the form of "fact sheets" or encyclopedia-type entries where most of the scientific claims are noncontroversial. For example, every page on carbon monoxide (CO) contained the scientific claims that CO is a colorless, odorless gas, is a product of incomplete combustion, and is dangerous to humans because it binds to hemoglobin in the blood. None of the pages backed these claims up with scientific studies, because these facts are so well-established for specialists as to not demand it. However, for students who are encountering these facts for the first time, it would be helpful to have evidence available even for these types of claims. These challenges will have to be met in some way if the web is to be used as a forum for students to practice domain-specific evaluations of evidence.

Analysis of Student Reviews in Project #2, Infectious Diseases

Summarization of Content in Project #2

In project #2, students were requested to write abstracts of their informational web pages. Writing an abstract differs from identifying claims, in that it is less an analysis than an overview of content. Teachers intended this field to be similar to a scientific abstract. Teachers talked to students about what the function of a scientific abstract is (previewing content for interested readers) and specific requests were also made as a result of analysis of reviews in project #1. In response to the lack of comprehensiveness observed in the project #1 claims field, we asked students to be more complete in coverage of their page's topics, and also suggested that using the page's headings might be a useful way of accomplishing this.

The focus of content analysis also shifted for this project's method of summarizing content. Instead of using the four-category scheme used in project #1 to identify strengths and shortcomings (see Table I), we took the next step and attempted to formulate a prototypical form for a good abstract, based partly on findings related to the four-category analysis. This prototypical form contained three components: a thesis statement describing the overall contents of the page, description of subtopics within the page, and elaboration of some subtopics with a few key pieces of information. An example of a review that includes these three features in succession is this one: "This is a web site about the research developments in leprosy (topic sentence). It tells a little about past treatments (subtopic) with chemotherapy, dapsone, and the new treatment, MDT (specific detail.)" The thesis sentence is necessary for giving a broad overview of content, and is also a feature of considerate texts in general. Inclusion of multiple subtopics is a response to the lack of comprehensiveness observed in identification of claims, recognizing that a good review is somewhat broad in its coverage of content. Inclusion of specific detail was added after observations of abstracts that contained the first two elements, but seemed shallow, lacking in depth of content.

Abstract fields from 25 reviews were available for analysis from project #2. Table III shows inclusion of the three aspects of a prototypical review in these abstracts. Accuracy was also checked. As with the claims, most abstracts were accurate, or at least avoided outright inaccuracy. There was only one in-

Table III. Frequency of Four Aspects of Abstracts Fields from Project #2

Aspect of review abstracts (n = 25)	Count	Percent
Accurate abstract content	24/25	96
Use of topic sentence	16/25	64
Description of multiple subtopics	24/25	96
Inclusion of specific content	5/25	20

accuracy, which was the result of students having difficulty with the reporting of statistics.

Sixty-four percent (16 of 25) of reviews contained a topic sentence. Ninety-six percent (24 of 25) of student-written abstracts included mention of two or more topics in the resource. Forty-four percent (11 of 25) of these drew from the headings present on the web pages themselves, while the other 52% (13 of 25) were written using the students' own words.

The third feature of the idealized abstract was that it go beyond listing topics, to inclusion of some of the specific scientific information from sources in their reviews. Using broad criteria for what constituted specific information, only 5 of 25 reviews included any specific information from reviewed resources in their abstracts. An example of specific information is a student review that said: "This article focuses on the evidence that exists that points to HIV as the infectious agent that leads to AIDS," when they might have written a less specific "This article is about HIV and AIDS." Writing reviews with an optimal mix of detail and main topic coverage is an ongoing challenge and further scaffolding needs to be developed.

Students in project #2 were, overall, able to accurately preview the content of the web resources they were reviewing. Key challenges are to have students capture the main points of web pages, make more consistent use of topic sentences, and teach them to integrate specific facts from resources along with more high-level summaries.

Identification of Source and Bias in Project #2

In project #2, students were asked to name the sources of the documents they were reviewing, and also asked whether any potential biases existed in these documents. These are nondisciplinary-specific methods of identifying source and bias. I will examine whether students were able to correctly identify sources at the correct level, whether students identified bias, and how they justified statements of no bias.

Table IV shows the distribution of sources, as identified by the researchers, of the web documents reviewed in project #2. This table is provided here as background information, and is discussed elsewhere (Bos, in press).

Note that this table identifies sources at a particular level, that of the sponsoring organization, but other possible sources could often be named. The researcher knew from informal experience that when students are verbally asked where they got web research materials, they often give answers like "we got this from the computer," "from Yahoo (a web search engine)," or "from the Web," and cannot say any more than this about the source of their information. We were interested to see whether students misidentified the source of their reviewed information in this way. Students might possibly identify as a source an organization that was too broad, such as web service providers, search engines, or indexing services. Alternatively, students might identify as a source something too specific, such as giving the name of the author of a on-line magazine article, without giving the name of the magazine itself. Note that some web-evaluation guides (e.g., Engle, 1997; Grassian, 1995) recommend naming of particular authors as the correct level of source identification. However, naming an unknown individual authors seems less helpful for evaluating credibility than a sponsoring organization, so the latter was recommended as the correct level of source.

Table V shows how well students identified the source of web pages, and whether they identified sponsoring organizations as opposed to other levels. Analysis is restricted to those pages for which a Web source document is available, leaving 25 published reviews from two sections.

This analysis shows that students were usually able to identify a source for Web pages they reviewed. Eighty-four percent (21 of 25) made mention of the source of the document. Sixty-eight percent (17 of 25) reviews mentioned the sponsoring organization of the page, which was considered the "correct" level of identification in this analysis. Of the remaining eight reviews, four failed to identify any source, and four identified a different level of source. Case-by-case discussion of theses eight student mistakes are included within the author's dissertation work (Bos, 1998).

Identification of Bias. In the "source" field of the review form, students were also asked to identify

Table IV. Sources of Reviewed Web Documents by Category

	Gov't (U.S.)	Gov't (foreign)	University	Nonprofit organization	Commercial (nonpublishers)	Commercial information provider	Individuals
Project #2	7	4	1	6	2	3	2

possible biases related to the sources of information. Students were given little direct instructions about identifying biases, and so were mostly left to their own ideas about what constitutes a bias, and how it might be identified. Analysis focused on whether students identified bias, and if not, how they supported their claim that the source was unbiased, as shown in the lower columns of Table V.

Twenty-one of 25 reviews (88%) analyzed did mention the issue of bias in response to the prompting questions on the review form. Only two, however, said that the page they reviewed had possible bias, while the rest dismissed the possibility. Of those who claimed no bias, 10 students stated that their pages had no bias without elaborating or justifying this. Seven claimed there was no bias because of the nature of the information, stating either that the information was purely factual, or otherwise was not susceptible to distortion, e.g., ''There are no commercial or political statements included in this article, the information is purely scientific.'' Four students claimed no bias because the identity of the source did not raise suspicion, e.g., ''This article was published by the Center for Disease Control, which is a fairly reliable source. There is not a lot of potential for bias because they are a non-commercial government based organization, whose purpose in this piece is to educate the public.'' Only two reviews did admit the possibility of bias; these are discussed in another paper.

Table V. Frequencies of Five Aspects of Students' Identification of Bias

Aspects of student review of source and bias (n = 25)	Count	Percent
Failed to name any source	4/25	16
Named sponsoring organization	17/25	68
Named different level of source	4/25	16
Mentioned issue of bias	21/25	88
Identified possible bias	2/25	8
Claimed no bias, without justification	10/25	40
Claimed no bias, due to nature of information	7/25	28
Claimed no bias, due to nature of source	4/25	16

Note: Some categories in this table overlap.

Overall, students rarely identified bias in web pages they reviewed. This probably points to the need to clarify what is meant by bias. It has been suggested that instead of being asked to look for ''bias,'' a term with negative connotations, students should be asked to consider the ''point of view'' of the published source. This phrasing may better elicit analysis of the perspective and potential selection of information by the publisher that were the intention of the review of bias. The role of perspective taking in review writing, along with specific examples of students' correct and incorrect identification of bias are discussed in another paper (Bos, in press).

DISCUSSION

This research presents a design experiment exploring how K–12 students can critically evaluate scientific Web resources. Eleventh-grade students in a project-based science curriculum wrote and published reviews of Web pages as a part of two projects, one on the topic of air pollution, and the second on the topic of infectious diseases. Three research questions are addressed:

1. Can students summarize the scientific resources that they find on the web?
2. Can students identify and evaluate evidence in the scientific resources that they find on the web?
3. Can students identify the source and potential biases of the scientific resources that they find on the web?

Can Students Summarize the Scientific Resources that They Find on the Web?

Summarization is a key strategic reading skill and is also an important part of a students' ability to critically evaluate resources. In project #1, students were asked to identify scientific claims, and in project

#2 students were asked to write a scientific abstract. Analysis of claims fields in project #1 showed that students' summaries were mostly accurate, but had other shortcomings, including being at too general or too specific a level, missing central claims, and especially failing to comprehensively cover claims (see Table I). In project #2, partly in response to these observations, both the summarization format and the instructions to the students were changed. The result seemed to be that students summarized content more comprehensively, as measured by mention of multiple subtopics (see Table III). From analysis of this second round of reviews, I sought to formulate a recommended form of reviews, which was determined as having three components: use a topic sentence, description of multiple subtopics, and inclusion of some specific detail. Summaries written in the second project usually included some statement of the main topic and description of multiple subtopics. However, few reviews included specific content statements, thus presenting this aspect as a challenge for future iterations. These two projects served to illustrate some of the particular challenges of students' summarizing content in web pages, and also served as the basis for identifying a three-component model of a good web page review. The end result of these two rounds of study, then, is awareness of a few key challenges for summarization, and a tested model for abstract-writing that seems worthwhile and feasible.

Can Students Identify and Evaluate Evidence in the Scientific Resources that They Find on the Web?

In project #1, students were asked to identify evidence for scientific claims, which is similar to the way disciplinary experts in a scientific field might evaluate credibility of new claims. Analyses show that this was a very challenging task, partly because reviewed web documents often did not contain well-structured, evidence-based arguments (see Table II). Even when well-structured evidence were present, students had great difficulty identifying and describing it in their reviews. Only 2 of 7 possible examples of well-structured evidence were identified by student reviewers. It appears from this study that if the web is to be used as a setting for students to practice evidence-based reasoning, modifications will have to be made in the way students are taught to recognize evidence. Also, it seems likely that web collections themselves would need to be tailored to this task. Most of the web resources students accessed were reference sources or "fact sheets," which provided information but did not support this information with scientific argumentation. The results here may stand as a challenge for web information providers in scientific domains to provide better structured resources.

Can Students Identify the Source and Potential Biases of the Scientific Resources that They Find on the Web?

Examining student reviews from project #2, it is evident that students can usually identify the sponsoring organization of pages they use (see Table V). This positive result suggests that examining source is a viable way for students to begin judging the credibility of the resources they use. Few students, however, identified potential biases in the documents they reviewed, making this an area for improvement (see Table V). It has also been suggested, related to this study, that instead of asking students to identify "bias," which has strong negative connotations, it might be better to ask students to identify the "point of view" of the web resource. This might help students think about more subtle aspects of bias such as possible inclusion/exclusion of details, as well as more obvious possibilities for fabrication, exaggeration, or slanting of information.

CONCLUSION

This paper presented data from two design studies, exploring the idea of student-written reviews as a means of teaching critical evaluation and getting students involved in publishing contributions in the distributed hypermedia resources of the World Wide Web. Data presented here may provide grounding for further development of the technical and pedagogical scaffolding of students' use of this vast and exciting new resource.

ACKNOWLEDGMENTS

I would like to thank Professors Joseph Krajcik and Elliot Soloway for mentoring, advice, and support during this research project. Many other members of the hi-ce research group (www.hi-ce.org) also made contributions. This work was supported by NSF/DARPA/NASA Digital Library Initiative (Cooperative agreement IRI-9411287), by a grant from the NSF NIE Middle Years Digital Library (RED-9554205) and by the University of Michigan.

REFERENCES

Alexander, J., and Tate, M. (1996). Checklist for an informational Web page. Wolfgang Memorial Library: Widener University. [on-line]. http://www2.widener.edu/Wolfgram-Memorial-Library/inform.htm

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., and Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist* 26(3 & 4): 369–398.

Bos, N. D. (in press). Giving back to the Web: Social filtering of scientific World Wide Web resources in high school science. *Journal of Science Education and Technology.*

Bos, N. D. (1997). Student publishing in a digital library: value-added contributions. In E. Soloway (Symposium chair) *Using online digital resources to support sustained inquiry learning in K-12 science.* Paper presented at the annual conference of the American Educational Research Association, Chicago, Il, April. ERIC ED 408579

Bos, N. D., Krajcik, J. S., and Soloway, E. (1997). Student publishing in a digital library: goals and instructional supports. In Bell, P. (Symposium chair) *Artifact-building in computer learning environments: supporting students' scientific inquiry.* Paper presented at the Annual meeting of the American Educational Research Association, Chicago, Il, March 24–28 1997. ERIC ED 408580

Bos, N. D. (1998). *Affordances of students using the World Wide Web as a publishing medium in project-based science learning environments.* Unpublished doctoral dissertation, University of Michigan, Ann Arbor. [On-line] http://www-personal.si.umich.edu/~serp/work/

Brown, A. L. and Day, J. D. (1983). Macrorules for summarizing texts: the development of expertise. *Journal of verbal learning and verbal behaviors* 22: 1–14.

Callison, D. (1993). Expanding the evaluation role in the critical thinking curriculum. In Goodin, M. E., and McNally, M. J. (Eds.), *School Library Media Annual 11.* Libraries Unlimited, Inc., Englewood, Calif.

Cunningham, S. J. (1997). Teaching students to critically evaluate the quality of Internet research resources. *SIGCSE Bulletin* 29(2).

Engle, M. (1997). Evaluating Web sites: criteria and tools. Reference Division, Olin-Kroch-Uris Libraries, Cornell University. [On-line] http://campusgw.library.cornell.edu/cgi-bin/helper.cgi?helper=/newhelp/toc.html

Grassian, E. (1995). Thinking critically about World Wide Web resources. UCLA College Library. [On-line] http://www.library.ucla.edu/libraries/college/instruct/web/critical.htm

Kintsch, W., and van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review* 85: 363–394.

Kuhn, D.L. (1991). *The Skills of Argument.* Cambridge University Press, Cambridge.

Marchionini, G. (1995). *Information seeking in electronic environments.* Cambridge University Press, New York.

National Research Council (1995). National Science Education Standards Washington, D.C.: National Academy Press. [On-line] http://www.nap.edu/readingroom/records/0309053269.html

Palincsar, A. S. and Brown, A. L. (1984). Reciprocal teaching of comprehension–fostering and comprehension-monitoring activities. *Cognition and Instruction* 1(2): 117–175.

Pitts, J. M. (1994). *Personal understandings and mental models of information: A qualitative study of factors associated with the information-seeking and use of adolescents.* Unpublished doctoral dissertation, Florida State University.

Pressley, M., and Afflerbach, P. (1995). *Verbal Protocols of Reading.,* Lawrence Erlbaum Associates, Hillsdale, NJ.

Ranney, M., Schank, P., Hoadley, C., and Neff, J. (1994). "I know one when I see one": How (much) do hypotheses differ from evidence? In Proceedings of the 5th annual American Society for Information Science Workshop on Classification Research, 139–156.

Smith, A. G. (1997). Testing the surf: criteria for evaluating Internet information resources. *The Public Access Computer System Review* 8(3). [On-line] http://info.lib.uh.edu/pr/v8/n3/smit8n3.html

Wallace, R., and Kupperman, J. (1997). Online search in the science classroom. In E. Soloway (Symposium chair.), *Using online digital resources to support sustained inquiry learning in K-12 science classrooms.* Symposium presented at the annual meeting of the American Educational Research Association, Chicago, Il, March, 1997. ERIC ED 407252

Appendix A

PROMPTS AND REVIEW CATEGORIES FROM PROJECT #1 (AIR POLLUTION)[2]

Claim

What scientific claims are made in this resource? [blank input field]

Evidence

What is the evidence for these claim? What scientific studies were conducted? [blank input field]

Organization

- How well organized is this resource?
- Can you find answers to specific questions without searching the entire resource? [blank input field]

Appearance

- Is this an attractively designed resource?
- Does it include especially good graphics or layout? [blank input field]

[2]Only the four review categories analyzed in this paper are shown for both projects.

PROMPTS AND REVIEW CATEGORIES FROM PROJECT #2 (INFECTIOUS DISEASES)

Abstract

- What information content is available here? Summarize content in enough detail that someone doing research on a specific topic could decide whether this site would help them.
- What audience would this site be helpful for? Is it too technical for some people, or not technical enough for others?
  [blank input field]

Source

- Who is publishing this page?
- Are there any potential biases or conflicts of interest due to who is publishing this page? Are there political or commercial statements mixed in with content information?
- If there are new claims presented (rather than common knowledge), what evidence or academic citation is presented to back up these claims?
  [blank input field]

Organization of Resource

- How well-organized is the information in this page? Would you be able to find a particular piece of information if you were looking for it?
- Is there a good central page where everything is accessible? Is there a map of the site? How would you describe the organization?
  [blank input field]

Appearance

- Is this an attractively designed resource?
- Are there graphics? What kind? (photos, illustrations, flow charts, graphs of data?) Do the graphics help you understand or organize the content?
- Do the layout, graphics, and text work together to effectively communicate?
  [blank input field]

APPENDIX B: APPEARANCE AND LAYOUT OF PUBLISHED REVIEW