Science students' critical examination of scientific information related to socioscientific issues

Extended version with extra citations from students' examinations of information

Condensed version published in *Science Education* **90**:632– 655, 2006

Stein Dankert Kolstø
University of Bergen
2005

Abstract

It is widely accepted that to be scientifically literate one needs to have the ability to make thoughtful decisions about socioscientific issues (SSI). This includes critical assessment of scientific claims and arguments involved. In this study we asked eighty-nine science education students with substantial academic qualifications in science, working in groups of two and three, to assess the reliability of scientific claims in an article of their own choice, but related to a socio-scientific issue, and to present their evaluation in a short text. In analyzing the students' texts, we focused on the criteria they had explicitly and implicitly used in their evaluations. Through a qualitative analysis, we identified thirteen different criteria focusing on empirical and theoretical adequacy, completeness of presented information, social aspects, and manipulative strategies. An inspection of the students' evaluations revealed that they drew upon knowledge of possible institutional interests, different signs of competence and an appreciation of concurrent expert views, but also methodological norms in science, specialized content knowledge and an appreciation of evidence and disclosure of sources. The number of criteria used and the quality of their application varied, indicating that critical examination of texts with a science dimension need to be emphasized in science teacher education.

Introduction

The ability to examine and make thoughtful decisions on SSI is recognized as an important goal for science education OECD, 2001; AAAS, 1989). In the Nuffield Seminar initiative Beyond 2000 (Millar & Osborne, 1998) it is emphasized that science education should help people to "respond critically to [...] media reports of issues with a science dimension" (p. 12). By definition, SSI involve scientific claims and arguments, in addition to the political, personal or ethical questions of what action to choose. Moreover, in many SSI, central scientific claims are also disputed. Decision-making on such issues, therefore, involves two main questions to be considered: one political/ethical and one scientific. One example here is the issue of irradiated food. There is the political question of whether irradiation of different foods should be legal or not. In addition, there is a scientific dispute involved on whether irradiated food has lower nutritional value. This is a scientific question. A student's view of this scientific question might be crucial to his/her opinion about the political and ethical issues regarding irradiated food (Kolstø, 2004). Thus, it is important that students are able to assess scientific claims and arguments encountered in SSI. The focus for this paper is therefore on the science dimension of SSI and not on the political/ethical questions involved.

In recent years the idea of emphasizing argumentation in science education has gained momentum. In their seminal paper on argumentation in science classrooms, Driver, Newton and Osborne (2000) claim that science education should foster students' ability to "analyze arguments relating to the social application and implications of science" (p. 297). Consequently, an important question for science education is how to support the students' critical examination of claims and arguments related to the science dimension involved in SSI. The science teachers have a crucial role here. In this study we have therefore analyzed science education students' ways of examining scientific claims and arguments. Especially when working on critical examination in science teacher education it is important to have some insight into the students' ways of assessing scientific information. This is a parallel to the need to be aware of students' preconceptions when designing constructivist teaching and learning sequences in some science content.

In addition, we need more knowledge of the non-expert's criteria for critical examination in real situations, and we need to understand the knowledge base that these criteria presuppose. Such knowledge will hopefully enable us to discuss different criteria in relation to their *relevance* for science education and their *availability* for students and for laypersons in general. Research within critical thinking has discussed the relevance of different criteria and suggested general criteria and skills (Facione, 1990; Siegel, 1988). However, this research has not focused on the kinds of utterances teachers and students meet on the Internet related to controversies with a science dimension.

Aikenhead (1990) states that "to be intellectually independent is to assess, on one's own, the soundness of the justification proposed for a knowledge claim" (p. 132). To study student's evaluation of the trustworthiness of science-related claims therefore implies exploring their intellectual independence empirically. However, Hartwig (1985) claims that nonscientists "can never avoid some epistemic dependence on experts" (p. 340), and that epistemic authority does not reside within one individual, but rests with the communities of experts. Hartwig's claim is in opposition to the idea that we ought to base our judgments on empirical evidence and not on the opinion of experts or other authorities. Hartwig and other authors (Bingle & Gaskell, 1994; Norris, 1995) have questioned whether highly specialized evaluation criteria for examining empirical evidence in relation to proposed claims are accessible to non-scientists. By identifying the evaluation criteria used by the students in this study, we hope to be able to shed light on the more general question of how and to what extent non-scientists are epistemically dependent on different kinds of experts.

Related studies

In the last decade, several studies have provided insights on students' evaluation of scientific claims and arguments. Focusing on secondary school students' evaluation of the plausibility of the conclusions of news reports of science, Philips and Norris (1994; 1999) found that the students tended readily to accept claims made in brief news reports presented. In a study focusing on evidence evaluation of media reports of contemporary scientific research Ratcliffe (1999) found that the majority of secondary school students distinguished between established facts and more uncertain claims, and recognized problems of generalizations based on insufficient evidence. Some students also commented upon methodological limitations or scientist' integrity. Korpan et al. (1997) asked university students to generate requests for information as needed in order to judge the plausibility of conclusions in constructed scientific news briefs. They found students most often focus on methodological aspects and possible explanation for presented findings. Fewer requests were made about the identity and credentials of the people associated with a study. Zimmerman et al. (1998) found university students' credibility ratings of specially constructed news briefs to be influenced by the area of research, plausibility of conclusion and amount of information about research methods included in the news briefs. In contrast to these studies, the present study asked the students to choose articles appearing in the public sphere, and not articles designed by the researchers. In addition, the scientific claims in the articles typically occurred in the context of an argumentation related to a point of view on a socioscientific controversy.

Within studies of public understanding of science, there have been a number of qualitative case studies where laypersons interact with scientists or scientific knowledge in relation to e.g. SSI (see Ryder, 2001). From these studies it is often possible to identify instances where laypersons have questioned scientists' conclusions. However, these studies do not include an analysis and explicit identification of criteria just by laypersons to judge scientific claims involved in the issues. One exception here is a study on grade 11 students' views of the trustworthiness of claims involved in a local SSI. Here Kolstø (2001b) found that the students partly took the trustworthiness of science-related claims for granted and partly sought to evaluate these focusing mainly on the source of the information competence and possible interests involved.

Studies on students' argumentation structures (Driver, Newton, & Osborne, 2000; Newton, Driver & Osborne, 1999) and fallacious thinking among students (Zeidler et al., 2003; Zeidler et al., 1992) also focuses on how students handle scientific claims. This research focus is on the quality or incompleteness of students' argumentation. In the present study, however, the students are to judge the trustworthiness of a scientific claim made in Internet articles *they*, the students, have inspected for inadequacies in argumentation or other characteristics. Also, instead of scrutinizing the completeness of students' arguments about strength and weaknesses in inspected articles, we looked at their focus of attention and repertoire of operational criteria

Before presenting our findings, we will present the theoretical perspectives within which our research question and analysis is embedded. In addition, we will discuss the differentiation between examination based on scientific criteria and examination based on contextual factors. After a presentation of methodological issues, we will present and discuss the criteria identified. Finally, we will discuss our findings in relation to the idea of epistemic dependence, the relevance of the identified criteria, and possible consequences for science teacher education and science curricula for scientific literacy.

Theoretical perspectives

Critical examination and critical thinking

There is agreement within the critical thinking movement that knowledge about a topic is ordinarily a necessary, but not sufficient condition for thinking critically (Ennis, 1989). However, there is nevertheless some debate on how critical thinking should be conceptualized (Bailin et al., 1999; Bailin, 2002; Ennis, 1989; Zeidler, 1992). Here we will take issue with Bailin (2002) who concludes that critical thinking presupposes (in addition to other intellectual resources) an understanding of background knowledge and of standards and criteria which apply in the discipline or context at issue. Consequently, in order to examine scientific claims and arguments critically, one need insight into scientific criteria for judging claims and arguments, e.g. accuracy of data, reliability of sources and validity of inferences.

In this study we focus on both aspects of critical thinking. We will look for the *criteria* used by the students in their critical examinations, and we will make inferences as to what *knowledge* the students draw upon when using the different criteria. A third dimension identifying good critical thinkers is the personal commitment to rational inquiry. However, this dispositional dimension was not covered in this study.

Critical examinations and the nature of science and SSI

Constructivist view of scientific knowledge emphasizes the difference between two kinds of science (Cole, 1992; Latour, 1987; Ziman, 1968). One kind, denoted as "core science" by Cole, is characterized by a stable consensus within the scientific community. The other kind, denoted as "frontier science" by Cole, is science in the process of being researched. At this stage of the production of scientific knowledge, hypotheses are being developed and scrutinized, and results from studies are presented to colleagues and discussed (Ziman, 1968). Subjective and unreliable frontier science is transformed into core science, or refused as not reliable, through different social processes characterized by publication, evaluation and argumentation.

One consequence of this view is that the reliability of scientific knowledge claims varies from unreliable or disputed claims from the frontiers of research to highly reliable consensual core science. Another consequence is that expert disagreement is not only legitimate, but it is an important part of the production of scientific knowledge.

But the presence of frontier science and expert disagreement is not confined to scientific communities. The scientific dimension of many SSI often involves contested scientific knowledge claims. Thus, the layperson who tries to reach an opinion on a scientific question involved in an issue has to make a similar kind of examination as the expert is supposed to do; the layperson, though, has a different, and in some regards, more restricted knowledge base. And because of the urgent character of many SSI, where one's decision "not to act" has social or political consequences, waiting for science to possibly come up with a consensual answer (on, for instance, the risk associated with power transmission lines) is not an option.

In this kind of context what examination criteria might be judged adequate and legitimate for a layperson to use?

Underdetermination and the idea of scientific criteria

In the literature, the phrase "scientific criteria" is often used to characterize the kind of epistemic criteria supposed to be used by scientists when examining scientific claims (see e.g. Ziman, 2000). In contrast, the phrase "contextual factors" is sometimes used to characterize examinations based on context specific claims -- for instance, the competence or possible vested interests of scientific experts involved.

In this study, however, the analysis of the students' examinations is based on the view that this differentiation between scientific and contextual factors is indistinct, and therefore not tenable to use when characterizing the students' examinations. In the following we will present the arguments on which we base our position.

Do scientists use certain criteria to evaluate knowledge claims from the frontier of science? If so, expert disagreement indicates that such criteria are not straightforward to use. In fact, the idea of the existence of a distinct set of "scientific criteria" is disputed. A main reason for this is the claim of underdetermination of theories by data (Quine, 1961; Abd-El-Khalick, 2003). One implication of this claim is that it is not logically possible to conclude that one hypothesis is true, while another is false, just by appealing to empirical data (Quine, 1961; Kuhn, 1977; Longino, 1990). For instance, even if data are consistent with a hypothesis, the claim of consistency is based on theories guiding interpretation of the data and the work of instruments. But if scientific criteria have to extend the appeal to empirical data, what are scientists appealing to when justifying a choice of theory?

Studying situations in the history of science, Longino and others have identified instances where they claim that the content of the science produced has been influenced by contextual values (Longino, 1983; Longino, 1990; Nelson & Nelson, 1997). However, this does not imply that contextual values *always* influence the content of scientific knowledge. It could be the case, as some philosophers of science have argued, that they "only" have identified instances of "bad science," i.e. instances where researchers have deviated from the norms of science.

Longino (1983) has suggested the use of the concepts "constitutive values" and "contextual values" to clarify the issue. She defines "the values governing scientific practice *constitutive values* to indicate that they are the source of the rules determining what constitutes acceptable scientific practice or scientific method" (p. 8). As examples of constitutive values she mentions truth, accuracy, precision, simplicity, breath of scope, and problem-solving capacity. The value denoted as "accuracy" is the one typically appealed to when the quality of methodology and empirical underpinning is examined. She defines contextual values as values that "belong to the social and cultural context in which science is done" (p. 8). The idea of contextual values, she argues, stems from "the beliefs that scientific practice ought to be independent of personal, social, and cultural values" (p. 7). These definitions make it possible to talk about the role of values in science and to discuss the legitimacy of the influence of different values, including contextual values, on scientific knowledge production.

Longino (1990) has also argued that "the relation between hypothesis and evidence is determined by background assumptions operative in the context in which data are being assessed" (p. 58). The choice of background assumptions is in general not possible to do only by appealing to the constitutive values mentioned above. Thus, she claims that choice of theory in science in principle might be influenced by contextual values.

The concepts of underdetermination and background assumptions have the following consequences. If interpretations of empirical data, in some or all instances, are partly based on contextual values, it is necessary to use contextual values to judge these interpretations. Moreover, to the extent that background assumptions are needed, who is to decide which background assumptions are legitimate to use? Longino has proposed e.g. "applicability to human needs." In principle it is possible for other scientists and for laypersons to apply other value-based background assumptions. Consequently, in our analysis we have to be open for instances where students use contextual values to critically evaluate background assumptions in studies.

Values in Post-normal science

Research seeking to inform political decision-making is often characterized by facing high system complexities. Issues related to health and the environment, like BSE (mad cow disease) and the climate issue, are typical examples. In many of these issues researchers are struggling with high uncertainties, due to the complexity of the phenomenon and its context. In addition, when highly important societal decisions are involved the role of values becomes very evident. Funtowicz and Ravetz (1993) claim that in these situations the researchers have to make decisions influenced by contextual values in order to produce knowledge relevant for political decision-making. Important science-related questions with a value dimension are, for instance, as follows: How can one simplify a problem and still produce policy-relevant information? Is the available or foreseeable scientific information in this case of a high enough quality to include it into the policy process at all? How precautionary do we wish to be?

The point here is *not* that the science involved is "purely subjective" or arbitrary, but that the complexity of the system makes it impossible to encompass the whole system from one perspective. In addition, the strength of the science-society interactions makes the role of prioritization apparent.

Funtowith and Ravetz (1993) have argued that in issues where both the system uncertainties and the decision stakes are high, an extended peer community is needed to perform quality-assurance of the whole scientific and political process. Funtowich and Ravetz use the term "post-normal science" in situations where the peer community is extended beyond the community of scientists. The possible role of laypersons in extended peer communities implies the need to discuss the legitimacy of different values and perspectives in the evaluation of scientific knowledge claims. If scientists are using contextual values in decisions related to their research, with or without an extended peer community involved, these decisions are legitimate to evaluate using alternative contextual values. Again the consequence is that we have to include in our analysis instances where students use contextual values to critically evaluate assumptions and decisions made during a study.

The role of trust in science

Trust is a basic entity in science. Hartwig (1991) states that "scientific propositions often must be accepted on the basis of evidence that only others have" (p. 706). A reviewer usually has to trust the data presented in an article; researchers in a team usually have to trust each testimony, and a researcher who wants to build on the results of others has to trust the judgment of those scientists who have reviewed and accepted these results. In addition Hartwig (1991) argues that “[r]eplication of experiments is not standard. Moreover, peer review can never detect plausible and internally consistent fabrication” (p. 706). He proposes the epistemological principle “the principle of testimony” which he claims is applicable both between experts and between laypersons and experts:

> If A has good reasons to believe that B has good reasons to believe p, then A has good reasons to believe p.” (p. 697)

He argues that A’s good reasons depend on whether B is truthful, competent and conscientious. It might be the case, for instance, that when engaged in a debate on a proposed new natural phenomenon, a scientist’s examination of claims and arguments put forward also involves an evaluation of an arguer’s competence and the possible influence of vested interests. And even if the peer review system is guided by the constitutive values, the single reviewer might also have additional interests related to career and affiliation that might have an impact on his/her examination in spite of his/her high ideals (Ziman, 2000). It is therefore not possible to draw a sharp line between the evaluation of scientific knowledge claims and the evaluation of the trustworthiness of the researchers involved in the production or dissemination of scientific results. If the personal trustworthiness of the researchers is fundamental in science, it is our conclusion that it is not tenable to dismiss the use of contextual information when evaluating scientific claims. Moreover, the inclusion of contextual criteria for judging contextual information on researchers’ trustworthiness becomes a necessity. This point has also been made by Bingle and Gaskell (1994), who based their opinion on their case study of the different possible ways to examine the controversy about the Galileo mission to Jupiter.

It is also important to be aware that laypersons seldom conduct thorough examinations of science, or have access to empirical data and details of the methods used (not to mention checking the correctness of the empirical data). Thus, the issue of trustworthiness of the researchers (and the journalists) and their ways of interpreting and presenting results becomes paramount.

The immediate consequence of the above discussions for our analysis of the students’ examination is that we will focus on their use of different scientific criteria as well as other kinds of criteria. We will also avoid differing between “content focusing examinations” and examination of sources of information when discussing our findings, as all criteria identified in fact aims at evaluating the content of a claim. Thus, we will discuss the relevance of different criteria identified without using the division between scientific and non-scientific criteria as a normative distinction.

Method

Data collection

A total of eighty-nine students, from teacher education departments in two different universities in Norway, participated in this study. All students attended courses in science education included in their teacher education. The courses involved were taught by some of the authors. The students were in their final year of study: Nearly half of the students held a master’s degree in science, and the rest held the Norwegian equivalent of a bachelor’s degree in science.

As a compulsory part of a teaching and learning unit on scientific literacy in their courses, the students had to write a few short texts based on work related to the inclusion of SSI in science teaching. The students worked in groups (n=31) of two and three in front of computers. They were asked to search the Internet for articles related to a socioscientific issue, pick one of interest, and perform a critical examination of the science dimension of that article. These activities comprised the first part of the unit, and, therefore, were not based on ideas proposed in lectures. However, two papers were made available for the students during the activity: "Questioning Authority" (Rampton & Stauber, 2001) and "Criteria for Assessing the Quality of Health Information on the Internet" (Rippen, 1999). Both papers suggest a variety of general criteria for assessing trustworthiness. The first emphasizes vested interests and "propaganda" (non-argumentative strategies used to try to convince people). The second, which focused on Internet sites providing health information, emphasized credibility (source, review process), content (accurate and complete), and presence of links in addition to site architecture and presence of disclosure. We did not ask the students, following the activity, whether they read the two articles or not. However, being present in the room where the students were working, we very seldom saw students looking into these papers. Importantly, the analysis showed that the students also used a range of criteria not included in the two papers.

This paper is based on texts written by the students, prompted by the following task:

> After providing a short summary of the article chosen, write a text where you evaluate the information and argumentation included in the article. Focus in particular on the trustworthiness of science-related claims that are put forward.

All the thirty-one groups of students responded to this task. It is important to be aware that the task specifically asks the students to focus on the science dimension of the chosen article. This probably explains why our student rarely focused on the moral and emotive aspects of the issues, as reported in other studies (Bell & Lederman, 2003; Sadler & Zeidler, 2005; Sadler & Zeidler, 2004a). However, three of the texts did not focus on the science dimension of the issue chosen, and these were therefore excluded from the sample. In addition, the student groups had to write a text where they commented on another groups' evaluation. These texts also included evaluation of the information of the articles chosen (e.g. what they thought the group should have seen or looked for in the article evaluated) and, subsequently, were included. The basis for this study, therefore, is the analysis of sixty texts: twenty-eight evaluations and thirty-two comments (these added up to more then twenty-eight as some groups commented on more then one evaluation). The texts are between a half and a full written page. These, and other texts written by the groups, were published by the students on the Internet in a kind of "newspaper" designed for this use.

The student groups chose articles for examination from different kinds of sources and related to a range of different issues. The articles chosen focused on a range of issues, namely health effects of food, alcohol or snuff (7), effects of reducing diets and of alternative medicines (5), radiation and health effects (3), global warming (4), pollution and biodiversity (6), whether cloning, traveling in time, reintroduction after captivity is possible (3). The different sources where the students found the articles they examined were the following: sites including sale or promotion of certain products (5); daily newspapers (7); university-based Internet magazines (3); internet magazines (commercial and published by school authorities) (3); United Nations and governmental bodies (3); environmental organizations (3) and sites informing about the views of an institution (4).

Analysis

The students' texts were analyzed using qualitative methods, as we wanted to gain insight into the kind of ideas they made use of when performing critical examinations in this particular context. The texts, and also a copy of the examined articles (for reference), were imported into the software Atlas.ti, which is designed to support the analysis of qualitative data.

The analysis consisted of two distinct phases. In the first phase, the analysis was performed on the basis of a tentative conceptual framework, where we tried to differentiate between examination of *content* and *sources* of information. Confronted with the students' texts, we realized that we needed a more elaborate theoretical framework in order to be able to characterize different criteria in a tenable manner, and performed an intensive literature review. Based on different criteria suggested in the literature (Bingle & Gaskell, 1994; Hardwig, 1985; Kuhn, 1977; Lehrer, 1977; Longino, 1983; Longino, 1997; McMullin, 1982; Norris, 1995, Rampton & Stauber, 2001; Rippen, 1999; Siegel, 1988; Ziman, 2000), the discussion presented under Theoretical perspectives above, and on preliminary empirical findings, we articulated a set of tentative categories for the main part of the analysis. The first category, quality of argumentation, refers to the use of empirical data and research findings in the argumentation toward a conclusion. Thus, all kinds of methodological criteria belong to this category. The category is based on Longino's (1997) constitutive value "empirical adequacy." The second category, compatibility with accepted theory, is inspired by the observation that the students sometimes compared a claim with their own understanding and also by Kuhn's (1977) criterion "consistency to related theories" (p. 323) currently held as true. In the section on findings, these two categories are conflated into one main category named empirical and theoretical adequacy. The third category, completeness of presentation, is based on the preliminary finding that the students often commented on the lack of arguments and references in the examined articles. This idea is echoed in Ziman's (2000) reminder that "in accordance with the norm of criticism, it [a research claim] has to be presented in a form capable of undergoing further communal tests" (p. 85).

The fourth category "social aspects of the sources of claims" is a broad one. It might, for instance, include qualifications or competence, the experts' honesty and possible conflict of interests (Siegel, 1988) and role of funding and gender (Bingle & Gaskell, 1994). Finally, social aspects might include "social information," a concept based on Lehrer's (1977) suggestion that "we must find a method for aggregating the opinions of individual experts" (p. 477). Through the continuing analysis, a fifth category focusing on manipulative use of language (Rampton and Stauber, 2001) was also identified.

Through the subsequent analysis we sought to identify more specifically what the students focused on in their examinations and other characteristics of the criteria they use. Using "the constant comparative method" (Strauss & Corbin, 1990), segments of students' texts were coded and tentatively sorted into groups. Through an alternation between "discovery mode" and "verification mode" (Guba, 1978), codes and categories were adjusted until they stabilized (Merriam, 1998). In the analysis we did not examine the "correctness" of the students' use of a criterion in the given context, but focused only on what idea or criterion the students used.

The validity of the identified criteria rests on the existence of at least one student examination that exemplifies the category. The completeness of the set of criteria rests on the absence of student examinations that do not fit into any of the identified categories. The coding was done by the first author. In order to evaluate the validity and the completeness of the categories developed, we invited a researcher outside the original project group to perform an inter-coder reliability control. The outside researcher read the theoretical section of the draft version of this paper. In addition the coding of four texts where inspected to clarify the definitions of the codes developed. Counting the number of co-occurrences of coding (of the remaining 56 texts), we obtained an inter-coder reliability of 69 percent (number of agreement vs. number of agreements and disagreements). Mismatches were partly due to difficulties in interpreting the students' texts regarding ideas involved, and partly due to misplaced codes. Through negotiation we reached agreement on 93 percent of the coded text units.

We believe the methodology used in this study enhances the inner validity (Merriam, 1998) of the findings, through the use of authentic SSI and free choice of issue and article for examination. Even if our results is constrained to a specific kind of context, it is possible to use them as a “working hypothesis” (Cronbach, 1975) when discussing teaching and learning of critical examination. However, in order to evaluate the outer validity of the findings, supplementing studies in contexts with other characteristics is needed. In the following, all quotes from students’ texts are translated from Norwegian by the authors.

Findings

Criteria focusing on empirical and theoretical adequacy

The analysis revealed that it was possible to organize the students’ evaluations regarding empirical and theoretical adequacy into the four different sub-categories listed in table 1.

Criteria focusing on empirical and theoretical adequacy	**Instances**	**Groups**
Quality of references	8	7
Consistency of argumentation	22	13
Face validity of argumentation	19	12
Compatibility with subject knowledge	7	4
In total	**56**	

Table 1: Criteria identified through the analysis that focuses on empirical and theoretical adequacy.

Quality of references

We found that several student groups commented on the quality of the sources and references given in examined articles. When explaining such evaluations, they used phrases like “relevant references,” “prominent scientific journals,” or “trustworthy and respected sources,” e.g.:

> Gr. 14: Because the article [...] makes references to a great many concrete scientific articles, we consider as a starting point the article to be trustworthy.

An interesting aspect here is how students, probably through their university studies, have socialized into an appreciation of scientific articles as highly reliable. The idea used by the students here seems to be that the trustworthiness of a claim is related to the quality or trustworthiness of the sources of the claim. In fact, this idea might be used for differentiating between claims based on peer reviewed scientific research and claims based on other kinds of evidence or beliefs. However, in order to use this criterion, the students also need knowledge to differentiate between characteristics for scientific journals and publishers, and differences between referred papers and grey literature, reports, newspaper articles and so forth.

In one instance an evaluation of the quality of references resulted in weakened credibility to be the conclusion:

> Gr. 20: [...] references to regulations and discharge requirements are made in vague terms. This detracts the text’s credibility somewhat.

Here it seems as if the students interpret the vagueness as providing room for the author to make a more subjective interpretation and presentation of the documentation than an exact representation would have allowed for. As one basic idea in science is to diminish the subjective element in data and interpretations (Ziman, 2000), this idea is in accordance with norms associated with scientific ideals. Awareness of this norm or idea, as tacit or explicit knowledge, is a precondition for using this criterion.

Consistency of argumentation

In a number of texts the students commented upon the logical correctness and consistency of an argument presented. Typically they examined whether the conclusion was legitimate in light of the referred evidence. Referring to the claim that a reduced ozone layer implies increased cancer risk, one group comments:

> Gr. 21: The article says nothing about the possibility that the increasing number of instances of cancer might have other causes, like for instance more frequent sunbathing.

Another group comments on the claim that the whale Keiko, which has been in captivity for many years, should be set free:

> Gr. 8: Scientific precedents, which document successful reintroduction of marine mammals, do not yet exist.

Again the students are criticizing the conclusion for going beyond available evidence. The application of the basic idea here was also found to result in increased trust when the conclusion did *not* go beyond the presented evidence. Referring to the claim made by The Norwegian Pollution Control Authority that dioxin can cause cancer, one group states:

> Gr. 9: As mentioned, we find the article very sober. We could for instance read that "Compared with many other environmental pollutants, we do have much knowledge about dioxins and the effects of these, but occurrences and sources are still not completely surveyed." One has been cautious with not being opinionated, and one gets the impression that all facts put forward are thoroughly supported by evidence. Because the page is so sober, it is much easier to accept the claims that actually are presented.

The idea used in these examples is that evidence is needed, and that the conclusion needs to follow logically from presented evidence. This criterion involves an appreciation of evidence and consistency, making it an example of application of the value "empirical accuracy." An interesting aspect of this category is that several groups were able to claim the presence of inconsistencies, which is a powerful rhetorical device. In addition, the criterion is obviously useful for identifying claims that are insufficiently backed by evidence. It can be inferred that, when using this criterion, the students draw on knowledge about inferences, complexities of variables and generalizing across contexts, and also the difference between various kinds of relationships, such as correlation and causality. One might ask, however, whether a master's degree or training in scientific methodology is needed to manage the use of this criterion, or whether it is within the reach of most people.

Face validity of argumentation

This category also focuses on the quality of arguments presented, but in contrast to the former, it does not involve any focus on consistency. In evaluations belonging to this category the students have pointed to a single piece of information or the general impression an argument gave, leading to the identification of these evaluations as based on the face validity of presented arguments or conclusions. One example is the use of exact figures:

> Gr. 18: The factuality of the text seems to be great, due to the numerical statements and evaluations, [...].

The main claim in the examined article, as referred in the group's text, was whether the use of nuclear power plants instead of plants using coal can reduce the global warming. The presence of exact figures is obviously regarded as positive. One possibility is to interpret this idea in light of Ziman's (2000) claim that one main strategy used in science to reduce the subjective element is quantification.

In other examples, the students point to the use of "a scientific method," or that it is claimed that "several methods" have yielded the same conclusion. In addition some groups stated that the argument "sounds reasonable," or that "detailed explanations" or "professional argumentation" was, or was not, present. The following example, related to the issue of whether homeopathy is backed by scientific evidence, shows how the presence of professional argumentation was valued:

> Gr. 13: However, the author of the article uses a distinct language where he makes use of scientific argumentation in relation to the scientific methodology used to judge homeopathy.

The idea the students seem to apply here is that certain "signs" indicate quality or trustworthiness of evidence, arguments or conclusions (Kolsto, 2001b). All these signs or indicators seem to resemble methodological norms common in science. However, we found no indications in the texts that the students had bothered to make a more thorough control of methods used. It might be the case that the necessary details enabling such thoroughness have not been easily accessible. The lack of details obviously restrains the possibility of doing thorough examinations based on knowledge of methodological norms in science. Probably these are also different for various disciplines, making it a demanding set of criteria to learn and to manage. However, the idea of trying to identify instances where the authors tried to diminish the subjective element, or to obtain space for more subjective interpretations, is a criterion worth closer inspection.

Compatibility with own theoretical knowledge

This category is based on examples where the students used their knowledge and understanding of scientific concepts and theories to judge the trustworthiness of arguments and claims. The following example shows a thoughtful examination of a claim found in a newspaper article. The article reports from a meeting where an American professor claimed that gene technology might result in super organisms that might outmaneuver everything else:

> Gr. 23: Many have put forward this claim as one of the most risky. How stable new genes are in the new organism is difficult to predict. It is proved that naturally occurring bacteria can transfer genetic material from one plant to another. This way you can not make a guarantee saying that a gene installed in one plant to increase its resistance against disease, herbicides or something of the sort can not be transferred to weed. In this way we might attain a super weed that will be difficult to conquer. Moreover, the gene for antibiotic is often used as a marker in such a modification process, even if this practice to a large extent is ended. These might in a similar way end up in bacteria that are harmful to us. This will be exceptionally dangerous, as it will not be possible to conquer with antibiotics.

The students here use their highly specialized knowledge in gene technology to evaluate the claim stated. Implicit in the evaluation is the idea that scientific claims put forward have to be compatible with current scientific understanding of the topic. This is in accordance with Kuhn's (1977) constitutive value "external consistence." The thoroughness of the students' use of this criterion varied however. Some students only hinted that this kind of evaluation had been performed:

> Gr. 15: The whole feature article has the mark of being technically correct and [...] a trustworthy plea.

The use of this idea of course presupposes relevant knowledge at a relatively high level. Actually, in this study only four of twenty-eight groups used this criterion, even if half of them held a master's degree in science. Moreover, it is not obvious that it is valid to use textbook knowledge from one's science education as a yardstick for judging the correctness of claims from the frontiers of science. The critical examiner might, for instance, be unaware of important details or new discoveries that could alter the judgment. However, an exception needs to be done here for people who really engage with an issue and seek deeper knowledge making them a kind of expert, like the farmer Mark Purday did in the BSE controversy

(Ravetz, 1997). Confronted with the variety of issues before us, this criterion is probably better suited to judge the competence of journalists and other non-scientists who are debating or referring to scientific claims.

Criteria focusing on completeness of information

In several of their examinations, the student groups focused on the lack or presence of factual information or arguments in the article. These examinations were grouped into the three categories listed in table 2.

Criteria focusing on completeness of information	**Instances**	**Groups**
Completeness of references	23	14
Completeness of an argument	21	12
One-sidedness in the presentation	14	10
In total	**58**	

Table 2: Criteria identified through the analysis that focuses on completeness of information presented in the examined article.

Completeness of references

The students often complained that an article lacked references for claims that were presented, implicitly indicating that this prevented any further examination of trustworthiness of a claim (and of course makes an evaluation of the quality of references impossible). E.g.:

> Gr. 28: We cannot find any direct references to these [claimed research findings], and this is a weakness of this page. [...] We want more links that could make it easier for us to check facts and claims.

Based on their awareness of the importance of references, the students here use "wanting references" as a criterion for evaluating trustworthiness. This criterion might prove relevant for differentiating between documented claims and mere guesses or claims only referred from others. An interesting aspect is that they applied this idea when examining texts from a variety of "non-scientific" sources. This criterion is easy to use, and an awareness of the need for reference to sources and further documentation should be within reach for most students.

Completeness of an argument

Students also complained that an argument, or an article, wanted explanations, methodological details, technical information, reflections or argumentation in general. As an example, one group commented on an article from an environmental organization arguing that incineration is preferable to the use of waste disposal sites. They made the following comment:

> Gr. 19: A discharge reduction of dioxin at 99.7% with new cleaning technology is unquestionably very good, but [the environmental organization] does not state any reasons for why 0.027g/MT [gram per megaton] incinerated waste is an acceptable value.

The students seems to use an idea saying that a claim is less trustworthy if the argument is not spelled out in enough details to be evaluated. This idea is in accordance with the explicitly valuation of claims or findings in science "to be presented in a form capable of undergoing further communal tests" (Ziman, 2000, p. 85) and "accompanied by an account of how and why it was obtained" (Ziman, 2000, p. 93). An interesting aspect of this criterion is that it does not evaluate quality per se, but only presence of information. Its use might, therefore, result in reduced trust in an article, but does not attack the claim itself. The criterion might nevertheless be judged as relevant for inclusion in science teaching, as it emphasizes the importance of requiring from authors the disclosure of arguments for further inspection.

One-sidedness in the presentation

In addition to complaining about wanting details in arguments, ten groups also commented explicitly on the lack or presence of counter-arguments in an article. Thus, these comments were not related to the information needed to evaluate an argument, but to the students' general impression of one-sidedness or exclusion of counter arguments:

> Gr. 27: The article has a one-sided focus. [The author] writes a great deal about advantages related to the use of nutritional supplement, but mentions little about advantages of changing eating habits [an alternative action].

> Gr. 4: We think that he sums up the different possibilities and consequences concerning cloning. Therefore the article makes us more confident to take a personal stand for or against cloning.

The idea implicit in this criterion is that the presence of both pro and con arguments in an article is needed to make it trustworthy. A somewhat weaker interpretation is that you cannot trust articles, which only present arguments in favor of the author's view, to tell the whole story. The use of this criterion probably stems from the knowledge or experience that all issues can be seen from more than one point of view. Based on this interpretation, the criterion is both important and within reach of most people.

Criteria focusing on social aspects

The students also used information about social aspects of sources of claims and arguments in their examinations. The five different categories representing criteria focusing on these aspects are listed in table 3.

Criteria focusing on social aspects	**Instances**	**Groups**
Possible underlying interest	23	12
Personal value-related qualities	3	3
Author's or expert's competence	27	15
Level of professional recognition	5	4
Level of expert agreement	16	11
In total	**74**	

Table 3: Criteria identified through the analysis that focuses on social aspects of sources of information.

Possible underlying interest

The students sometimes called attention to possible interests that could have influenced a study, findings, evaluations or presentation. An important difference between this category and the one related to *one-sidedness in the presentation* is that here the students not only looked at the argumentative balance in the text, but made inferences from what they read to what might have caused a possible bias, e.g.:

> Gr. 23: It should not be concealed that the specialized competence that the gene technologists possess can be used as a "smoke screen" in order to secure their own work. Especially in relation to the fact that gene technologists within the industry in general need to get their products and ideas sold, in order to make a living of their expertise.

The kind of interests the students are considering is related to economy and workplace (as in this example), professional prestige and position for experts, institutional interests, and loyalty to friends. In addition, one group commented that the information from The Norwegian Radiation Protection Authority concerning radiation from cell phones was "characterized by trying to avoid public concern." One group also showed awareness of possible cultural influences by stating that

> Gr. 28: [s]cientific communities are often predisposed by accepted theories and attitudes in their surroundings, and a strong skepticism towards everything that is not scientifically proved.

The idea the students use in this kind of evaluation seems to be that a claim or an argument is less trustworthy if it seems like certain interests might have colored its content. The impression from most of the evaluations concerning interests is that these are based on the presence of clues in the texts, for instance that a brand name and logo is used. However, in the citations included, the students appear to have some knowledge about institutional characteristics which exceeds the idea of commercial interests, and which enables them to take a more critical stand. This echoes the finding in a study of sixteen-year-old students who seemed to have too little knowledge about institutional aspects of modern science to be able to use the idea of possible influencing interests adequately (Kolstø, 2001b). The relevance of interests as a criterion for examining scientists' evaluations is convincingly demonstrated by Bingle and Gaskell (1994).

Personal value-related qualities

In the study by Kolstø (2001b) mentioned above, it was also found that some students used the idea that experts who seemed to share their values, e.g. emphasizing the precautionary principle, were more trustworthy. Also, scholars have pointed to the relevance of examining an author's or expert's personal qualities or attitudes, like conscientiousness (Norris, 1995) and honesty (Siegel, 1988). In this study we found that two groups expressed the idea that when authors showed a "critical attitude" towards "both sides" in an issue, this enhanced the trustworthiness in the article. Commenting on an article about the validity of "Homeopathy versus Medicine," one group wrote:

> Gr. 13: The critical attitude towards both homeopathy and scholastic medicine is one of the things which contribute to our impression of seriousness.

The rare use of criteria such as integrity, conscientiousness, honesty and critical attitude in this study could be due to the students' shallow knowledge about the authors from only reading one article. However, the inspection of the personal qualities of an expert or a debater should be regarded as important in critical examination in relation to controversies. Whether an ability to consider personal value-related qualities is suitable as a goal for science education is probably a more debatable issue.

Author's or cited expert's competence

Not surprisingly, we often found the students to evaluate the competence of the author or cited expert. What is more interesting, however, is the kind of information the students used to judge competence. When judging an article to be trustworthy, the students rather frequently pointed to relevant education (master's degree or PhD) and current occupation. Examining an article about cloning, and one about the claim that it is healthy to drink two glasses of red wine every day, the students gave the following comments respectively:

> Gr. 4: The cell biologist [name] has a professional background that makes us confident about the content of the article.

> Gr. 24: The article emerges as trustworthy as it is medical practitioner [name] that states this. He is director at the [name] [medical] clinic and therefore he has some credibility.

An interesting aspect of the idea of using competence as a criterion is that the students' evaluations are, with few exceptions, based on trust in the information given in the article.

In addition to education and occupation, some groups used the place of publication to judge trustworthiness. The trustworthiness of newspaper articles and journalists' competence was questioned by two groups. Trust was attributed when the article was published in a reputable scientific journal, at a Web site from a governmental authority, or (in one instance) from an environmental organization. E.g.:

> Gr. 22: The content of the article is reliable. The journal where it is published inspires trust, because not just anyone is allowed to get published in [it]. As a non-expert one has to be able to trust articles in [it].

An interesting aspect of this last citation is that the journal is a kind of popular magazine presenting research done at a Norwegian university for a wider public, and not a peer-reviewed research journal. This exemplifies that, in order to be able to use the place of publishing as a criterion, some knowledge of different journals, publishers and organizations are necessary.

In general the students seemed to put great confidence in the criterion "competence." They typically used it to make conclusive comments on trustworthiness, and not only to indicate a possible conclusion. The relevance of qualifications as a criterion is indisputable and is also suggested by Siegel (1988). However, the idea of facts as indisputable and researchers as neutral and objective is not tenable anymore. Therefore, the criterion should be used in combination with other criteria.

Level of professional recognition

We also found that the students examined competence by evaluating an expert's level of professional recognition. Commenting on information about health effects of a commercial product called noni juice, published on a Web site with relations to dealers of this juice, one group commented:

> Gr. 25: It is made a reference to Dr. [name] as an acknowledged biochemist, but he has only published works about noni. It therefore emerges as improbable that he really is an acknowledged biochemist, and not a biochemist working on a contract for noni juice.

The idea here is that professional recognition signals quality or competence, and when this is lacking the expert's claims are less trustworthy. However, we found only a few instances where this kind of criterion was used, and it was typically found in texts where the students commented upon each other's critical examinations. Nevertheless, one of the criteria for judging experts suggested by Norris (1995) is "the role and weight of prestige in the scientific community" (p. 216). Using this criterion presupposes knowledge of different experts' prestige in science, or strategies to get this kind of information. The criterion might be used in situations characterized by the presence of dilettantes with a personal hypothesis, but is less adequate in the common situations with competent experts on "both sides."

Level of expert agreement

The final criterion identified in this category is examination of *level of expert agreement*. In contrast to the criterion *level of professional recognition*, which focuses on the *expert's* standing within the scientific community, the criterion *level of expert agreement* is about the standing of the science-related *claim*. The criterion echoes Lehrer's (1977) suggestion of including aggregated social information about the opinions of individual experts when evaluating a scientific theory. Based on the students' evaluations in this study, the category is defined also to include, on the one hand, the idea that a claim is more trustworthy if supported by one or several additional researchers, and, on the other hand, the idea that consensual or "core science" is reliable. The following citations illustrate how the idea that presence/absence of expert agreement respectively increases/decreases a claim's trustworthiness, is used by the students (italics added):

> Gr. 17: The article makes use of *established knowledge* about positive and negative effects of alcohol.

> Gr. 8: According to [the researcher], this whale [Keiko] is not suited for a life outside captivity, and costs very much money. This corresponds to *other researchers' utterances and experiences*.

> Gr. 12: Many of the pages that deal with the blood type diet are based on the same book. [...] *Professionals* with knowledge about this topic *have criticized the research* in this book for not being up to standard.

The idea of using the role and weight of consensus in science to judge experts has also been suggested by other authors (Kolstø, 2001a; Norris, 1995). The criterion presupposes knowledge about, and appreciation of, the role and importance of critique, argumentation and consensus in science. It is a demanding criterion to use, as it presupposes that the non-scientist in a decision-making position takes notice of and aggregates different experts' evaluations and mutual criticisms.

Criteria focusing on manipulative strategies

In a few examinations the student groups focused on the author's choice of words, arguments and pictures, and examined whether such choices was done to deliberately manipulate readers. In the six instances identified, the students claimed their trust in the author, expert or claim was weakened due to appeal to emotions or values not relevant for the author's claim. Commenting on the mentioned article about research related to "noni juice," found on a Web site financed by companies with commercial interests in noni juice, one of the groups stated:

> Gr. 25: The history of the noni fruit is used to show that it is a "natural" and therefore healthy product. They are again appealing to popular feelings and to the popular opinion that everything that is old is good.

The implicit idea seems to be that a claim is less trustworthy if the author chooses to use a language that deliberately appeals to emotions. The idea of examining rhetorical use of words and phrases we found in four of the sixty texts, which we find surprisingly infrequent. Probably, this criterion is discussed and used more frequently and varied in everyday contexts than the students in this study did. According to Rampton and Stauber (2001), awareness of emotional use of language is important when confronted with authors with hidden agendas. The ability to detect language appealing to emotions and values is probably also within reach of most people.

Number of examinations identified per group

As a group the students used a range of criteria in their examinations, and a total of 194 instances of examinations were identified through our analysis. However, the number of critical remarks per student group varied. Figure 1 shows that some groups only made one or a few critical remarks, while other groups made more then ten.

Did some groups make few critical remarks because their article lacked information enabling the application examination criteria? Figure 1 indicates that this possibility is not a sufficient explanation. In our analysis we also identified criteria focusing on completeness of information (Com in figure 1). If we inspect on the half of the groups were we identified fewest instances of critical examinations, the following pattern emerges: seven groups did not comment on lacking information in the article inspected, and four groups only made one such comment. Consequently, several groups which used few criteria either did not recognize incompleteness of information, judged the information as sufficient, or laced awareness of completeness of information as a criterion.

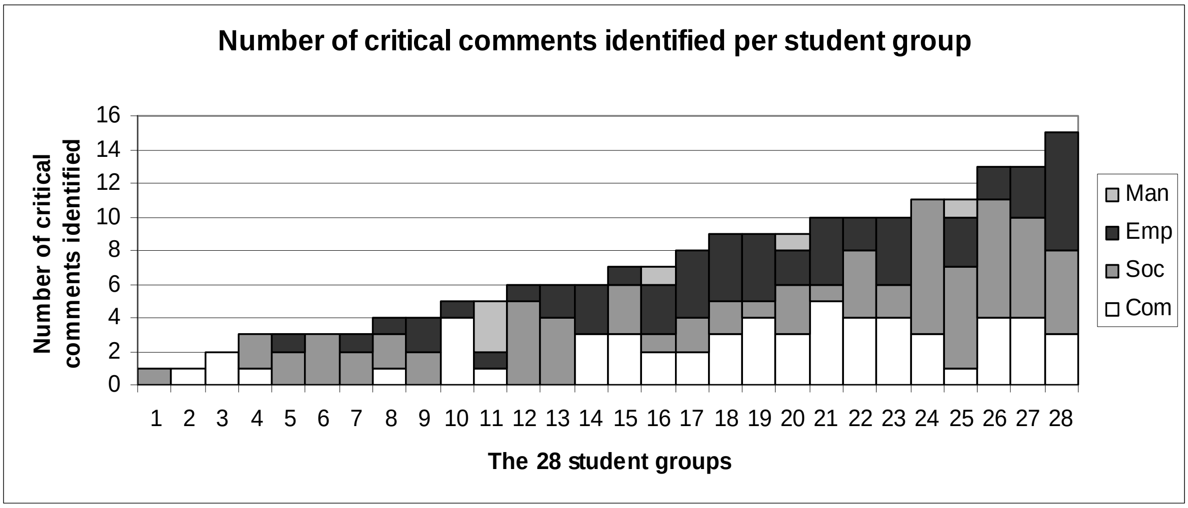

Figure 1: The number of critical comments made for the different student groups and for different types of criteria. In the figure Man refer to manipulative strategies, Emp to empirical and theoretical accuracy, Soc to social aspects and Com to completeness of information.

Discussion

In this article we have presented and discussed a group of students' critical examination of the science dimension of articles related to SSI of their own choice. We found that the students focused both on empirical and theoretical adequacy, on the completeness of the presented information, on social aspects and on manipulative strategies. Examination criteria identified as related to empirical and theoretical adequacy were quality of references, consistency of argumentation, face validity of argumentation and compatibility with own theoretical knowledge. The criteria related to completeness of information were completeness of references, completeness of argumentation and one-sidedness in presentation (e.g. lacking presentation of counter arguments). The criteria found to be related to social aspects were author's or cited expert's competence, professional recognition, possible underlying interest, author's or expert's personal qualities and level of expert agreement. Finally, the criterion manipulative strategies regarding appeals to emotions was identified.

It is important to be aware that the identification of these criteria was done in a specific context. If the context had not been educational, if the students had selected other articles, and if students with other experiences and knowledge had been involved, other or additional criteria might have been found. However, the inner validity of the criteria identified is not diminished by this lack of completeness and outer validity. They "existed out there" as the criteria are based on actual examinations made by students in the context of SSI.

Also, we have not made any systematic normative analysis in order to judge the relevance of the applications of the identified criteria. However, through our reading of examined articles and students' text, we did not recognize any examinations as strictly misplaced, even if e.g. those categorized as face validity of argumentation where rather superficial.

Our findings imply that the students involved in this study did not use certain criteria discussed in the literature. We did not find instances where students explicitly or implicitly made appeals to values suggested as constitutive for science other than empirical adequacy and external consistency/theoretical adequacy. The values "internal consistency" (within a theory), "simplicity," "breath of scope" and "fruitfulness" suggested by Kuhn (1977) were thus not found, and neither was the values "novelty," "ontological heterogeneity," "applicability to current human needs" or "diffusion of power" suggested by Longino (1997) from a feminist perspective. Regarding possible underlying interest, the role of gender (Bingle & Gaskell, 1994) was not found to be touched upon. Regarding competence the role and weight of successful competition for research grants (Norris, 1995) was not used. Charisma (Bingle & Gaskell, 1994) as a criterion was only hinted at, through the comment that at first glance an expert seemed to be "very sympathetic." In addition, experts' conscientiousness (Norris, 1995) was never explicitly evaluated. Based on this study it is not possible to state whether the students did not find application for these criteria in the examined articles, whether they did not value them, or whether the students lacked awareness of the ideas expressed in these criteria.

In a Taiwan study Yang (2004) found that students expressing uncertainty on their thoughts about a SSI tended to attribute their uncertainty to insufficient information. This finding is echoed in this study. A main finding is the students' explicit critique of the information deficiency in the examined articles, making a more thorough examination difficult. This includes sources and references, evidence and details of arguments, methodological issues and also information about social aspects. The identified criteria or evaluation strategies, therefore, have to be interpreted as indicating how the students tried to make critical examinations in spite of shortage of information. Nevertheless, only two out of twenty-eight groups of students searched for additional information, in order to make a more thorough examination.

Another characteristic of the students' examinations was that they seldom made concluding statements regarding credibility; the only exception was when they used competence as a criterion. This special emphasis on competence could be due to their background as science students, and a belief that professional competence implies neutrality, objectivity and honesty.

We also found that thirteen of the twenty-eight groups expressed clearly whether they had trust in the examined article or main scientific claim, and twelve groups indicated in vague terms a decision on trustworthiness. Probably a decision about trustworthiness sometimes evolves by comparing the outcome of examinations based on several criteria, as the following extract indicates:

> Gr.29: It is made a reference to Dr. [name] as an acknowledged biochemist, but he has only published works about noni. It therefore emerges as improbable that he really is an acknowledged biochemist, and not a biochemist working on contract for noni juice. Down the list [of references] there are other researchers and works mentioned, and there is probably a need for being somewhat critical to some of this, as the first mentioned researcher seems to be paid by noni juice. Some of this research has been on TAHITIAN NONI® Juice, and not on any noni extract. This indicates that the producer of noni pays for this research, and this might therefore influence the results achieved. Explicit reference information like journal and volume is missing from all entries on the list of scientific works provided.

This group examines both quality of references, professional recognition and possible underlying interest. In addition the group is using the criteria completeness of references, consistency of argumentation and appeals to emotions in their text. They conclude that the implicit claim that noni juice is healthier than other juices and fruits is not convincing.

Epistemic dependence and scientific criteria

Do our findings indicate that laypersons are epistemically dependent on experts (Hardwig, 1985; Norris, 1995), or is it possible to state that these students were intellectually independent of the experts whose claims were examined? On the one hand, the students did at least try to assess the soundness of justifications proposed for knowledge claims, virtually fulfilling Aikenhead's (1990) definition of intellectual independence referred in the introduction. But on the other hand, the students' examinations of e.g. quality of references and competence build extensively on information given in the articles. This illustrates that in the given context the students chose to trust what was presented, instead of cross-checking all pieces of information. And the students never tried to examine the relation between empirical evidence and proclaimed findings, probably because this would have presupposed an examination of research reports and not e.g. newspaper articles. Within the context of newspaper and Internet articles, some groups did manage to examine claims in light of presented evidence, the possible influence of different social aspects, and the general quality of articles based on the completeness of the presented information. However, the quality of the examinations would have improved substantially if information had been crosschecked, using e.g. the possibility to search the Internet for further evidence, counter arguments, antagonists' evaluation of arguments, or extensive institutional information. But again, the context was educational and not a personal commitment to examine a controversy. The students' texts where not marked as the activity was part of a learning process. Within the three lessons allocated for the writing of the texts (plus voluntarily in between the lessons), the students had to make their prioritizations. The idea of crosschecking was not discussed with the students before afterwards. These factors might explain why evidence of crosschecking was not found in the texts in spite of the fact that the students worked on-line.

The question of epistemic dependence is related to the question raised by several authors, of whether the criteria scientists use in examining claims are accessible for laypersons (Bingle & Gaskell, 1994; Hardwig, 1985; Norris, 1995). In this study some groups of students did manage to criticize the relevance and strength of presented evidence in relation to claims stated, and to examine whether scientific claims were compatible with their own understanding of core science. Whether this kind of examination is within reach of non-majors in science depends probably as much on the content of their science education, and their engagement with the examined issue, as on their cognitive abilities.

The relevance of the different criteria identified

The students involved in this study held a bachelor's or a master's degree in one of the natural sciences. They where confronted with complex issues involving political/ethical and scientific aspects in interplay. Are the criteria the science education students used in their critical examination acceptable for inclusion in science teaching?

For the purpose of comparison, we inspected information for peer reviewers published on the Internet by the well-known journals Nature and Science, and by the American Institute of Physics, American Geophysical Union and Institute of Physics. We found a range of criteria emphasized. Among these were technical failings, adequately established main claim and whether the study was set in the context of previous work, echoing the criteria focusing on empirical and theoretical adequacy found in this study. The criteria for peer reviewers also emphasized the availability of data sets and that all relevant works were cited, echoing some of the criteria focusing on characteristics of presentation found in this study. Finally, the criteria for peer review emphasized (in addition to other criteria) disclosure of affiliations and funding, which echoes the criteria related to competence and interests. Certainly several of these criteria exceed empirical and theoretical adequacy.

In our discussion of theoretical perspectives, we claimed that a strict differentiation between content-focused criteria and contextual factors might not be fruitful, let alone adequate, due to the strength of the science-society interactions in SSI. This is also an argument for the relevance and acceptability of all criteria found in this study, including criteria related so social aspects. Zeidler et al. (1992) identifies as "fallacies of relevance" attacks on a claimers' credibility and appeals to popularity (but not explicitly level of agreement among experts). It is our view that this identification is too general, and that evaluation of the credibility of a person is relevant when instrumentation, measurements and so forth are not available for inspection. It should be evident from the examples provided that all examinations made by the students were done in order to assess the trustworthiness of the *content* of the claims considered. This also implies that a division between content-focusing and author- and context-focusing criteria is problematic. Without judging the students' application of the different criteria, it is our conclusion is that the criteria the students used are legitimate as such, even if several of the criteria could have been used more thoroughly by considering additional sources and information. Norris (1995) has claimed that

> [...] the proper attitude for non-experts to have toward scientific experts is not to feel a need to judge the claims that they [the experts] make in their area of expertise, but, rather, to be disposed to judge the grounds for their claim to expertise before willingly consenting to recognize that expertise. (p. 213)

But both this study and Ryder's (2001) analysis of case studies on laypersons' interaction with SSI indicate the possibility that, to some degree, non-experts are able to examine the empirical adequacy of scientific claims. Therefore, our answer to the question of whether science education should focus on scientific criteria or on awareness of contextual factors is that both are relevant and legitimate.

Implications for science teacher education

A main goal for this study was to explore what kinds of critical examinations are possible when facing articles of own choice but related to SSI. The teacher education students involved had graduated or hold a master degree in science. The criteria identified indicate the kind of critical examinations science education students, with no special course in critical thinking, might be able to teach. Our results show that the science education students involved indeed were capable of performing critical examinations of such articles using a broad range of criteria. However, a closer inspection reveals several challenges for science teacher education.

Firstly, the quality of the students' evaluations varied considerably. While some student groups carefully evaluated the authors' competence or the correctness of scientific claims made, others' evaluations was more superficial. One example here is a group who wrote that "the article claims that the higher consumption of alcohol in the population, the more addicts, and this seems logical." However, they never scrutinized this claim any deeper, and the logic here is definitely not watertight, but needs empirical underpinning. Another group stated that the article was written by a person holding a doctoral degree and therefore the author knew what he was writing about. However, the claim to competence was from the article itself, and the students did obviously not know in what discipline the author hold his degree.

Secondly, also the number of criteria used differed considerably between the groups. As figure 1 show, several groups made use of few criteria. An inspection of the texts written by the different groups revealed that the length of the texts where more then four criteria was found was on the average of the same length as the texts where less then four criteria were identified used. Nevertheless, seven of these groups did not criticize on the completeness of the information provided in the article examined. Our conclusion is that this study indicates that critical examination, including criteria and knowledge enabling the use of these, needs to be taught explicitly in science teacher education. This claim is supported by a Delphi-study on critical thinking (Facione, 1990) which states that within any course, critical thinking should be taught *explicitly*. An explicit focus might bring more science teachers up to a proper level for teaching critical examination in the science classroom. Hopefully results from this study can provide thought provoking examples and criteria to discuss and reflect upon in courses for science teachers.

In this study we also found that students judged it relevant to use criteria focusing on social aspects of science and scientists. Application of these criteria presupposes knowledge of institutional aspects of science. This indicates a need for more explicit focus on institutional and social aspects of science in science teacher education in order to make this kind of criteria accessible for more science teachers. In general, the broad range of criteria used by the students in this study indicates that science teacher educators need to discuss whether a broader range of criteria, extending the focus on empirical and theoretical accuracy, should be regarded as relevant for inclusion in science teacher education

An important finding in this study is the students' critical comments on incompleteness of information in examined articles. However, we did not find many signs in the students' texts of strategies to cope with this situation. In the information society there are several possibilities for examining claims and arguments using multiple sources of information. An emphasis on such strategies seems to be important for future science teacher education.

Implications for science curricula

The findings in this study revealed that the students draw upon different kinds of knowledge and ideas in their examinations. An evaluation of what kind of knowledge is prerequisite for critical examination of different kinds of texts is important when designing science curricula for scientific literacy. This study indicates that for critical examination of newspaper and Internet articles related to SSI, scientific content knowledge is not sufficient.

In order to evaluate the empirical and theoretical adequacy of claims, the students made use of knowledge about different sorts of sources of scientific information, including research journals, and knowledge about methodological norms in science, e.g. regarding inferences and generalizing across contexts. Some groups also used their understanding of content knowledge to examine scientific claims. Understanding of relevant content knowledge was also found by (Sadler & Zeidler, 2004b) to correlate positively to the quality of informal reasoning regarding SSI.

When evaluating completeness of presentation, the students used their awareness of the need for references to sources and further documentation, and the idea that arguments need to be presented with enough details to make critical examination possible. In addition, they used the idea that all issues can be seen from more than one point of view. When focusing on social aspects, the students used their knowledge of possible interests of institutions providing scientific information, and also an appreciation of a source's critical attitude. In addition, they used their knowledge of how to recognize competence and an expert's prestige in science, and their awareness of the role and importance of consensus in science. Thus, the students draw upon their knowledge of general methodological norms in science, science content, social processes in science, and institutional aspects of science.

These findings support studies (Sadler, 2004; Kolstø, 2001a; Ryder, 2001) which conclude that insights into the nature of science are relevant for examination of the science dimension of SSI. Other studies have indicated that the correlation between a decision-makers' understanding of NOS and decisionmaking on SSI is weak or missing (Bell, 2003; Bell & Lederman, 2003; Zeidler et al., 2002). However, these studies have focused on the moral or political decsisionmaking, while the present study has focused solely on students' examination of scientific claims and arguments.

Our findings on the relevance of a range of insights about science are in stark contrast to at least the profile of Norwegian curricula for integrated science, which emphasis content knowledge. However, when the different topics are introduced, the kind of reasons stated for its inclusion is that it will provide a foundation for students' assessment of information, e.g.:

> The students will acquire some knowledge of molecular biology, development biology and genetics, to make them able to understand and examine information about the development of gene technology and biotechnology and be able to make decisions on some of the ethical issues here. (KUF 1993, p. 7)

The implicit model for scientific literacy seems to be that critical examination is based on understanding of content knowledge. However, in this study we found that students with university courses in science also make use of a range of criteria not related to content knowledge. We will maintain that if scientific literacy is to include critical examination of

SSI, an extended model of the relation between prerequisite knowledge and the scope of critical examination is needed. Science curricula for scientific literacy need to include knowledge about methodological norms, social processes and institutional aspects of science. This conclusion is supported by Ryder's (2001) analysis of a number of qualitative case studies of laypersons' interactions with science and scientists. He concludes that knowledge of the nature of science and scientific knowledge are at least as important as scientific content knowledge for laypersons' assessment of science-based arguments.

Acknowledgement

This study was a part of the project "Danning, informasjonsvurdering og argumentering i naturvitenskap" (Liberal education, information assessment and argumentation in science). The project was supported by a grant from the Norwegian Ministry of Education via the PLUTO project.

References

AAAS (American Association for the Advancement of Science). (1989). *Science for All Americans. Project 2061*. New York: Oxford University Press.

Abd-El-Khalick, F. (2003). Socioscientific issues in pre-college science classroom. In D. L. Zeidler (Ed.), *The Role of Moral Reasoning on Socioscientific Issues in Science Education*. Dordrecht: Kluwer Academic Publishers.

Aikenhead, G. (1990). Scientific/technological literacy, critical reasoning, and classroom practice. In S. P. Norris & P. L. M. (Eds.), *Foundations of Literacy Policy in Canada*. Calgary: Detselig.

Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Conceptualizing critical thinking. *Journal of Curriculum Studies*, 31(3), 285-302.

Bailin, S. (2002). Critical Thinking and Science Education. *Science & Education*, 11, 361-375.

Bell, R. L. (2003). Exploring the role of nature of science understandings in decision-making. In D. L. Zeidler (Ed.), *The Role of Moral Reasoning on Socioscientific Issues in Science Education* (pp. 63-79). Dordrecht: Kluwer Academic Publishers.

Bell, R. L., & Lederman, N. (2003). Understandings of the nature of science and decision making on science and technology based Issues. *Science Education, 87*, 352-377.

Bingle, W. H., & Gaskell, P. J. (1994). Scientific literacy for decisionmaking and the social construction of scientific knowledge. *Science Education*, 72(2), 185-201.

Cole, S. (1992). *Making Science. Between Nature and Society*. Cambridge, Massachusetts: Harvard University Press.

Cronbach, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, 30, 116-127.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287-312.

Ennis, R. H. (1989). Critical thinking and subject specificity: clarification and needed research. *Educational Researcher*, 18(3), 4-10.

Facione, P. A. (1990). Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction: Research Findings and Recommendations (The Delphi Report): Prepared for the Committee on Pre-College Philosophy of the American Philosophical Association. ERIC ED 315 423.

Funtowicz, S. O., & Ravetz, J. R. (1993). Science for the post-normal age. *Futures*, 25(7), 739-755.

Gayford, C. (1993). Discussion-based group work related to environmental issues in science classes with 15-year-old pupils in England. International Journal of Science Education, 15(5), 521-529.

Geddis, A. N. (1991). Improving the quality of science classroom discourse on controversial issues. Science Education, 75(2), 169-183.

Guba, E. G. (1978). *Toward a Methodology of Naturalistic Inquiry in Educational Evaluation* (Vol. 8). Los Angeles: University of California.

Hardwig, J. (1985). Epistemic dependence. *The Journal of Philosophy*, 82(7), 335-349.

Hardwig, J. (1991). The role of trust in knowledge. *The Journal of Philosophy*, 88(12), 693-708.

Jasanoff, S. (Ed.) (2004). States of Knowledge: The Co-production of Science and the Social Order. London: Routledge.

Kolstø, S. D. (2001a). Scientific literacy for citizenship: tools for dealing with the science dimension of controversial socio-scientific issues. *Science Education, 85*(3), 291-310.

Kolstø, S. D. (2001b). "To trust or not to trust, ..." - pupils' ways of judging information encountered in a socio-scientific issue. *International Journal of Science Education, 23*(9), 877-901.

Kolstø, S. D. (2004). Students' argumentations: Knowledge, values and decisions. In E. K. Henriksen & M. Ødegaard (Eds.), *Naturfagenes didaktikk - en disiplin i forandring? Det 7. nordiske forskersymposiet om undervisning i naturfag i skolen* (pp. 63-78). Kristiansand: Høyskoleforlaget AS.

Korpan, C., A., Bisanz, G., L., Bisanz, J., & Henderson, J., M. (1997). Assessing literacy in science: evaluation of scientific news briefs. *Science Education*, 81(5), 515-532.

Kortland, K. (1996). Decision-making on science-related social issues: the case of garbage in physical science - a problem-posing approach. In G. Welford & J. Osborne & P. Scott (Eds.), Research in Science Education in Europe. Current Issues and Themes (pp. 115-124). London: Falmer Press.

Kuhn, T. (1977). *The Essential Tension*. Chicago: University of Chicago Press.

KUF. (1993). *Læreplan for videregående opplæring. Naturfag* (Curriculum for upper secondary. Integrated science). Oslo: Kirke-, undervisning- og forskningsdepartementet (Ministry of Education and Research).

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society*. Milton Keynes: Open University Press.

Lehrer, K. (1977). Social information. *Monist*, LX(4), 473-487.

Longino, H. (1983). Beyond "bad science": Sceptical reflections on the value-freedom of scientific inquiry. *Science, Technology and Human Values*, 8(1), 7-17.

Longino, H. E. (1990). *Science as Social Knowledge. Values and Objectivity in Scientific Inquiry*. Princeton: Princeton University Press.

Longino, H. E. (1997). Cognitive and Non-Cognitive Values in Science: Rethinking the Dichotomy. In L. H. Nelson & J. Nelson (Eds.), *Feminism, Science and the Philosophy of Science* (pp. 39-58). Dordrecht: Kluwer Academic Publishers.

McMullin, E. (1982). Values in science. In P. D. Asquith & T. Nickles (Eds.), *Proceedings of the Philosophy of Science Association 2*. East Lansing: Philosophy of Science Association.

Merriam, S. B. (1998). *Qualitative Research and Case Study Applications in Education.* San Francisco: Jossey-Bass.

Millar, R., & Osborne, J. (1998). *Beyond 2000: Science Education for the Future*. London: Nuffield Seminar Series: Interim Report V3.

Nelson, L. H., & Nelson, J. (Eds.). (1997). *Feminism, Science and the Philosophy of Science*. Dordrecht: Kluwer Academic Publishers.

Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. *International Journal of Science Education*, 21(5), 553-576.

Norris, S. P. (1995). Learning to live with scientific expertise: toward a theory of intellectual communalism for guiding science teaching. *Science Education*, 79(2), 201-217.

Norris, S. P., & Philips, L. M. (1994). Interpreting pragmatic meaning when reading popular reports of science. *Journal of Research in Science Teaching*, 31(9), 947-967.

OECD. (2001). *Knowledge and Skills for Life. First Results From PISA 2000.* Organisation for economic co-operation and development.

Phillips, L. M., & Norris, S. P. (1999). Interpreting popular reports of science: what happens when the reader's world meets the world on paper? *International Journal of Science Education*, 21(3), 317-327.

Quine, W. V. O. (1961). *From a Logical Point of View*. New York: Harper Torchbooks.

Rampton, S., & Stauber, J. (2001). *Trust us, we're Experts - How Industry Manipulates Science and Gambles with your Future*. New York: Penguin Putnam. Chapter 11: Questioning Authority.

Ratcliffe, M. (1996). Pupil decision-making about socio-scientific issues, within the science curriculum. International Journal of Science Education, 19(2), 167-182.
Ratcliffe, M. (1999). Evaluation of abilities in interpreting media reports of scientific research. *International Journal of Science Education*, 21(10), 1085-1099.
Ravetz, J. R. (1997). Simple scientific truths and uncertain policy realities: implications for science education. *Studies in Science Education*, 30, 5-18.
Rippen, H. (1999). Criteria for assessing the quality of health information on the Internet, [World Wide Web]. Health Summit Working Group [2003, 01.07]. Available: hitiweb.mitretek.org/docs/policy.html
Ryder, J. (2001). Identifying science understanding for functional scientific literacy. *Studies in Science Education*, 35, 1-44.
Sadler, T. D. (2004). Student conceptualisation of the nature of science in response to a socioscientific issue. *International Journal of Science Education, 26*(4), 387-409.
Sadler, T. D., & Zeidler, D. L. (2004a). The morality of socioscientific issues: Construal and resolution of genetic engeneering dilemmas. *Science Education, 88*(1), 4-27.
Sadler, T. D., & Zeidler, D. L. (2004b). The significance of content knowledge for informal reasoning regarding socioscientific issues: Applying genetics knowledge to genetic engeneering issues. *Science Education, 89*(1), 71-93.
Sadler, T. D., & Zeidler, D. L. (2005). Patterns of informal reasoning in the context of socioscientific decision making. *Journal of Research in Science Teaching, 42*(1), 112-138.
Siegel, H. (1988). Rationality and epistemic dependence. *Educational Philosophy and Theory*, 20, 1-6.
Simonneaux, L. (2001). Role-play or debate to promote students' argumentation and justification on an issue in animal transgenesis. International Journal of Science Education, 23(9), 903-928.
Strauss, A. L., & Corbin, J. (1990). *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*. Newbury Park, California: Sage.
Wertheimer, N., & Leeper, E. (1979). Electrical wiring configuration and childhood-cancer. American Journal of Epidemiology, 109(3), 273-284.
Yang, F.-Y. (2004). Exploring high school students' use of theory and evidence in an everyday context: the role of scientific thinking in environmental science decision-making. *International Journal of Science Education*, 26(11), 1345-1364.
Zeidler, D. L., Osborne, J., Erduran, S., Simon, S., & Monk, M. (2003). The role of argument during discourse about socioscientific issues. In D. L. Zeidler (Ed.), *The Role of Moral Reasoning on Socioscientific Issues in Science Education*. Dordrecht: Kluwer Academic Publishers.
Zeidler, D. L., Walker, K. A., Ackett, W. A., & Simmons, M. L. (2002). Tangled up in views: Beliefs in the nature of science and responses to socioscientific dilemmas. *Science Education, 86,* 343-367.
Zeidler, D. L., Lederman, N. G., & Taylor, S. C. (1992). Fallacies and student discource: conceptualizing the role of critical thinking in science education. *Science Education*, 76(4), 437-450.
Ziman, J. (1968). *Public Knowledge: An Essay Concerning the Social Dimensions of Science*. Cambridge: Cambridge university Press.
Ziman, J. (2000). *Real Science. What it is, and What it means*. Cambridge: Cambridge University Press.
Zimmerman, C., Bisanz, G. L., & Bisanz, J. (1998). Everyday scientific literacy: Do students use information about the social context and methods of research to evaluate news briefs about science? *The Alberta Journal of Educational Research*, 44(2), 188-207.