THE UNIVERSITY OF ALBERTA

SCIENTIFIC LITERACY:
EVALUATION OF MEDIA REPORTS OF SCIENTIFIC RESEARCH

BY

CONNIE KORPAN

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

DEPARTMENT OF PSYCHOLOGY

EDMONTON, ALBERTA

FALL, 1994

ISBN 0-315-95060-9

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Connie Korpan

TITLE OF THESIS: Scientific Literacy: Evaluation of Media Reports of Scientific Research

DEGREE FOR WHICH THESIS WAS PRESENTED: Master of Science

YEAR THIS DEGREE GRANTED: Fall, 1994

(SIGNED) ____________

PERMANENT ADDRESS:

#210, 10634-80 Ave

Edmonton, Alberta

T6E 1V5

DATED ____________

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled SCIENTIFIC LITERACY: EVALUATION OF MEDIA REPORTS OF SCIENTIFIC RESEARCH submitted by CONNIE KORPAN in partial fulfillment of the requirements for the degree of MASTER OF SCIENCE IN PSYCHOLOGY.

....

supervisor

....

....

Date..21 June 1994..................

Abstract

One aspect of scientific literacy was investigated, how university students evaluate the plausibility of media reports of scientific research. Evaluation was assessed by asking students to read four news briefs in various domains and to generate a list of questions they would need answered to determine whether each research conclusion was true. They were also asked to justify each question. Of particular interest was whether the frequency and nature of students' questions and justifications were related to their educational background in science. Fourth-year Psychology majors were more likely than first-year students and fourth-year English majors to evaluate issues concerning design, measures, statistics and boundary conditions, indicating that majoring in Psychology may sensitize students to such issues when evaluating media reports of research. Fourth-year students in either major were more likely than first-year students to evaluate the social context of the reported research, indicating that such knowledge may be acquired through general university experience or through extra-curricular experiences with science. Performance also varied with the research domain described in the news briefs. Future implications in terms of assessment of scientific literacy are discussed.

Acknowledgements

I would like to express my gratitude to Dr. Gay Bisanz for your guidance throughout my program. Your unfailing support and patience has been greatly appreciated. Also, thanks to Drs. Jeffrey Bisanz and J.P. Das for your valuable feedback on my thesis. Finally, I wish to thank Dr. Terry Taerum for your advice regarding the statistical analysis used in this study.

Table of Contents

List of Tables

Scientific Literacy:
Evaluation of Media Reports of Scientific Research

The launching of Sputnik I in 1957 marked the end of a long period of stability in the school science curriculum (Hodson, 1985). This event led to the concern that the USA and Canada were falling behind in the international science-technology race and that our political and economic security would be jeopardized if the school system failed to produce well-trained scientists (Mullis & Jenkins, 1988; Shortland, 1988). Many different projects for curriculum reform were developed, but a common feature among many of them was a shift away from teaching science as an established and complete body of knowledge towards teaching science as a human activity (Howe, 1971). The implication of such curricular reform was to teach children how to think and act like scientists. The goals were to make children aware of what scientists do and to encourage children to pursue science at an advanced level (Hodson, 1988).

Since the end of the 1960's, the goals of science curriculum planners have changed slightly (DeBoer, 1991). Preparing interested students for a career in science remains a priority. However, it became increasingly evident that all students need science literacy skills to function responsibly and effectively in the world (DeBoer, 1991; Mullis & Jenkins, 1988). Furthermore, in 1984, the Science Council of Canada found that many children and adults were not attaining adequate levels of scientific literacy. The 1986 NAEP Report concerning science achievement in the United States described similar findings. These findings set the stage for another call for reform in the science curriculum; a curriculum was needed that would produce scientifically literate citizens.

There are many components to scientific literacy. These components include content knowledge (i.e., concepts and conceptual networks) in various scientific disciplines, an understanding of the nature of scientific knowledge, an understanding of the nature of scientific methodology and reasoning, and an understanding of the social context of science (Aikenhead 1987, 1988; Deboer, 1991). A scientifically literate person is one who can use this knowledge to cope with the personal, social and environmental consequences of the advancements in scientific research (Snow, 1987). A scientifically literate person can also use such knowledge for individual decision-making (e.g., choosing medical treatment, evaluating product safety) and for decision-making that has societal and environmental implications, such as supporting medical research and purchasing environmentally-safe products (National Science Teachers Association Position Statement, 1982).

In the current study, I investigated one aspect of scientific literacy that has received little attention, how individuals evaluate the plausibility of media reports of scientific research. In particular, I investigated the kinds of questions that university students want answered when evaluating the truth of research conclusions, similar to those reported in the media. The media has a strong impact on what people believe about science (Aikenhead, 1988). How individuals react to these media reports may be influenced by their ability to ask questions that help them to distinguish good from poor research (Koelsche, 1965; Smith, 1974). Therefore, in the assessment of scientific literacy, it is important to determine the knowledge that enables students to evaluate scientific research as it is presented in the popular media (Deboer, 1991; Koelsche, 1965).

In this study, the students who participated differed in terms of type of university training (i.e., non-majors vs. Psychology majors vs. English majors) and amount of university training (i.e., first-year vs. fourth-year students). Such differences in university training were expected to affect the degree to which students are sensitized to various scientific issues. Because the students may have had comparable levels of STS (Science-Technology-Society) training at the high school level, however, they were expected to have at least rudimentary understanding of many scientific issues. The main objective of an STS-based curriculum is to teach students about *authentic science*. Instruction of authentic science entails instruction on the *actual nature* of science, rather than the *idealized* version of science (i.e., the way science *ought to be*) (Martin, Kass, & Brouwer, 1990). Knowledge of the actual nature of science is critical to science literacy because this knowledge may affect how one evaluates scientific reports. It should be emphasized at this point that the STS curriculum is only *recommended* to teachers. National and provincial curriculum objectives are not necessarily reflected in the science classroom. Therefore, it is possible that a number of students who participated in the current study have not been exposed to the STS curriculum.

In the following discussion, I review some components of an STS curriculum. For each component discussed, I also describe the nature of science training at the university-level, if it is different from secondary-level instruction. University training does not always elaborate upon secondary-level instruction in science. I also present research on how science instruction influences students' understanding of some components of science. Finally, because a new coding scheme was developed to categorize student responses in the current study, I briefly discuss issues that surround the assessment of scientific literacy. I also describe a pilot study that was conducted to develop and refine the coding scheme.

STS Curriculum and University Science Training

Over the last few decades, science and technology have increasingly pervaded society and society has influenced the direction of scientific and technological research (Martin et al., 1990). This situation has motivated science curriculum planners to increase scientific literacy by designing an STS curriculum (Aikenhead, 1987, 1988). An STS curriculum may include the following components: content knowledge across various scientific disciplines; the nature of scientific knowledge; the nature of scientific theories and models, scientific methodology and scientific reasoning; and the social aspects of science. This list is not exhaustive nor are these components mutually exclusive. I have chosen to discuss the components that I feel knowledgeable students would use when evaluating media reports of scientific research. For expository purposes, I have chosen to discuss each component separately, but it should be made clear that issues relevant to one component are influenced by issues that are relevant to the remaining components.

Content Knowledge/Nature of Knowledge

Teaching students content knowledge (i.e., facts) across various scientific domains is not new to the science curriculum (Deboer, 1991). In the discipline-oriented curriculum that dominated science instruction in the 1950's to 1980's, it was recognized that people need a knowledge base in which to incorporate new knowledge (Posner, Strike, Hewson, & Gertzog, 1982). Teaching students content knowledge is still a curricular goal. However, under the STS curriculum, it is recognized that scientifically literate individuals must possess more than scientific knowledge; they must also understand the *nature* of scientific knowledge. In other words, they understand the limitations of scientific knowledge (e.g., limitations concerning collection and interpretation of data). The implications of teaching students the limitations of scientific knowledge is that they should evaluate this knowledge rather than accept it at face value.

Under the STS curriculum, one important characteristic of scientific knowledge that may be stressed is that it is tentative (Aikenhead, 1987, 1988; Brower & Singh, 1983; Mullis & Jenkins, 1988; Ryan & Aikenhead, 1992). That is, students may be taught that scientific knowledge is a product of scientific investigations that *currently* bears the seal of approval from the scientific community (Shortland, 1988). If the scientific community deems a research finding to be valid, it is temporarily considered to be scientific knowledge, but may be replaced or elaborated as a result of other investigations (Aikenhead, 1987; Fleming, 1988). Assuming that most students in this study received instruction on the nature of scientific knowledge, it was expected that most students would inquire about the scientific community in which it is generated.

At the university level, the tentativeness of scientific knowledge is often stressed, especially in science courses that cover the history of the discipline. For example, students who receive instruction on the history of psychology are often made aware of various paradigm shifts that have occurred in different psychological domains (e.g., from behavioral to cognitive frameworks). Furthermore, students who receive instruction in contemporary psychology are often made aware of the various debates that have occurred within different research areas (e.g., reading: whole language vs. phonics) and how these debates may be resolved via consensus. Assuming that the Psychology majors in this study received this university training, it was expected that they would be more likely than first year students and English majors to inquire about consensus among the scientific community when evaluating the research described in the news briefs.

The Nature of Scientific Theories

Before the implementation of the STS curriculum, students were left with the impression that scientific theories are fixed, unalterable truths (Hodson, 1988). After science literacy was deemed a curricular goal, teachers were encouraged to tread a very narrow line when teaching students about scientific theories. They were to attempt to teach currently accepted theories but not so strictly that students could not on occasion evaluate and reject the theories and generate or consider alternative theories (Hodson, 1988). Teaching students about the true nature of scientific theories and models may help motivate students to evaluate scientific theories. Thus, in the current study, it was expected that most students would ask questions about the theories behind the research described in the news briefs. Furthermore, assuming that the three groups of students have similar knowledge of or experience with the domains discussed in the news briefs, no group differences were expected with regard to how often students would evaluate the associated theories.

Nature of Scientific Methodology

Teaching students about scientific methodology is not unique to the STS curriculum. The process-oriented curriculum introduced in the late 1950s emphasized the five/seven step methodology purported to be used by scientists (Deboer, 1991; Ryan & Aikenhead, 1992). However, under the STS curriculum, students may learn about the *nature* of methodology including (a) there is not *one definitive method* but a range of methods available to scientists (Martin et al., 1990; Medawar, 1984), (b) scientists will adopt the methods they consider most appropriate to the domain being investigated (Hodson, 1985), (c) scientists modify and develop methodology as scientific knowledge and technology change, (d) scientists may use flawed methodology that make resultant data and conclusions questionable, and (e) scientists who do conduct their research in a careful and thoughtful manner still do not have access to *absolute truth*

(Mullis & Jenkins, 1988).

At the university level, students who pursue a science-oriented degree are typically exposed to rather sophisticated aspects of research methodology, either through methodology courses or by reading the primary research literature. The type of methodology they are exposed to depends on whether they major in a deterministic science (e.g., chemistry, physics) or probabilistic science (e.g, psychology) (Lehman, Lempert, & Nisbett, 1988). For example, at the University of Alberta, most fourth-year psychology majors become familiar with issues regarding (a) design (e.g., controlling for confounding factors, problems with repeated measures), (b) procedure (e.g., consideration of the influence of research setting), (c) subjects (e.g., sample size, representativeness, subject awareness of the nature of the research), (d) measures (e.g., reliability, validity) and (e) the importance of using replicable procedures consistently across testing times. University students who do not pursue a science-oriented degree (e.g., English majors) and first-year university students will not likely receive such detailed instruction on scientific methodology.

In the current study, I expected that fourth-year Psychology majors would ask more sophisticated questions about methodology than fourth-year English majors and first-year students. However, I expected first-year students and fourth-year English majors to ask at least rudimentary questions about methodology, by virtue of their secondary-level instruction in science.

Scientific Reasoning

There are various *intuitive,* abstract rules of reasoning that are important to the scientific process, including methodological rules and statistical rules (Fong, Krantz, & Nisbett, 1986; Fong & Nisbett, 1991; Lehman, Lempert, & Nisbett, 1988; Lehman & Nisbett, 1990; Nisbett, Krantz, Jepson & Kunda, 1983). These rules are intuitive and abstract in the sense that people acquire a rudimentary understanding of them simply through exposure to various daily life experiences. The fact that these rule systems are intuitive and abstract has two educational implications. First, if one can acquire an understanding of these intuitive rules via exposure to various circumstances in daily life, then even brief, formal training in proper rule application should lead to improved scientific reasoning. Second, if these rules exist at an abstract level, one should expect transfer of training in rule application across a variety of domains (Fong et al., 1986; Fong & Nisbett, 1991, Lehman et al., 1988, 1990; Nisbett et al., 1983).

Fong et al., (1986, 1991) have discovered that people acquire rudimentary forms of a number of statistical rules by virtue of being exposed to a everyday events that involve variability and uncertainty (i.e., are probabilistic in nature). As a result of such exposure, people acquire a sense for the law of large numbers (i.e., sample values resemble population values as a direct function of sample size and an inverse function of population variability) and regression to the mean (i.e., extreme values for a sample are less likely to be extreme when new samples of the population are observed). Because these rules exist in rudimentary form, they are used imperfectly. However, high school and university students who received brief training in the application of these rules demonstrated increased frequency and quality of statistical reasoning in a wide variety of everyday problems. Furthermore, improved use of statistical rules was seen for three types of problem structures: probabilistic problems that clearly incorporate random variation (e.g., involve randomizing devices such as the classic gumball urn model), objective problems that are readily codable but do not have explicit cues about randomness (e.g., sports events or academic events), and subjective problems that are about judgments concerning interpersonal events and are therefore less codable. Thus, Fong et al., (1986, 1991) concluded that the use of statistical rules was domain independent. Finally, transfer of training was observed regardless of whether the training was achieved via rule-training (i.e., subjects were taught the formal properties of the law of large numbers) or if it was achieved via examples-training (i.e., subjects were presented with examples that illustrate how the law of large numbers could be applied to make inferences about everyday life). Such training effects were seen because the rules already existed in the subjects' repertoire.

Lehmen, Lempert, and Nisbett (1988) conducted similar research about the effect of graduate-level training (e.g., medicine, psychclogy, chemistry) on reasoning about everyday-life and scientific problems. One hypothesis was that type of graduate training would differentially affect statistical and methodological reasoning because each discipline is highly specialized with respect to the inferential rules emphasized. They found that compared to training in chemistry, three years of graduate training in disciplines such as medicine and psychology led to better performance in statistical and methodological reasoning for both scientific problems and everyday-life problems. Training in the psychological and medical fields, which are probabilistic in nature, alerted students to issues such as the riskiness in basing inferences on small samples, the problems posed by confounding variables in inferring causality, and the problems that arise when studying causes that are neither necessary nor sufficient. Graduate training in a deterministic science such as chemistry did not lead to improved reasoning performance; their understanding of these types of reasoning was at the same level as first-year graduate students.

Lehman and Nisbett (1990) extended this line of research to study the effects of undergraduate training in the natural sciences, humanities, and social sciences on the use of statistical and methodological rules of reasoning. Social sciences training (i.e., Psychology) produced large improvements in statistical and methodological reasoning. Training in the natural sciences (e.g., Chemistry) and in the humanities (e.g., English) produced a marginal, but statistically significant improvement on methodological reasoning.

In summary, even brief formal training in the application of statistical and methodological rules increases the frequency and quality of statistical and methodological reasoning in a wide variety of everyday life problems. Evaluating news briefs about research constitutes an everyday problem whereby statistical and methodological reasoning is applicable. Based on the research outlined above, I expected that Psychology majors would ask more and better questions about statistical and methodological issues than English majors and first-year students. Because these rules are intuitive in nature, I expected first-year students and English majors to ask at least some rudimentary questions about statistical and methodological issues. Finally, because these rules are abstract, I expected to see transfer of training effects across all the news briefs.

Social Aspects of Science

Contrary to popular belief, science is not an enterprise that is isolated from society and conducted by objective individuals (Snow, 1987). In reality, there is a social dimension to science which exists at three interrelated levels: the individual researcher; the community of researchers within a field of research; and the interaction between science and the rest of society.

Individual Researcher

Before the STS curriculum was incorporated into the school system, scientists were often presented as strictly logical beings who approach their task emotionlessly--stifling their own biases and interests (Kuhn, 1970; Mitroff, 1974; Tweeney & Walker, 1990).

Students taught under the STS curriculum may be made aware of the personal dimension of science; scientists are human beings who are fallible and sometimes biased in their thinking (Mitroff, 1974). Indeed, there are some scientists who are incompetent, ill-trained, and engage in unethical research practices (Edmonton Journal, Nov 13, 1993). Assuming that most students in this study have been exposed to the STS curriculum, it was expected that they would inquire about the competence and the motivations of the researcher when reading or hearing about research findings, especially for research that seems implausible. No particular group differences were expected.

The Scientific Community

The scientific community, composed of scientists who share a common interest in a particular area of research, has a strong impact on the research process and the dissemination of research findings (Fleming, 1988). It is the scientific community that establishes the criteria of adequacy regarding experimental design, observational techniques, data collection, theoretical argument, etc. (Snow, 1987). Thus, the scientific community provides the arena within which scientific consensus is established and scientific disagreement is settled. Furthermore, these criteria may be used in the peer-review process to determine which research is accepted into scientific journals. The stricter the criteria for acceptance into the journal, the more faith one can place on the quality of research and the credibility of the reported research conclusion.

In high school, students may learn about the role of the scientific community in establishing and disseminating scientific knowledge. Furthermore, Psychology majors are frequently taught about the issues related to the scientific community (e.g., consensus) in courses on historical and contemporary psychology. By virtue of this additional training, I expected that Psychology majors would ask more questions about the scientific community than first year students or English majors when evaluating the research described in the news briefs.

Science-Society Interaction

Science is embedded within a broad societal matrix. Therefore, it is shaped by and in turn helps shape the values of society (Martin, et al., 1990; Mullis & Jenkins, 1988). An awareness of this aspect of science is considered by some researchers to be a definitive aspect of scientific literacy (e.g., Gallagher, 1971; Hofstein & Yager, 1982). For this reason, the science curriculum (i.e., STS curriculum) has been significantly restructured to include discussion on the reciprocal relationship between science and society (Aikenhead, 1987, 1988; Fleming, 1987; Martin et al., 1990). For example, under the STS curriculum, students may be taught about the *mission-oriented* nature of science (Fleming, 1987). In other words, they may be taught that scientists do not conduct research only for the sake of knowledge itself; research is also conducted to provide knowledge that will have a positive influence on society. Assuming that students in this study have been exposed to the STS curriculum, I expected that they would inquire about the relevance of the research described in the news briefs.

In summary, science curriculum planners and educators have been faced with the challenge of improving scientific literacy of all students. To the extent that these objectives have been met in individual classrooms, students are taught about the structure of various scientific disciplines (i.e., content knowledge and the structure of knowledge), the epistemology of science (i.e., nature of scientific knowledge, theories, methodology, and reasoning), and the social aspects of science. Another challenge faced by science curriculum planners and educators has been to assess scientific literacy as a function of science education. Such a task has not been an easy one. There are a number of interpretive issues that must be addressed when assessing students understanding of science. In the next section, I will discuss these issues and then describe the coding schema that was developed for the current study.

Assessment of Science Literacy

Issues of Assessment

Over the last 40 years, a number of standardized instruments have been developed to assess students understanding of science (e.g. NAEP-Science, Science Process Inventory). With the exception of performance assessment, most assessment instruments have been focused on students' knowledge of scientific vocabulary and facts and have followed a multiple choice or Likert format (Champagne & Newell, 1992; Chiappetta, Fillman, & Sethna, 1991). This approach is problematic for curricular and psychometric reasons.

Curricular Issues

There are a number of negative curricular implications in using assessment tools focused on scientific facts. First, they perpetuate the myth that scientific knowledge is an established, complete body of knowledge. In reality, scientific knowledge is an evolving body of knowledge that is the product of investigative and consensual processes.

Second, this focus is based on the assumption that there is only one right answer to a scientific question. Furthermore, a correct response is assumed to reflect true understanding of the test item. When errors are made, no distinction is made between flawed recall, correct thinking but misinformation, or correct information but flawed thinking (Champagne et al., 1992). Thus, these kinds of assessment tools have little prescriptive value in terms of curricular modifications.

Third, assessment tools that focus on vocabulary and scientific facts are not only misrepresentative of science, they are incomplete. Such tools neglect the other components of scientific literacy outlined in the previous section of this paper. Fortunately, assessment tools focused on science literacy are slowly being developed and introduced to teachers (e.g., VOSTS, Aikenhead, 1987, 1988; Aikenhead & Ryan, 1991). Unfortunately, teachers seem hesitant to use them because they are more difficult to use. Scientific literacy, defined as being knowledgeable about the many components of science and an ability to apply this knowledge to the real world, is difficult to measure. For this reason, some teachers may fall back on the old assessment tools (Chiapetta et al., 1991). Multiple choice tests and Likert scales are easier to grade and it is easier to teach-to-test when the test items focus on *scientific facts.* Curricular reform efforts will be undermined if these problems are not addressed.

Psychometric Issues

Psychometric issues must also be considered when developing instruments that assess scientific literacy. The instruments must provide a reliable and accurate assessment of scientific literacy. Although reliability has not been a concern in the measurement of scientific knowledge, measuring students' true understanding of scientific issues has been a major problem (Aikenhead, 1987; 1988; Aikenhead & Ryan, 1991; Champagne et al., 1992). As alluded to earlier, a correct answer on a test item has been assumed to reflect true understanding of the test item. However, language is often used differently by students and assessors and this mismatch leads to misinterpretations of students' perceptions and understanding (Lederman & O'Malley, 1990). Multiple choice and Likert responses offer only a guess at student beliefs and the chances of the evaluator guessing accurately are remote (Aikenhead, 1987). For assessment tools that use written responses, ambiguity derives from the fact that some students write incomplete and inarticulate responses (Aikenhead, 1988).

Attempts to assess and teach science literacy skills have been thwarted by problems inherent in conventional assessment tools (Aikenhead, 1988). This situation has resulted in a recent call for the development of alternative assessment instruments. To address the curricular and psychometric problems outlined above, Aikenhead and his associates (Aikenhead, 1988; Aikenhead, Fleming, & Ryan, 1987) have developed an instrument called *VOSTS* (i.e., Views of Science-Technology-Society). VOSTS include items that cover both the epistemological and social aspects of science. Furthermore, the multiple choice items were derived not only from a theoretical or researcher-based viewpoint, they were also derived empirically from student viewpoints (Aikenhead, 1992). Thus, VOSTS provides a more comprehensive, accurate, and naturalistic picture of students' knowledge than conventional assessment instruments. Like VOSTS, the coding schema developed for the current study (i.e., Evaluation of Scientific Research or *ESR*) has incorporated these features. ESR differs from VOSTS in that it was designed to code students' written responses. In particular, it was designed to allow investigators to categorize the types of questions students want answered when evaluating science materials (e.g., media reports of scientific research). It was also designed to categorized justifications provided for asking each question. ESR will now be described in detail.

Evaluation of Scientific Research (ESR)

Scientific literacy entails more than possessing knowledge about the components of science discussed earlier (i.e., content knowledge; knowledge about the nature of scientific knowledge, theories, methodology reasoning and knowledge about the social aspects of science). Scientific literacy also entails an ability and inclination to apply such knowledge while engaged in evaluative thinking and decision-making. Although numerous instruments have been developed to assess people's knowledge of science, I am not aware of any instruments that have been formally developed to measure people's ability or inclination to use such knowledge for real-world tasks. The challenge in developing ESR was to develop an instrument that is (a) comprehensive (i.e., can assess knowledge of all components of science), (b) more accurate (i.e., more able to capture the meaning of peoples' questions, and (c) can be used to measure peoples' ability or inclination to use such knowledge in a variety of real-life tasks in a variety of scientific domains (e.g., medicine, environment). The following exposition is a discussion of how the ESR instrument meets these three criteria.

Comprehensiveness

When reading about scientific research, people may ask a number of evaluative questions that vary in terms of topic and in terms of specificity. The *TOPICS* and *KNOWLEDGE DOMAINS* section of ESR (see Appendices A and B) was designed to capture the nature of people's questions and justifications for asking each question.

Each evaluative question was assigned to one of nine major topic headings: Social Context (1.0), Agent (2.0), Methods (3.0), Data/Statistics (4.0), Relevance of Agent/Research On the Agent (5.0), Related Research-Beyond the Study Described (6.0), Other (7.0), Ambiguous But Relevant to the Research Described (8.0), and Offtask (Irrelevant to the Study) (9.0). Each of these topics is further divided into subtopics (see Appendix A). Thus, the TOPICS section of ESR is hierarchical in nature. General or rudimentary questions are assigned to the most general level of the appropriate category (e.g., What method did they use? was assigned Topic 3.0-Methodology). Specific questions are assigned to the most specific subcategories into which they clearly fit (e.g., Did they use a cross sectional design was assigned Topic 3.1.1-Type of Design). (For detailed descriptions of each topic and subtopic, see Korpan et al., 1994).

ESR-TOPICS is comprehensive in that it captures both the epistemological aspects of science (i.e., Agent, Methods, Data/Statistics) and the social aspects of science (i.e., Social Context, Relevance of Agent/Research, Related Research). The inclusion of these major headings was based on normative models of scientific research and intuitions about how scientists think. Furthermore, ESR-TOPICS is comprehensive in that three of the major headings (i.e., Other, Ambiguous But Relevant to the Research Described, and Offtask) capture peoples' unique questions.

Accuracy

Accuracy of ESR was addressed by the incorporation of a second coding schema (i.e., ESR-KNOWLEDGE DOMAINS) that is used to categorize people's justifications for asking each evaluative question (see Appendix B). Asking people to justify their questions serves two purposes. First, it helps determine whether people have *true knowledge* of the scientific issues around which they frame their questions. For example, some students may ask about *measurement reliability* simply because they heard the term used in their classes. Asking subjects to justify their questions about measurement reliability may help determine if they know why its important. Second, the justifications help clarify language ambiguities.

For example, questions are often hard to categorize because they are incomplete; people's questions sometimes consist of one or two words (e.g., "reliability?"). Furthermore, it is not certain that people use terminology in the same way that researchers do. When subjects are asking about one aspect of science (e.g., measurement reliability), they may be really be asking about something else (e.g., measurement validity). All in all, having people justify their evaluative questions helps alleviate these problems.

ESR-KNOWLEDGE DOMAINS is similar to ESR-TOPICS in that it is hierarchical in nature. It consists of six major headings: Knowledge About "Good" Scientific Practices and Conventions (1.0), Knowledge About the Social Context of Science (2.0), Knowledge About the Functions of Science in Society (3.0), Other (4.0), Ambiguous (5.0), and No Identifiable Knowledge Domain (6.0). Each major heading is further subdivided into more specific justifications (see Appendix B). (For detailed descriptions of each heading and subheading, see Korpan et al., 1994).

ESR-KNOWLEDGE DOMAINS is also similar to ESR-TOPICS in its comprehensive nature. It can be used to categorize justifications that are about epistemological aspects of science (i.e., Knowledge about Good

Scientific Practices and Conventions), sociological aspects of science (i.e., Knowledge about the Social Context of Science, Knowledge about the Functions of Science in Society) and peoples' unique justifications (i.e., Other, Ambiguous, No Identifiable Knowledge Domain).

Measuring the Ability/Inclination to Use Knowledge

The final criteria used in the development of ESR is that it would be useful for measuring people's ability or inclination to use their scientific knowledge in a variety of real-life tasks. This ability or inclination is a keystone of scientific literacy. ESR can be used to categorize evaluative questions and justifications that arise when people read media reports about research in various domains (e.g., medicine, environment). It can also be used to categorize questions and justifications that arise when people read primary sources of scientific research as well as when they wish to do a quality-check on their own research.

ESR has undergone a number of revisions since its inception. The original ESR was based primarily on normative models of scientific research and intuitions about how scientists think. This early form was tested in a pilot study. A number of modifications in the instrument have resulted, including the addition of major topic headings that capture peoples' unique questions, the reorganization of subtopics, and the addition of ESR-KNOWLEDGE DOMAINS. Furthermore, the task used to test students scientific literacy has undergone revisions as a result of the pilot study. Therefore, before describing the current study, it will be informative to first describe the pilot study.

The Pilot Study

Knowledge of treatment studies, a type of research that is common to many scientific disciplines, was examined in the pilot study (Korpan, Sadesky, Mabbott, Henderson, Bisanz, & Bisanz, 1992). As in the current study, the empirical goals of the pilot study were to determine (a) what scientific issues university students are sensitized to by virtue of their education, and (b) when and whether students engage in evaluative thinking about these aspects when reading reports of scientific research. The theoretical goal was to describe the underlying knowledge structures that are involved in the evaluation of scientific research.

Using a questionnaire format, 60 first- and second-year university students rated the credibility of conclusions in four domains (Diet, Textbooks, Pesticides, Dreams). Next, news briefs were read in which these same conclusions were reported in the context of treatment studies but with no methodological details provided. Each news brief had the following format: (a) researchers report a research finding, (b) a general concern or issue regarding the research area is described, and (c) an independent group promotes the importance of the research finding for addressing the general concern or issue (see Appendix C).

After reading each news brief, subjects were asked to write down what additional information they would need in order to determine whether the conclusions of these reports were true. This approach allowed us to ask whether domain specificity is an issue when students read about and evaluate scientific research. For example, do students ask similar or different evaluative questions when reading about research in different domains? Domain specificity has been an important issue in education theory and practice. Teachers who wish to teach their students to become scientifically literate need to know how they can teach their students to generalize their evaluative skills to domains not covered in the classroom.

Students generated a wide variety of requests for information, indicating that several areas of knowledge are used in evaluating media reports. Three factors seemed to influence the type of requests made: (a) presumed relatedness of the topic to the natural and physical sciences taught in school, such as biology and chemistry; (b) presumed personal experience with the topic; and (c) rated plausibility of the conclusion. Ratings for relatedness of the topic to the natural and physical sciences and ratings for personal experience with the topic were not collected in the pilot study. These two factors were derived from our examination of student requests. Only plausibility ratings were collected.

Ratings

Relatedness to the Natural and Physical Sciences

Students were more likely to ask about theory after reading news briefs about research that are closely related to the natural and physical sciences (i.e., how pesticides affect the environment and how diet affects cancer recovery) than after reading news briefs describing research not closely related to the natural and physical sciences (i.e., how crystals affect dreams and how textbooks influence grades) (See Table 1). For the *Diet* and *Pesticides* news briefs, 65% of the students asked about theories; 48% and 33% of students asked about underlying theories for the *Dreams* and *Textbooks* news brief, respectively. Perhaps students felt that a sound theory for the *Dream* research could not exist and that choice of textbooks are based on empirical findings rather that on theories.

The opposite trend seemed to occur with regard to methodology; students asked more methodological questions for the *Dream* and *Textbook* news briefs as compared to the *Diet* and *Pesticides* news briefs. For example, for the *Dream* and *Textbook* news briefs, questions about dependent measures were asked by 35% and 33% of the students, respectively. However, for the *Diet* and *Pesticide* news briefs, dependent measures were asked about by 10% and 22% of the students, respectively. Perhaps students assumed that methods are well established and reliable in the latter two research areas.

Personal Experience with the Topic

Due to the fact that the subjects in the pilot study were university students, we presumed that they would have had more experience with the topic of the *Textbook* news brief as compared to the topics of the other three news briefs. The *textbook* news brief elicited many requests about methodological variables (e.g., how grades are influenced by other variables, such a quality of teachers). Perhaps students' personal knowledge served as a source of insight about confounding factors. Many students (67%) also wanted to know more details about the textbook. Perhaps they felt that if they had this information, they could ask more specific evaluative questions about the textbook. In contrast, for the *Diet*, *Dream*, and *Pesticide* news briefs, questions about the nature of the treatment were brought up by only 25%, 15% and 35% of the students, respectively.

Plausibility of News Brief

The *Dream* news brief received a low plausibility rating compared to the other three news briefs. This news brief elicited questions that reflected a skepticism about the credentials of the people involved in the research. Twenty percent of the students asked about the qualifications of the researchers who conducted research on *Dreams*.. Qualifications of the researchers who conducted research on *Diet*, *Pesticides*, and *Textbook* were questioned by 15%, 8% and 8% of the students, respectively. Questions about the groups who promoted the importance of the research followed a similar trend. Finally, with regard to the effects of crystals on dreaming, 63% of students made requests for *proof*. Questions about supporting data for the *Diet*, *Pesticides,* and *Textbook* news briefs were made by 42%, 33%, and 47% of the students, respectively.

For all .he news briefs, there were fewer requests for information concerning the use of control groups and very few requests for information about the type of statistical analysis used. It is possible that first and second-year students have not had enough statistical training that would enable them to ask relevant questions about statistical issues.

Despite these findings, there were a number of problems inherent in the pilot study that made data interpretation difficult or limited to the study. These problems included: (a) limited ratings data, (b) inability to deal with ambiguous requests, and (c) uncontrolled subject selection. In the following section, these problems and the manner in which they were resolved in the current study will be outlined. The methods of the current study will then be described in more detail.

Interpretive Problems and Resolution

Limited Ratings

It is theoretically more interesting to describe how evaluation of news briefs is influenced not by topic per se, but by factors that transcend topics (e.g., Plausibility, Knowledge/Experience, Relatedness to the natural and physical sciences, Interest). With the exception of Plausibility, ratings on these factors were not collected in the pilot study. In the current study, one hypothesis was that the type of evaluative questions asked would be influenced by these factors. In order to test the hypothesis, four news briefs that represent orthogonal combinations of Plausibility (high, low) and Relatedness (high, low) were used. Selection of such news briefs was based on a norming study.

In the norming study, 72 first-year university students were presented with 20 news briefs, each describing a research conclusion in a different domain. For each news brief, students rated the conclusion in term of the three factors using a seven point Likert scale: PLAUSIBILITY: *How likely do you think it is that the conclusion is true?* (1=Not Very Likely, 7=Very Likely); EXPERIENCE-KNOWLEDGE: *How much experience with of knowledge of the topic of the conclusion do you have?* (1=Little Experience or Knowledge, 7=Much Experience or Knowledge) and RELATEDNESS: *How closely related is the conclusion to topics covered in the natural and physical sciences such as Physics, Chemistry, or Biology?* (1=Not Very Related, 7=Closely Related). For a sample of such ratings, refer to Appendix D.

Based on the ratings collected in the norming study, it seemed that the factor of experience/ knowledge of the domains described in the news briefs is a subject-related factor. This factor was not used in the selection of news briefs for the present study because it could not be controlled. It should be noted, however, that although the Experience-Knowledge dimension could not used to *select* the news briefs, this factor was still of interest in the current study because it was suspected that it might affect the types of evaluative questions asked. Thus, students in the current study were still asked to rate each news brief in terms of their experience with and knowledge of the topic. Another subject-variable suspected to influence the types of evaluative questions asked was interest in the topic. Consequently, subjects in the current study were asked to rate their interest in the topics described in each news brief.

Plausibility-ratings and Relatedness-ratings did vary as a function of topic of the news briefs. Thus these two factors were used to select four distinct news briefs (i.e., distinct in terms of the ratings) to be used in the current study.

The four news briefs selected for the current study were focused on research in the areas of Pesticides, Dreaming, Meditation, and the Environment (see Appendix E). Each news brief had the following average ratings for plausibility and relatedness, respectively: Pesticides = 5.58, 5.32 (High High); Dreams=1.82, 2.18 (Low, Low); Meditation=5.17, 2.96

(High, Low); and Environment=2.61, 5.11 (Low, High).

Ambiguous Requests for Information

Another problem in the pilot study was that students were not asked to justify their requests. Unfortunately, requests were often ambiguous and difficult to categorize. In the current study, subjects were asked to explain how knowing the answer to questions would help them evaluate the truth of the conclusion (i.e., justify their requests). Such justifications cleared up ambiguities of language use.

Subject Selection

In the pilot study, subjects were students who were enrolled in an introductory psychology class. Amount and type (i.e., major) of education and age were not controlled. Thus, it was difficult to draw conclusions as to how these subject-related factors influenced the types of requests made. In the current study, participants had to meet particular educational and age criteria.

First-year students were included to gather baseline information about the evaluative skills students have upon leaving high school and before receiving university training. Recruitment was restricted to students who had not taken any courses in statistics or research methodology. Also, to ensure that first-year students

methodology. Also, to ensure that first-year students

were comparable in terms of age, recruitment was restricted to first-year students who graduated from high school between 1975 and 1976.

Fourth-year Psychology and English majors were included in the current study to address whether university education in different disciplines affects the types of issues evaluated when reading scientific materials. It was presumed that sensitivity to various scientific issues is influenced by the amount of training individuals have had with regard to science. The recruitment of fourth-year Psychology majors was restricted to students who have taken at least one university course in statistics. It was also presumed that these students have had at least some university exposure to research methodology. By virtue of their university education, it was expected that they would make more references than the other two groups to statistical and methodological issues while evaluating the news briefs. Such a finding would corroborate the findings of Fong et. al (1986) as well as those of Lehman et. al. (1990).

The recruitment of fourth-year English majors was restricted to students who had not taken courses in statistical reasoning or research methodology. Because they have had little or no university level training in the nature of scientific research, it was expected that their evaluative performance would not be significantly different from first-year university students.

Finally, recruitment of fourth-year English and Psychology majors was restricted to students born between 1972 and 1973 to assure that they were comparable in terms of age.

II. CURRENT STUDY: METHODS

Subjects

Three groups of university students participated: 24 first-year students born between 1975 and 1976; 24 fourth-year English majors born between 1972 and 1973; and 24 fourth-year Psychology majors born between 1972 and 1973. In each group, half the students were male and half were female. First-year students participated in order to fulfill a course requirement in introductory psychology. Fourth-year English and Psychology majors were recruited in their respective classes and were paid ten dollars to participate in the study. Participation restrictions were made clear to the students at the time of recruitment.

Materials and Procedures

Test Booklet

Students were asked to complete a test booklet consisting of three parts. Some of the data collected in the booklet was not analyzed for the current study. First I describe Parts I, II, and III of the test booklet (see Appendices D to J). Next, I describe the procedures used to code the data collected in the test booklet.

Test Booklet-Part I

Students read four statements that described a possible research finding in four domains: (a) PESTICIDES: *Using a certain type of insecticide could be an important factor in causing a decline in robins' mating behavior;* (b) DREAMS: *Wearing a certain type of crystal could be important for increasing the frequency of dreams about future events;* (c) MEDITATION: *Practicing a certain type of meditation could be important for increasing the sense of well being in seniors;* and (d) ENVIRONMENT: *Fueling vehicles with a gasoline containing a certain type of poisonous chemical could be important for decreasing current levels of air pollution*.

After reading each statement, they were asked use a seven-point Likert scale to rate the following four factors: (a) PLAUSIBILITY: *How likely do you think it is that the statement you just read is true?* (1=Very Unlikely; 7= Very Likely); (b) EXPERIENCE/KNOWLEDGE: *How much experience with or knowledge of the general topic of the above statement do you have?* (1=Little Experience or Knowledge; 7=Much Experience or Knowledge); (c) RELATEDNESS: *How closely related is this statement to topics covered in natural and physical* *sciences such as Physics, Chemistry, or Biology?* (1=Very Unrelated; 7=Very Related); and (d) INTEREST: *How interested are you in the general topic of the above statement* (1=Very Uninterested; 7=Very Interested). Refer to Appendix D for an example of the ratings task.

Students were also asked to explain their ratings. These ratings and explanations are not presented in the present report. The order of statements (i.e., Environment, Pesticides, Dreams, Meditation) was counterbalanced across students.

Students were given 20 minutes to complete Part I. They were not allowed to go on to Part II until the 20-minute period was up.

Test booklet-Part II

Each subject read four news briefs that described research in the same domains as Part I: Environment, Pesticides, Dreaming, and Meditation (see Appendix E). No methodological details were provided in these news briefs. The format of each news brief was as follows: (a) a general concern or issue regarding the research area was described, (b) researchers report a finding, and (c) an independent group promotes the importance of the research finding for addressing the general concern or issue.

The conclusion put forth by the promoters was underlined. Also, the underlined conclusion was similar, but not identical to the statements that students had rated in Part I. The statements of Part I and the underlined conclusion in Part II differed in that the underlined conclusion in Part II was presented as an *actual* research finding involving a particular treatment agent (e.g., ...*fueling vehicles with gasoline containing this poisonous chemical (i.e., Quipmanol) is important for decreasing current levels of air pollution).* The statement in Part I presented a *possible* outcome that results from applying the class of treatment agents (e.g., *Fueling vehicles with a gasoline containing a certain type of poisonous chemical could be important for decreasing current levels of air pollution).*

After reading each news brief, students were asked to rate the underlined conclusion using the same four ratings they provided in Part I of the study (i.e., Plausibility, Experience/Knowledge, Relatedness, and Interest). They were not asked to explain these ratings.

After providing the four ratings for each respective news brief, students were given the following instructions:

> Suppose that this conclusion is very important to you and that you must determine whether it is true. Please generate a list of as many questions as you can that you would want to have answered before you

> decided whether the conclusion made by (Promoters Name) is true. Also, for each question you list, please indicate how you think the answer to that question would help you to evaluate the conclusion in the news brief.

Students were then presented with the following prompts:

> What is the first question you would want answered? How would an answer to this question help you to decide whether the underlined conclusion in the news brief is true? What is the next question you would want answered? etc.

Students were given 20 minutes to complete each news brief (ratings, questions, and justifications). They were not allowed to go on to the next news brief until the 20 minutes were up. Also, they were not allowed to go back to previous news briefs. For a sample of the question and justification-generation task, refer to Appendix F.

The news briefs were counterbalanced across students. The order of the news briefs paralleled the order of the statements presented in Part I.

<u>Test Booklet-Part III</u>

Part III consisted of four questionnaires (see Appendices G to J).

Questionnaire 1 was used to gather information about students' educational background in various scientific domains, both in high school and in university. They were also asked to indicate their age, sex, and year graduated from high school (see Appendix G).

Questionnaire 2 was used to determine students awareness of the limitations of science. Using a True-False format, students were asked to respond to five statements that tap issues about the tentativeness of scientific knowledge, the probabalistic nature of scientific conclusions, and the meaning of statistical analysis reported by scientists (e.g., statistical significance levels) (see Appendix H).

In Questionnaire 3, students were asked to rate their general interest in various science-oriented activities (i.e., pursuing a science-oriented career, reading about science-related topics, watching TV programs on science topics, participating in science-oriented hobbies). For each activity, they rated their interest using a seven-point Likert scale (1 = Very Little Interest; 7 = Very Much Interest) (see Appendix I). Responses from this questionnaire were not analyzed in the current study.

Questionnaire 4 was used to gather information about each students' belief in the paranormal. They were presented with 25 statements regarding various aspects of the paranormal and were asked to rate their agreement with the statement using a five-point Likert Scale (1 = Strongly Disagree; 3 = Undecided/Don't Know; 5 = Strongly Agree) (see Appendix J). Data from this questionnaire were analyzed but were not judged to be of importance to the current study.

Students were allowed to use as much time as they needed to fill out Part III. All students were given the four questionnaires in the same order.

Coding the Questions/Justifications

Detailed procedures for coding questions and justifications are outlined in the coding manual (Korpan et al., 1994). Four general procedures were used to assign codable responses. There were times, however, when students did not record their responses as requested. To code these responses, three additional coding principles were used.

In this section, the four general coding procedures are first described, followed by the three coding principles. Finally, topic and knowledge domain assignment are elaborated.

General Procedures

Four general procedures were followed when coding the responses (i.e., questions and justifications): (1)

The coder read each question. (2) The main idea in the question was identified and assigned a topic code (see Appendix A-ESR TOPICS). (3) The justification was then read. (4) The central idea underlying the justification was identified and assigned a knowledge domain code (see Appendix B-ESR KNOWLEDGE DOMAINS). These four procedures were carried out for every response for all four news briefs.

Coding Principles

The drafting of these four general procedures was based on the assumption that students would express a single idea in response to the first prompt (i.e., *What is the first/next questions that you would want answered?*) which could be given a single topic code. It was also assumed that students would provide a single justification after reading the second prompt (i.e., *How would an answer to this question help you to decide whether the underlined conclusion in the news brief is true?*) which could be given a single knowledge domain code. These ideas were embodied in Principle 1. Students did not always respond to these two prompts in a straight forward manner. To deal with these problems, two additional coding principles were adopted. The three coding principles were as follows:

Principle 1: Assume the subject is doing the task.

This principle had implications for (a) assigning codes for topics and knowledge domains, (b) identifying the link between the questions and the associated justifications, and (c) partitioning the responses into units that represent separate questions and justifications.

Principle 2: Topics codes should be assigned based on information in the questions only, whereas knowledge domain codes should be made based on the information in both the question and the associated justification.

When classifying ideas underlying each question, use of information from the associated justification was prohibited. Use of this information would have undermined reliability and affected validity. When classifying ideas underlying the associated justification, use of the question information was essential. It is conventional in discourse not to repeat an idea that has just been mentioned. In this case, interpreting the justification required establishing the referential coherence between nouns in the question and pronouns in the justification.

To give a knowledge domain coding, the coder was to ask, "How does the justification explain the question?".

Principle 3: There are three levels of responses. Responses may be (a) rudimentary or ambiguous with respect to the topics/knowledge *domain categories listed in ESR, (b) representative of one of the categories listed in ESR, or (c) representative of an "other" category.*

Rudimentary responses were the most naive types of responses. They addressed ideas in a simple, superficial way (e.g., What methods did they use?). There was not enough information in the response to assign it to subcategories. These responses were assigned at the most general level with respect to the category (e.g., Topic 3.0-Methods for rudimentary questions about methodology and Knowledge Domain 1.0-Knowledge about Good Scientific Practices for rudimentary justifications regarding methodology).

Ambiguous responses were vague responses. They seemed to be related to more than one category, but there was not enough information in the response to assign them to a single category or divide them into separate idea units. (e.g., How did they come to this conclusion?). Ambiguous questions were categorized as Topic 8.0-Ambiguous but Relevant to the Research Described. Ambiguous Justifications were categorized as Knowledge Domain 5.0-Ambiguous.

Responses coded as "Other" addressed a specific idea that was not included in ESR. Such questions were categorized as Topic 7.0-Other, whereas such justifications were categorized as Knowledge Domain 4.0-

Other.

In summary, to assign either a topic or knowledge domain code, the coder was to identify the top level category that best fit the question or justification and then ask three questions: (1) Is this response rudimentary or ambiguous with respect to the categories listed in the ESR instrument?, (2) Does this response fit one of the categories listed? and (3) Does this response belong to another category not specified in ESR? Every response belonged to one of these three groups.

Assignment of Topics

Topic assignment of each question was based on the gist or main idea of the question. When assigning the topic code, care was taken to take each question as it was written; the coder was not to add information during interpretation as this could alter topic assignment.

After reading each question, the coder first decided what *top level* category under ESR-TOPICS best captured the gist (e.g., 1.0-Social Context; 2.0-Agent; 3.0-Methods; 4.0-Data/Statistics; 5.0-Relevance of the Agent/Research on the Agent; 6.0-Related Research; 7.0-Other). A question was coded 7.0-Other if it was specific but could not be coded in categories 1.0 to 6.0 (i.e, these categories do not reflect the gist of the response). If a question was ambiguous with respect to these top-level categories (e.g, reflected either Methods *or* Data), it was assigned 8.0-Ambiguous But Relevant to the Research Described. If it was apparent that a question was unrelated to the news brief, it was coded as 9.0-Offtask.

Once the top-level category was assigned, the question was then assigned to a subcategory if possible. Coders used the most specific level of classification to which the question *clearly* fit; if the question did not fit clearly into one subcategory, only the top-level category was assigned. Sometimes a question reflected two or more clearly separable ideas (e.g., How many subjects were there in the sample and was the sample representative of the population). In this case, each idea in the question was given a separate topic code.

<u>Assignment of Knowledge Domains</u>

Justifications were assumed to reflect the knowledge structures that were tapped when students generated their questions; they reflected an understanding of why the information requested for in the question is important for evaluating scientific research.

The basic procedure for identifying the knowledge domains underlying the questions was similar to the procedures for assigning topics to the questions. Knowledge domain assignment was based on information that was present in both the questions and the justification. Sometimes, information from the question was required to interpret the justification.

Justifications were first assigned to knowledge domains at the *top level* of categories listed in ESR-KNOWLEDGE DOMAINS (e.g., 1.0-Knowledge About Good Scientific Practices and Conventions, 2.0 Knowledge about the Social Context of Science, 3.0-Knowledge about the Functions of Science in Society). Again, care was taken not to read more into the justification than was written. If the justifications reflected ideas not listed under knowledge domains 1.0 to 3.0 but represented a clear idea, it was assigned 4.0-Other. If the justification fit into more than one top level category, it was assigned 5.0-Ambiguous. If the justification was unclear, or if a justification was not provided for a question, it was assigned 6.0-No Identifiable Knowledge Domain.

If possible, justifications were further assigned to a subcategory under each category. Coders used the most specific level of classification to which the justification clearly fit. If it was not clear to what subcategory the justification belonged, only the top-level category was used.

There were some instances where students did not provide a justification for a question. Instead, they requested new information by asking another question (e.g., Q:"How old were the subjects?"; J: "Were they randomly selected?"). In this case, each question was given a topic coding and assigned a knowledge domain coding of 6.0-No Identifiable Knowledge Domain. There were also instances where students simply elaborated upon the question rather than providing a justification (e.g., Q: "How old were the subjects?"; J: "Where they old or young?"). In this case, the question and justification were given one topic code and a knowledge domain code of 6.0-No Identifiable Knowledge Domain was assigned. Finally, there were some instances where students gave *both* a question and its justification in response to the first prompt and no response to the second prompt (e.g., Did they use a control group to control for confounding variables?). Responding to the second prompt would be redundant. In this case, both the topic and knowledge domain code was assigned.

Reliability of Categorization

Three coders, including the principle coder (myself) coded one-third of the responses. Only the principal coder coded the remaining two-thirds of the responses.

Inter-coder reliability for top-level category assignment of topics and knowledge domains (e.g., Topics 1.0-Social Context vs 2.0-Agent etc.), averaged 96.1%. Disagreements were settled by discussion.

Inter-coder reliability for second-level category assignment (e.g., Topics 1.1-People vs. 1.2-Source of Information etc.), ranged from 67.4% to 89.8%. Most disagreements were in regard to subtopics 3.1 to 3.7 (i.e., methodology questions). Disagreements were settled by discussion.

III. RESULTS AND DISCUSSION

How individuals evaluate the plausibility of media reports of scientific research was investigated in the current study. The empirical goals were to determine (a) what scientific issues university students are sensitized to by virtue of their education and (b) under what conditions students engage in evaluative thinking about these issues when reading media reports of scientific research. These goals were achieved by examining the types of questions students asked when evaluating the news briefs and how these questions were affected by subject-related factors (e.g., type and amount of education) and article-related factors (e.g., domain of the news briefs). The theoretical goal was to determine the nature of the knowledge structures that are presumably involved in the evaluation of the media reports. One step towards this goal was examining the justifications that students gave for asking each question.

In the following section, the ratings data for the news briefs and questionnaire data are presented and discussed first. Next, the manner in which group-related factors (e.g., age, education) and article-related factors (e.g., ratings) influenced the generation of questions is described and discussed (under the section entitled Topics). Finally, the manner in which group-related factors and article-related factors influenced the generation of justifications is described and discussed (under the section entitled Knowledge Domains).

News Briefs: Ratings Data

Students rated each reported conclusion in terms of four dimensions: (a) Plausibility, (b) Knowledge and/or Experience, (c) relatedness to topics covered in natural and physical sciences (i.e., Relatedness), and (d) Interest. Students were also asked to explain their ratings (see Appendix D and E). Because reading the news briefs had little effect on these ratings, only the second set of ratings will be described.

For each rating dimension (i.e., Plausibility, Knowledge/Experience, Relatedness, and Interest), 3(Group) x 2(Sex) x 4(News Brief) MANOVAs with repeated measures on the last variable revealed that although no group or sex differences existed, there were significant news brief (i.e.,article) effects. Mean ratings for each news brief, summed across group, are presented in Table 2.

Plausibility

The plausibility ratings collected in the current study were similar to the plausibility ratings collected in the norming study, indicating that perhaps plausibility is primarily a text-variable rather than a subject-variable.

Students' rating for plausibility varied across news briefs, $F(3,207)=114.11$, $p < .001$. Ratings for the Pesticides and Meditation news briefs were significantly higher than the plausibility rating for the Environment and Dreams news briefs, $F(1,207)=289.45$, $p < .01$. Furthermore, students rated the Environment news brief as being more plausible than the Dreams news brief, $F1,207)=53.20$, $p < .01$. The Pesticides and Meditation news briefs were rated as being similarly plausible (see Table 2).

The following quotes exemplify the reasoning underlying the plausibility ratings:

Dreams:

There is no biological basis behind this finding. How can crystals possibly affect dreaming? The people who made this claim are a bunch of quacks.

Environment:

Although I know that gasoline additives can greatly decrease the amount of pollution emitted by vehicles, it doesn't make sense to me that a "poisonous" chemical would reduce air pollution--don't poisons generally "increase" pollution?

Pesticides:

There is a lot of research out there that shows that other pesticides, like DDT, have a terrible effect on wildlife. I'm not surprised that any pesticide would have a similar negative effect.

Meditation:

Meditation is an ancient technique that has been shown time and time again to have a positive effect on people's emotions. I am sure that meditation would increase "anybody's" sense of well-being.

It seems that plausibility ratings are higher if students believe that well-established research exists that is comparable to the research described in the news briefs (i.e., Pesticides, Meditation). Ratings are lower when the research finding seems counterintuitive but still tied to well-established research (i.e., Environment). Finally, if students cannot retrieve well-established research with which to compare the research described in the news briefs, they will give the reported research conclusion a low plausibility rating.

Knowledge/Experience

Students' knowledge of or experience with the four topics referred to in the news briefs was not equivalent, $F(3,207)=8.76$, $p < .001$. They indicated that they had more knowledge of meditation than the other three topics, $F(1,207)=18.43$, $p<.01$, and that their level of knowledge of pesticides, dreams, or the environment were similar (see Table 2). The following quotes exemplify the reasoning underlying the knowledge or experience ratings:

Pesticides:

> *I don't know much about robins and their mating habits. I don't know about pesticides either.*

Dreams:

> *I have heard that people use crystals for mystical sorts of purposes--but I have never heard of any research done on the effect of crystals on dreaming.*

Environment:

> *I don't know anything about additives and I have only limited knowledge about cars and pollution emissions.*

Meditation

> *I have heard about Transcendental Meditation from my friends*

who practice it and they say it works.

It appears that this rating were influenced by students' awareness of similar treatment and treatment effects.

Relatedness

The relatedness ratings in the current study were similar to the relatedness ratings provided in the norming study, suggesting that this dimension may be primarily a text-variable rather than a subject-variable.

The four topics referred to in the news briefs were perceived as being differentially related to the natural and physical sciences, $F(3,207)=137.48$, $p < .001$. Students considered the research described in the Pesticides and Environment news briefs to be significantly more related to the natural and physical sciences than the research described in the Dreams and Meditation news briefs, $F(1,207)=378.82$, $p < .01$. Furthermore, the research described in the Meditation news brief was rated as being more related to the natural and physical sciences than the research described in the Dreams news brief, $F(1,207)=32.89$, $p < .01$ (see Table 2).

The following quotes exemplify the reasoning underlying the relatedness ratings.

Pesticides:

In biology class, when we studied ecology, I learned about the negative influences that pesticides have on wildlife. This *research is similar to what we were taught in class.*

Environment:

Chemicals and additives--this research seems to be the sort of research that chemists would conduct.

Meditation:

Although psychologists would probably be the sort of scientists who would study this phenomenon, I'm sure that research could be done on the brain mechanisms underlying this effect.

Dreams:

No "scientist" with any self-respect would conduct this flaky type of research.

It seems that this rating was based on classroom exposure that students may have had to similar research or on their intuitions of what scientists study.

<u>Interest</u>

Students had more interest in some topics than in others, $F(3,207)=13.01$, $p < .001$. Their interest in dreams was significantly lower than their interest in pesticides, meditation, and environmental issues, $F(1,207)=38.00$, $p < .01$. Interest for the latter three topics were similar (see Table 2).

The following quotes exemplify the reasoning underlying the interest ratings.

Pesticides:

I am interested in nature and am quite concerned about animal *extinction. However, I am not really interested in reading about pesticides.*

Meditation:

I may consider practicing in meditation in the future, but for now I would rather engage in other forms of relaxation and increase my sense of well being in other ways.

Environment:

I am concerned about air pollution--but not gasoline or poisons

Dreams:

I would like to be able to tell the future-but I wouldn't bother trying these crystals because I don't believe they work.

It seemed that the interest ratings were influenced by two factors, interest in the general domain and interest in the particular research findings. Students seemed interested in the general domain, but uninterested in the particular research findings. These two factors combined resulted in a medium interest-rating. For the Dreams news brief, students interest seemed to be especially influenced by the plausibility of the research conclusion.

Subject Characteristics

Each participant in this study completed four questionnaires: (a) Questionnaire 1-Background Information (see Appendix G), (b) Questionnaire 2-Knowledge about the Limitations of Scientific Data (see

Appendix H), (c) Questionnaire 3-Interest in Science (see Appendix I), and (d) Questionnaire 4-Belief in the Paranormal (see Appendix J). Results from only Questionnaires 1 and 2 will be discussed. Questionnaire 3 (i.e., Interest in Science) was not analyzed in the current study. The purpose of Questionnaire 4 (i.e., Belief in the Paranormal) was to determine whether individual's belief systems affected the issues students consider when they evaluate scientific news briefs. The results were analyzed but judged not to be of sufficient interest to warrant discussion.

Questionnaire 1: Background Information

In this questionnaire, students indicated their age and the number of high school and university courses they completed in the natural sciences and physical (e.g., biology, chemistry, physics), social sciences (e.g., psychology, sociology), and statistics (see Appendix G).

Age. Age was a variable that was controlled for at the time of recruitment. First-year students (M=17.83 years) were younger than fourth-year Psychology majors (M=21.54 years) and fourth-year English majors (M=21.71 years). There were no sex differences.

Education. With regard to the number of *high school* classes completed in science courses, there were no significant group or sex differences. The average number of completed high school courses were as follows: biology: M=3.35, SD=1.52; chemistry: M=3.63, SD=1.42; physics: M=2.74, SD=1.91; and psychology: M=.57, SD=1.21. All students who participated in this study fulfilled the basic matriculation requirements of the University of Alberta. Because all students have had some education in science, it was expected that all students would consider a variety of scientific issues when evaluating the news briefs.

At the *university-level*, however, the three groups differed with regard to the number of completed science related courses, especially courses in the natural and physical sciences, $F(2,66)=4.917$, $p < .01$; the social sciences, $F(2,66)=199.833$, $p < .001$; and statistics, $F(2,66)=108.429$, $p < .001$. There were no significant sex differences.

Fourth-year Psychology majors completed significantly more university courses in the natural and physical sciences (M=3.63, SD=3.18) than first-year students (M=2.04, SD=1.99) and fourth-year English majors (M=1.58, SD=1.62), $F(1,66)=9.47$, $p < .01$. They also completed more university courses in the social sciences (M=18.13, SD=5.69) than first-year students (M=1.54, SD=.759) and fourth-year English majors (M=.958, SD=1.08), $F(1,66)=399.60$, $p < .01$. First-year students and fourth-year English majors were comparable in the number of science courses (natural-physical and social sciences) completed at the university level. First-year students completed few science courses simply because it was their first-year of university. English majors did not take any more science courses than required to meet the diversification requirements of the University of Alberta.

Most of the fourth-year Psychology majors who participated in this study completed a number of courses in the natural and physical sciences as part of their degree requirement. Furthermore, most of these students completed several psychology courses that spanned a variety of research domains (e.g., physiological, cognitive, social psychology). Therefore, it was presumed that they have had more exposure to research that is typical of the social sciences.

The number of statistics courses completed at the university level was a variable that was controlled for at the time of recruitment; fourth-year Psychology majors were required to have completed at least one statistics course to participate. First-year students and fourth-year English majors were required to have taken no university level statistics courses. All students who participated in this study met this requirement. The first-year students and fourth-year English majors completed no statistics courses, whereas the fourth-year Psychology majors completed one or more statistics courses (M=1.38, SD=.75).

Due to the fact that the fourth-year Psychology majors in this study have had more university-level training in science than first-year students and fourth-year English majors, it was expected that they would consider a wider range of scientific issues when evaluating the news briefs. Because the first-year students and fourth-year English majors in this study had comparable levels of science training, it was expected that they would perform similarly on the evaluation task. If they did not perform similarly, world knowledge or general university experience were considered to be possible causal factors.

Questionnaire 2: Knowledge About the Limitations of Scientific Knowledge

In order to evaluate scientific knowledge, one must be aware of its limitations. It was presumed that individuals who lack this awareness would not have the mind-set necessary to evaluate the epistemological aspects of research. The five true/false items that constituted this questionnaire concerned the limitations of scientific data and the kinds of inferences that can be drawn from scientific data (see Appendix H). A "false" response indicated an awareness of a particular type of limitation alluded to by the item (i.e., true=0; false=1).

Because Fourth-year Psychology majors have had more scientific training, it was expected that they would be more likely to give a "false" response. An ANOVA conducted on the pooled score (possible range=0 to 5) revealed a significant group effect that supported the above prediction, $F(2,66)=3.242$, $p < .05$. Fourth-year psychology students were more likely to respond "false" (M=4.21) than first-year students (M=3.63) and fourth-year English majors (M= 3.58), $F(2,69)=3.242$, $p < .05$. There were no significant differences between first-year students and fourth-year English majors. Thus, it seems that university training in Psychology has a positive influence on students' awareness of the limitations of scientific data.

Based on the five items used in this questionnaire, it was difficult to determine precisely the manner in which fourth-year Psychology students were more knowledgeable, however. Each item, analyzed separately via Kruskal Wallis one-way ANOVAs, revealed no significant group differences. All students had at least some appreciation of the types of limitations alluded to in each item. High school training in science may be responsible for this awareness. Because students in this study demonstrated an awareness of the limitations of data, it was expected that all would be able to evaluate the epistemological aspects of research described in the news briefs.

Keeping in mind the pattern of results outlined above, the evaluative questions (i.e., Topics) and justifications (i.e., Knowledge Domains) that were generated are described and discussed in terms of article effects and group effects. Sex effects are also discussed where applicable.

Topics

To facilitate comprehension of the data presented under *Topics*, the reader may wish to refer to the topic codes presented in Appendix A. Three types of main effects were investigated: (a) Group (i.e., first-year students, fourth-year Psychology majors and fourth-year English Majors); (b) Article (i.e., Pesticides, Dreams, Meditation, and Environment); and (c) sex. The dependent variable was the number of questions that were generated for each topic for each article. For each of the nine general topics (i.e., 1.0-Social Context, 2.0-Agent, 3.0-Methods, 4.0-Data, 5.0-Relevance of the Research, 6.0-

Related Research, 7.0-Other, 8.0-Ambiguous, and 9.0-Offtask), 3(Group) x 2(Sex) x 4(Article) MANOVAs with repeated measures on the last variable were conducted. Because of the hierarchical nature of the coding schema, several MANOVAs were conducted to reflect the level with which questions may have been coded. For example, five MANOVAs were conducted for Topic 1.0-Social Context, which consists of five levels of categories. First, all questions given a coding between 1.0 to 1.6 inclusive were pooled and analyzed as a general or first level category (i.e., 1.0-Social Context: Pooled). This score represents a compilation of both rudimentary and specific questions. Second, MANOVAs were conducted on questions after they were pooled into a score representing the second level of the hierarchy. For example, all questions that were given a coding between 1.1.1 and 1.1.5 inclusive were pooled and analyzed as Topic 1.1-People: Pooled. Similarly, all questions that were given a coding between 1.2.0 and 1.2.3 inclusive were pooled and analyzed as Topic 1.2-Source of Information: Pooled. Third, MANOVAs were conducted on questions after they were pooled into a score representing the third level of the hierarchy. For example, questions given a coding of 1.1.1.1 to 1.1.1.6 were pooled and analyzed as Topic 1.1.1-Researchers/

Experimenters: Pooled. Fourth, MANOVAs were conducted on each specific question type (e.g., Topic 1.1.1.1-Researcher Identification). Finally, MANOVAs were conducted on questions that were coded as Rudimentary because they were too rudimentary to be more specifically coded.

For Topics 3.0 (Methods) and 4.0 (Data), four types of MANOVAs were conducted to reflect each of the four levels in the hierarchy. Three types of MANOVAs were conducted to reflect each of the three levels of Topics 2.0 (Agent), 5.0 (Relevance of the Agent/Research), and 6.0 (Related Research), Finally, for Topics 7.0 (Other), 8.0 (Ambiguous) and 9.0 (Offtask), each of which consists of a single level, one MANOVA was conducted.

In the following discussion, MANOVAs will be discussed only for topics where the number of questions were sufficient to allow interpretation. In most cases, results from only the first and second levels of analysis will be reported. Floor effects, which were frequent for topics beyond the second level in the hierarchy, did not warrant a MANOVA. For example, rudimentary and ambiguous responses usually occurred too infrequently to warrant analysis. Nonsignificant results are reported if they are interesting (i.e., were not expected to be nonsignificant) and if the number of responses was sufficient to allow interpretation. Nearly all of the significant results were main effects for either group or article. A handful of interactions are also described. All interactions occurred for Topic 3.0-Methods and only for topics at the third level of coding. Finally, only significant effects for sex will be discussed. The means for the pooled analyses for all nine topics for groups and articles are shown in Tables 3 and 4, respectively.

Topic 1.0: Social Context

Group Effects: Pooled

By virtue of their curricular and extra-curricular experience with science, it was expected that all participants would ask about the social context of science. Indeed, in the current study, most students in all three groups generated this type of question. When this type of question was pooled, however, there was a significant group effect, $F(2,66)=4.35$, $p < .05$ (see Table 3). Fourth-year majors generated this type of question more often than first-year students, $F(1,66)=8.23$, $p < .01$. There were no significant differences between the two groups of fourth-year students.

Differences in performance between fourth-year majors and first-year students were not due to formal education in *science* because as they indicated in Questionnaire 1 (i.e., Background Information), first-year students and English majors did not differ in terms of high school or university science education. It may be that fourth-year majors in both disciplines acquired their knowledge about the social context of research through their general university experiences. Over a four year period, students may recognize that their area of study is taught and researched by a community of scholars who vary in training and biases. Differences in performance between first-year students and fourth-year students may also be due to extra-curricular experiences with science. For example, the media often reports research findings, some of which are suspect. It may be that with time (i.e., age) people acquire a skepticism of such reports.

Group Effects: Other Trends

1.1 People. Students frequently inquired about the people explicitly referred to in the news briefs (i.e., researchers, promoters) although prevalence varied across the three groups, $F(2,66)=3.75$, $p < .05$. Fourth-year Psychology majors ($M=3.04$, $SD=2.13$) and English majors ($M=2.79$, $SD=2.84$) asked this type of question more often than first-year students ($M=1.33$, $SD=1.71$), $F(1,66)=7.37$, $p < .01$. There were no significant differences between the fourth-year students. Typical questions included:

Who were the researchers?

Where did these guys get their training?

Again, world knowledge or general university experience is presumed to be responsible for this group effect.

No questions concerning the creators of the agent (Topic 1.1.3) and other experts (Topic 1.1.4) were generated.

1.2-1.5. Students in this study rarely generated other types of social context questions. They rarely asked about the source of information (Topic 1.2), funding (Topic 1.3), and never asked about the affiliated research institution (Topic 1.4) or about the source of research question (Topic 1.5).

Article Effects: Pooled

It was expected that plausibility would affect the frequency with which social context questions would be generated. Specifically, it was expected that students would question the credibility and motivation of the researchers who conducted the research and the groups who promoted the research after reading low-plausibility research (as did students in the pilot study). This prediction was supported by a significant article effect, $F(3,198)=4.09$, $p < .01$ (see Table 4). Students asked questions about social context after reading the Dreams and Environment news briefs significantly more often than after reading the Pesticides and Meditation news brief, $F(1,198)=11.67$, $p < .01$. There were no significant differences between the Dream and Environment news briefs, nor were there significant differences between the Pesticides and Meditation news briefs.

The Dreams and Environment news briefs were distinct from the Pesticides and Meditation news brief in terms of the plausibility ratings; the Dreams and Environment news briefs received lower ratings than the Pesticides and Meditation news briefs. Thus, when students did not believe that the research conclusion could be true, they tended to ask questions about the social context surrounding the research.

Article Effects: Other Trends

Plausibility particularly seemed to affect the number of questions asked about the people associated the research (Topic 1.1), $F(3,198)=3.17$, $p < .05$. Students asked this type of question significantly more often for the Dreams (M=.722, SD=.859) and Environment (M=.694, SD=.929) news briefs compared to the Pesticides (M=.528, SD=.691) and Meditation (M=.444, SD=.729) news briefs, $F(1,198)=9.58$, $p < .01$. There were no significant differences between the Dreams and Environment news briefs nor between the Pesticides and Meditation news briefs. Typical questions included:

> *Why would a "scientist" study the effects of crystals on dreaming?*
>
> *What do Autos for the Future have to gain from this research?*

It appears that students question the credentials and motivations of the people associated with the research when they have little faith in the research conclusions.

Topic 2.0: Agent/Theory

Group Effects: Pooled

Because the domains of the news briefs were not related to types of university training investigated in this study (i.e., Psychology, English), no particular group was expected to ask more questions about the agents used nor the mechanisms underlying the agent-effects. Indeed, all three groups generated this type of question frequently and equally often (see Table 3). It is possible that there would have been a significant group effect if different types of science majors participated in this study and if news briefs relevant to their area of specialty had been used. Students would have enough domain specific knowledge from which they could generate a number of theoretical questions.

Group Effects: Other Trends

When specific questions about agent/theory were analyzed, the results followed a similar pattern. There were no significant group effects regarding agent identification (Topic 2.1), agent mechanisms (Topic 2.2), or alternative agents (Topic 2.3). It should be noted that most agent/theory questions concerned agent mechanisms. All three groups were interested in how the agents purportedly caused the outcome. Questions concerning agent identification were also frequent. Questions about alternative agents were relatively infrequent.

Article Effects: Pooled

It was expected that students would be more likely to ask this type of question after reading about counterintuitive research. This prediction was not supported when the questions regarding agent and theory were pooled; students generated this type of question frequently and equally often after reading the four news briefs (see Table 4).

Article Effects: Other Trends

2.1 Agent Identification/Properties. There was a significant article effect with regard to the number of questions generated about agent ID and properties, $F(3, 198)=7.03$, $p < .01$. This type of question was more frequent for the Environment news brief (M=.556 SD=.748) than for the Pesticides (M=.281 SD=.422), Dreams (M=.292 SD=.488) or Meditation (M=.306 SD=.493) news briefs. There were no significant differences between the latter three news briefs. A typical question was:

What is Quipmanol?

One reason why this question was asked was to evaluate the theory underlying the research. In their ratings, many students indicated that reducing air pollution levels using a poison was counterintuitive. Thus, one reason they asked this question was to evaluate the soundness of the theory (Knowledge Domain 1.2). For example:

Q: *What is Quipmanol?*

J: *I want to know how it is possible for a poison to reduce air pollution.*

2.2 Agent Mechanisms, Effects/Side Effects. Students' intuitions about agent-mechanism seemed to influence the frequency with which this type of question was generated, $F(3,198)=10.19$, $p < .01$. Students inquired about agent mechanism significantly more often after reading the Environment news brief (M=1.000 SD=.856) compared to the Pesticides (M=.542 SD=.691) Dreams (M=.597 SD=.763) or Meditation (M=.389 SD=.693) news briefs, $F(1,198)=27.02$, $p < .01$. There were no significant differences among the latter three news briefs. A typical question was:

How does a "poison" reduce air pollution?

It was expected that, because the Dreams news brief received the lowest plausibility rating, many students would generate questions inquiring about the agent/theory underlying the research. Contrary to expectations, the frequency of these questions for the Dreams news brief was no greater than for the other news briefs. Perhaps students believed that a theory regarding the effects of crystals on dreaming could not exist and thus could not be evaluated. Perhaps they believed that research concerning the effect of crystals on dreaming was not *scientific* and thus not driven by any type of theory.

Topic 3.0: Methods

Group Effects: Pooled

It was presumed that in high school science courses, most students are taught that research methodology can be flawed or inappropriate and thus questionable. Thus, it was predicted that most students would ask questions about the methodology underlying the research described in the news briefs. This prediction was supported. Students from all three groups generated a variety of methodological questions (see Table 5).

By virtue of their university-level training in science, it was also expected that fourth-year Psychology majors would generate more questions that spanned a broader range of methodological issues. When the methodological questions were pooled, this prediction was supported by a significant group effect, $F(2,66)=8.70$, $p < .01$. Fourth-year Psychology majors generated methodological questions significantly more often than first-year students and fourth-year English majors, $F(1,66)=17.22$, $p < .01$ (see Table 3). There were no significant differences between first-year students and fourth-year English majors. Thus, training in science at the university-level seems to have sensitized Psychology majors to methodological issues that underlie research. To understand this trend, it is necessary to look at specific methodological questions.

Group Effects: Other Trends

As shown in Table 5, when the subtopics under 3.0-Methods were examined, this trend did not always hold up. Fourth-year Psychology majors did not demonstrate increased sensitivity to all methodological issues. For example, although several questions regarding Topics 3.2 (Agent Delivery) and 3.3 (Subjects) were generated, there were no significant group effects. Also, for Topics 3.5 (Consistency of Methods) and 3.6 (Replicability of Methods), no questions were generated by any of the groups. Only for Topics 3.1 (Design) and 3.4 (Measures) did significant group effects exist.

3.1 Design: Pooled. In university, science students (natural and social sciences) may be exposed to design issues either through research experience or by reading the primary research literature. In high school, however, students are less likely to design their own studies or read any research literature. Therefore, upon graduation, they probably would not be as familiar with design issues as, for example, fourth-year Psychology majors. Similarly, the training that English majors receive would not likely enable them to generate evaluative questions about design. Thus, fourth-year Psychology majors were expected to generated this type of question more often than the other two groups. When pooled, there was a significant group effect that supported this prediction, $F(2,66)=3.93$, $p < .05$ (see Table 5). This type of question was generated by fourth-year Psychology majors significantly more often than first-year students and fourth-year English majors, $F(1,66)=7.33$, $p < .01$.

With regard to the number of questions asked about control of confounding factors (Topic 3.1.4), there was also a significant group effect, $F(2,66)=3.22$, $p = .046$. Fourth-year Psychology majors asked this type of question significantly more often than first-year students and fourth-year English majors, $F(1,66)=6.27$, $p < .05$. This effect was complicated by a significant group by article effect, however, $F(6,198)=2.20$, $p = .045$. Only after reading the Dreams news brief did fourth-year Psychology majors generate this type of question more than the other two groups, $F(1,198)=7.99$, $p < .01$. All three groups similarly and frequently generated this type of question for the other news briefs. It seems that all three groups have at least an intuitive understanding of the issue of confounding factors. It is not clear, however, why fourth-year Psychology majors were especially likely to evaluate this issue after reading the Dreams news brief.

It should be noted that questions regarding other design issues (i.e., Topics 3.1.1-Type of Design, 3.1.2.-Research Context, and 3.1.3-Duration of Research) were not asked very frequently by any of the three groups.

3.4 Measures: Pooled. In high school, students are unlikely to generate their own measures when conducting research. Thus, they are not likely to get much exposure to issues surrounding measurement. By the time they reach the fourth-year of their program, however, most Psychology majors have had extensive exposure to measurement tools and issues, either by reading the primary research literature, or through classroom instruction. Thus, a significant group effect was expected. When the questions regarding measures were pooled, this prediction was supported, $F(2,66)=5.45$, $p = .006$ (See Table 5). Overall, fourth-year Psychology majors asked significantly more questions about measures than first-year students, and fourth-year English majors, $F(1,66)=10.79$, $p < .01$.

With regard to specific questions about measurement identification (Topic 3.4.1), there was also a significant group effect that supported the prediction, $F(2,66)=5.82$, $p = .005$. This type of question was asked by fourth-year Psychology majors (M=1.833, SD=1.341) significantly more often than by first-year students (M=1.000, SD=1.285) and fourth-year English majors (M=.708, SD=.955), $F(1,66)=10.63$, $p < .05$.

Surprisingly, compared to the other two groups, fourth-year Psychology majors were not more likely to generate evaluative questions regarding quality of measures (Topic 3.4.3). In fact, this type of question was relatively infrequent. It was expected that university-level training in science would sensitize fourth-year Psychology majors to measurement issues, such as reliability and validity. Based on the question-generation data, one would conclude that this is not the case.

Other questions regarding measures (i.e., Topic 3.4.2-Manipulation Check and Topic 3.4.4-Additional Measures) were rarely or never generated.

3.0 Methods: Rudimentary. Finally, it was expected that by virtue of their training, fourth-year Psychology majors would be less likely to generate rudimentary methodological questions. This prediction was not supported; all three group similarly and frequently generated questions that received this coding. A typical question was:

What methods did they use?

It is possible that first-year students and fourth-year English majors generated this type of question because, for some methodological issues, they did not have the kind of knowledge necessary to generate specific methodological questions. Also, it is possible that fourth-year Psychology majors generated this type of question as a set up for a more specific justification.

Article Effects: Pooled

Because this topic is so heterogeneous with regard to the subtopics it incorporates, there were no expectations regarding the frequency with which methodological question would be generated across news briefs. When all methodological questions were pooled, however, there was a significant article effect, $F(3,198)=25.55$, $p < .0001$ (see Table 4). Generally, students generated methodological questions significantly less often after reading the Environment news brief than after reading the Pesticides, Dreams, and Meditation news briefs, $F(1,198)=6.08$, $p < .05$. The latter three news briefs did not differ significantly. To explain this finding, analysis at the second level was necessary.

Article Effects: Other Trends

As shown in Table 6, when analyses were conducted on specific methodological issues, the particular nature of the article effect varied with the methodological issue being considered. It should be noted, however, that students generally asked specific methodological questions least often after reading the Environment news brief. It should also be noted that there was a significant article effect for each of the major methodological subtopics. Each subtopic will be discussed in turn.

3.1 Design: Pooled. With regard to the number of questions asked about the research design used in the studies, there was a significant article effect, $F(3,198)=14.68$, $p < .01$ (see Table 6). Students asked this type of question significantly more often after reading the Pesticides news brief than after reading the Dreams, Meditation, and Environment news briefs, $F(1,198)=32.58$, $p < .01$. Furthermore, students asked this type of question significantly more often after reading the Dreams news brief compared to the Meditation and Environment news brief, $F(1,198)=4.42$, $p < .05$. There were no significant differences between the Meditation and Environment news briefs. It is not clear why this result occurred. For example, there does not appear to be a cue in the Pesticides news brief that would cause students to ask about design.

With regard to specific questions about design, there was a significant article effects only for questions about the control of confounding factors (Topic 3.1.4), $F(3,198)=10.72$, $p < .0001$. Students generated this type of question after reading the Pesticides and Dreams news briefs significantly more often than after reading the Meditation and Environment news briefs, $F(1,198)=27.86$, $p < .0001$. There were no significant differences between the Pesticides and Dreams news brief, nor between the Meditation and Environment news brief).

This type of question seemed to be directed towards the animate nature of the subjects and their surroundings. Some typical questions asked for the

Pesticides, Dreaming, and Meditation news brief include:

Could ozone problems be the cause of reduced mating behavior rather than the pesticide?

Did they check to see that the subjects were not making their dreams come true just to please the researcher?

*Could it be that the seniors' sense of well-being was increased simply because they were receiving**was increased simply becau* *more attention, and not because of the meditation?*

Students may have asked this type of question most frequently for the Pesticides because not only were the subjects animate, but their natural environment could be extremely variable.

3.2 Agent Delivery: Pooled. There was a significant article effect with regard to the number of questions asked about agent delivery (i.e., procedures), $F(3,198)=9.09$, $p < .01$ (see Table 6). Students asked this type of question significantly more often after reading the Pesticides news briefs than after reading the Dreams, Meditation, and Environment news briefs, $F(1,198)=25.26$, $p < .01$. There were no significant differences between the latter three news briefs.

Agent delivery in the Dreams news brief (wear crystal), Meditation news brief (meditate), and Environment news brief (put additive in gasoline tank) may have seemed obvious to the students. In the Pesticides news brief, however, the manner in which the robins were exposed to Permaldrin was not obvious. A typical question was:

> *Were the robins directly exposed to Permaldrin, or were they exposed to it by eating insects sprayed with Permaldrin?*

It should be noted that regardless of research domain, questions about setting (Topic 3.2.2) and the time of year in which the research was conducted (Topic 3.2.3) was rarely and never generated, respectively.

3.3 Subjects: Pooled. There was a significant article effect with regard to the number of questions asked about the subjects who participated in the research described in the news briefs, $F(3,198)=16.38$, $p < .01$ (See Table 6). Generally, students asked this type of question significantly more often after reading the Meditation news brief than after reading the Dreams, Pesticides, and Environment news briefs, $F(1,198)=30.08$, $p < .01$. Furthermore, students asked this type of question significantly more often after reading the Dreams news brief than after the Environment and Pesticides news briefs, $F(1,198)=15.21$, $p < .01$. There were no significant differences between the Environment and Pesticides news brief.

The subjects who participated in the Dreams and Meditation studies were human, but nonhuman in the Pesticides and Environment studies. Perhaps students tended to ask more questions about the human participants simply because they could identify with them more and therefore think to ask more questions about them. As shown in Table 7, the majority of subject-related questions were too infrequent to warrant analysis. Most questions, which concerned the characteristics of the subjects (Topic 3.3.2), were influenced by information included in the news briefs. $F(3,198)=14.57$, $p < .01$ (see Table 7). Students asked this type of question after reading the Meditation news brief significantly more often than after reading the other three news briefs, $F(1,198)=41.52$, $p < .01$, which in turn did not differ. In the Meditation news brief, one characteristic of the participants alluded to was that they were senior citizens. This information seemed to provide students with a cue to ask questions about age-related factors. Typical questions included:

Exactly how old were these seniors?

How healthy are these people?

How active are these senior citizens?

<u>3.4 Measures: Pooled</u> There was a significant article effect with regard to the number of questions

asked about measures, $F(3,198)=5.04$, $p < .001$ (see Table 6). Students asked this type of question significantly more often after reading the Dreams and Meditation news briefs than after reading the Pesticides and Environment news briefs, $F(1,66)=47.19$, $p < .01$. There were no significant differences between the Dreams and Meditation news briefs nor between the Pesticides and Environment news briefs. Some typical questions include:

How did they measure sense of well-being?

How did they measure accuracy?

Perhaps students recognized the difficulty in measuring emotions, which are often capricious and context-bound. Dream content, based on self-reports is liable to error and bias. Furthermore, because research concerning Pesticides and the Environment was rated as being highly related to domains researched by natural or physical scientists, students may have considered the outcomes to be relatively more measurable.

Topic 4.0: Data/Statistics

Group Effects

For fourth-year Psychology majors, the prerequisite for participating in this study was the completion of at least one statistics course. First-year students and fourth-year English majors were not allowed to participate if they had any statistical training. By virtue of this difference in statistical training, it was expected that fourth-year Psychology majors would generate more questions than first-year students or fourth-year English majors about data and statistics (i.e., type and appropriateness of statistical analyses). This prediction was not supported. When the data/statistics questions were pooled, there were no significant group effects. In fact, none of the students generated questions about statistics (Topic 4.2). All of the questions concerned data. Questions about data were equally and frequently by all three groups (See Table 3). Based on the results from this analysis, one would conclude that undergraduate training in statistics does not sensitize one to statistical issues that underlie research.

Article Effects

It was predicted that students would make more requests for proof (i.e., data) after reading news briefs that described an implausible research finding. This prediction was not supported. When pooled, students asked about supportive data frequently and similarly for all four news briefs (see Table 4). When specific questions about data were evaluated, however, there were significant article effects that followed a complicated pattern.

Plausibility was a factor that seemed to influence the frequency with which students asked for information about absolute data (Topic 4.1.1), $\underline{F}(3,198)=9.00$, $\underline{p} < .01$. They asked this type of question after reading the Dreams news brief ($\underline{M}=.361$, $\underline{SD}=.635$) more often than after reading the Pesticides ($\underline{M}=.097$, $\underline{SD}=.298$), Meditation ($\underline{M}=.111$, $\underline{SD}=.316$), and Environment ($\underline{M}=.056$, $\underline{SD}=.231$) news briefs, $\underline{F}(1,198)=12.28$, $\underline{p} < .01$. There were no significant differences between the latter three news briefs. Students indicated (with their ratings) that such an outcome (i.e., accurate dreams about the future) is unlikely, with or without the crystal. Before they were willing to believe such a conclusion, they would need supporting evidence.

The degree to which the research domains described in the news briefs were related to the natural and physical sciences seemed to be predictive of how often students asked for information concerning comparative data (Topic 4.1.2), $\underline{F}(3,198)=4.50$, $\underline{p} < .01$. Students asked this type of question after reading the Pesticides ($\underline{M}=.208$, $\underline{SD}=.442$) and Environment ($\underline{M}=.208$, $\underline{SD}=.409$) news briefs significantly more often than after reading the Dreams ($\underline{M}=.028$, $\underline{SD}=.165$) and Meditation ($\underline{M}=.125$, $\underline{SD}=.333$) news briefs, $\underline{F}(1,198)=10.30$, $\underline{p} < .01$. There were no significant differences between the Pesticides and

Environment news briefs or between the Dreams and Meditation news briefs. The former two domains were rated as being more related to the natural and physical sciences than the latter two domains. Perhaps students felt that the outcomes reported in Pesticides and Environment news briefs could be more easily measured and thus more easily compared than the outcomes described in the Dreams and Meditation news briefs.

Sex Effect

There was a significant sex effect with regard to the number of questions asked about the data or statistics associated with the research described in the news briefs, $F(1,66)=4.59$, $p < .05$. Females ($M=2.111$, $SD=2.039$) asked this type of question significantly more often than males ($M=1.417$, $SD=1.025$). There is no logical explanation for this finding as the amount of statistical training was not significantly different for males and females.

Topic 5.0: Relevance of Research

It should be emphasized that, with the exception of Topic 5.1 (Generalizability/Specificity of Agent Effect), answers to questions about relevance of the research would not help one evaluate the *truth* of research conclusions. This topic was added to ESR because many students asked about relevance in the pilot study.

Group Effects

When pooled, questions about relevance of the research were frequently and similarly asked by all three groups (see Table 3). When more specific questions about relevance of the research described in the news briefs were looked at, the trend continued; there were no group differences for any of the subtopics under Topic 5.0. Questions about generalizability (Topic 5.1) were asked similarly and frequently by all three groups. This finding was surprising because Psychology majors are more likely than the other two groups to receive instruction on the implications of boundary conditions.

All three groups similarly and frequently asked about practicality, utility, and the function of the agent (Topic 5.2). Typical questions included:

Do only seniors benefit from this meditation, or would it help younger people too?

Can anyone practice this meditation in their own homes?

The nature of the task may have precluded students from generating other questions concerning relevance. For example, questions about recency of the research (Topic 5.3) may have been rare because students assumed that the research was recent, as media reports are typically current. Questions about the impact of the research (Topic 5.4), contribution of the research to the field (Topic 5.5), and audience familiarity (Topic 5.6) may have been rare because students were directed to assume that the research conclusion *was* very important to *them*.

Article Effects: Pooled

Domain of news brief was not expected to affect the frequency with which this type of question would be generated. When questions about relevance were pooled, however, this prediction was not supported, $F(3,198)=11.64$, $p < .01$ (see Table 4). Students generated this type of question after reading the Environment news brief significantly more often than after reading the Pesticides, Meditation, and the Dreams news briefs, $F(1,198)=30.16$, $p < .01$. Furthermore, students generated this type of question after reading the Pesticides and Meditation news briefs significantly more often than after reading the Dreams news brief, $F(1,198)=4.60$, $p < .05$. To understand this trend, results from the second level of analysis should be considered.

Article Effects: Other Trends

When specific questions concerning relevance of the research were investigated, there was a significant article effect for Topic 5.1 (Generalizability of

Effect), $F(3,198)=6.69$, $p < .01$ and Topic 5.2 (Practicality Utility, Function of Agent), $F(3,198)=29.82$, $p < .01$. In both cases, information in the news briefs seemed to cue students to ask about these issues.

5.1 Generalizability/Specificity of Effect. Students generated this type of question significantly more often after reading the Meditation (M=.292, SD=.488) news brief than after reading the Pesticides (M=.125, SD=.264), Dreams (M=.069, SD=.256), and the Environment (M=.083, SD=.278) news briefs, $F(1,198)=19.67$, $p < .01$. There were no significant differences between the latter three news briefs.

The fact that participants of the Meditation research were senior citizens may have provided students with a cue to ask whether this effect would occur for other types of subjects (e.g., younger).

5.2 Practicality, Utility, Function of the Agent. Students were more concerned about practicality after reading the Environment news brief (M=.667, SD=.872) than after reading the Pesticides (M=.056, SD=.231), Dreams (M=.083, SD=.278), or Meditation (M=.028, SD=.165) news briefs, $F(1,198)=6.46$, $p < .05$. There were no significant differences between the latter three news briefs. Typical questions included:

Can anyone add Quipmanol to their gas, or does it have to handled by professional?

Considering that Quipmanol is poisonous, how safe is it to use?

Although this type of question was occasionally generated for the other news briefs (e.g., *How difficult is it to learn to do this meditation?*), the fact that the agent (i.e., Quipmanol) used in the Environment research was poisonous may have cued many students to ask about practical issues, such as safety.

Topic 6.0: Related Research

Group Effects: Pooled

The research community is more likely to be discussed in science courses if the teacher discusses with students the various scientific debates and revolutions that have occurred within the domain. For example, such teachers are more likely to discuss how supporting data from other studies (or lack thereof) and consensus among the scientific community (or lack thereof) play a role in the acceptance (or rejection) of scientific viewpoints. Teachers who teach normal science (i.e., facts, methods that are currently accepted) may be less likely to discuss the scientific community. At the high school level, the decision to teach students about scientific debates and revolutions probably varies from teacher to teacher. By the time Psychology majors (at the University of Alberta) reach the fourth-year of their program, it was expected that they would have much exposure to various scientific debates and their resolution (i.e., currently accepted viewpoint(s)). Therefore, it was expected that fourth-year Psychology majors would generate more questions about related research than first-year students or fourth-year English majors. When questions about related research were pooled, however, this prediction was not supported. As shown in Table 3, all three groups similarly and infrequently generated this type of question. One one hand, it is possible that these students did not possess an understanding of the communal nature of science and thus did not think to ask about related research. On the other hand, the instructions provided to the students may have precluded them from asking about other research. Including science experts (i.e., scientists) in future studies would address the latter possibility.

Group Effects: Other Trends

When specific question regarding related research were investigated, this trend continued; all three groups similarly and infrequently generated questions about related research.

Most of the questions that were generated concerned whether other researchers conducted similar research (Topic 6.1) and supporting data (Topic 6.3). The remaining questions were evenly distributed over subtopics Topics 6.2 (Method of study), 6.4 (Consensus) and 6.5 (Related Research-Other).

Article Effects

It was expected that students would generate this type of question more often after reading low-plausibility research (i.e., Dreams) than after reading the other three news briefs. For example, it was expected that students would want to determine if the researchers who conduct the Dreams research were unique in their research interests and findings. This prediction was not supported. When pooled, it was apparent that this type of question was similarly and infrequently generated for each article (see Table 4). Again, students may not have generated this type of question very often because they were told to focus on the research reported in the news brief.

Topics 7.0-Other, 8.0-Ambiguous, and 9.0-Offtask

Group Effects and Article Effects

As seen in Tables 3 and 4, questions were infrequently assigned Topics 7.0 (i.e., other), 8.0 (i.e., ambiguous but relevant to the research) and 9.0 (offtask). Too few questions were given these codings to comment on any trends other than the fact that students usually generated questions about topics that are incorporated in ESR and usually remained on-task.

Summary

The empirical goals of this study were to determine (a) what issues students are sensitized to by virtue of their education and (b) under what conditions students engage in evaluative thinking about these issues when reading media reports of scientific research. These two goals were achieved by examining the types of questions students asked when evaluating the news briefs and how these questions were affected by subject-related factors (type of education, age, sex) and article-related factors, respectively.

Overall, all three groups generated a wide variety of questions, indicating that several areas of knowledge are used when evaluating reports of scientific research. Furthermore, they applied their knowledge when evaluating all four news briefs, indicating that they are able to generalize their knowledge to various domains. Characteristics of the news briefs did have a selective effect on the frequency with which many types of issues were evaluated, however.

By virtue of their training in research methods, it was expected that fourth-year Psychology majors would be more sensitized to the methodological issues that underlie research. To some extent, this prediction was supported by the question generation data; they were more likely to ask about research design and measures than the other two groups presumably because these issues are given more extensive coverage in university science classes. It was also expected that by virtue of their statistical training, fourth-year Psychology majors would be sensitized to the statistical issues that surround research. Based on the data from the question generation task, however, one would conclude that statistical training does not increase the frequency with which Psychology majors engage in statistical reasoning when evaluating research. This conclusion may be too hasty, however. As mentioned earlier, students on occasion generated ambiguous questions, making categorization difficult. Such ambiguities may cause one to underestimate students' knowledge about a given topic. Also, students were sometimes circuitous when generating their responses. For example, students sometimes asked about the researchers to find out what measures were used. To circumvent the problem of underestimating students' scientific knowledge, it is necessary to analyze the justifications that were generated for the questions.

Knowledge Domains

The theoretical goal of this research was to determine the nature of the knowledge structures that are presumably involved in the evaluation of scientific media reports. Examining the justifications given for each question is one step towards this goal. In particular, the justifications were used to explicate the underlying meaning of the associated questions. They were especially helpful for questions that were rudimentary or ambiguous. Thus, the justification generation task is a more sensitive measure of peoples' knowledge than the question generation task. To facilitate comprehension of the data concerning knowledge domains, the reader may wish to refer to the listing of knowledge domain codes presented in Appendix B.

Three types of main effects were investigated: (a) Group (i.e., first-year students, fourth-year Psychology majors, and fourth-year English Majors); (b) Article (i.e., Pesticides, Dreams, Meditation, and Environment); and (c) sex. Interaction effects were also investigated. The dependent variable was the number of justifications that were generated by individuals for each knowledge domain for each article.

For each of the six general knowledge domains (i.e., 1.0-Knowledge About Good Scientific Practices and

Conventions, 2.0-Knowledge about the Social Context of Science, 3.0-Knowledge About the Functions of Science in Society, 4.0-Other, 5.0-Ambiguous, and 6.0-No Identifiable Knowledge Domain), 3(Group) x 2(Sex) x 4(Article) MANOVAs with repeated measures on the last variable were conducted. As before, several MANOVAs were conducted to reflect the level with which the justifications may have been coded in the knowledge domain hierarchy. For example, four MANOVAs were conducted for Knowledge Domains 1.0-Good Scientific Practices, and 2.0-Social Context, each of which consists of four levels of categories. First, MANOVAs were conducted on the justifications after they were pooled into a score representing the general knowledge domain (i.e., first level). For example, all justifications that were given a coding between 1.0 to 1.9, inclusive, were pooled and analyzed as a general category, 1.0-Good Scientific Practices: Pooled. This pooled score represents both rudimentary and specific forms of the justifications. Second, MANOVAs were conducted on the justification after they were pooled into a score representing the second level in the hierarchy (e.g., 1.1: Adequacy of Research Design, 1.2: Theoretical Explanation, 1.3: Control of Confounding Factors, etc). Third, MANOVAs were conducted on each specific knowledge domain type (e.g., 1.3.1:

Control of Confounding Factors-Subjects). Finally, MANOVAs were conducted on justifications that were coded as Rudimentary because the justifications were too rudimentary to be more specifically coded.

For knowledge domain 3.0 (Functions of Science in Society) three types of MANOVAs were conducted to reflect each of the three levels in the hierarchy. For knowledge domain 4.0 (Other), 5.0 (Ambiguous) and 6.0 (No Identifiable Knowledge Domain), each of which consists of a single level, one MANOVA was conducted.

In the following discussion, MANOVAs will be discussed for knowledge domains where the frequencies were sufficient to allow interpretation. Results from more specific analysis will be reported only if frequencies were sufficient. Finally, nonsignificant results will be reported if they were interesting (i.e., were unexpectedly nonsignificant). Nearly all of the significant results were main effects for either group or article. There was only one significant group by article interaction for justifications concerning design (Knowledge Domain 1.1). Finally, only significant effects for sex will be discussed.

Knowledge Domain 1.0: Good Scientific Practices

Group Effects: Pooled

This knowledge domain is quite heterogeneous. The sub-domains included under this knowledge domain are conceptually related to Topics 2.0 (Agent/Theory), 3.0 (Methods), 4.0 (Data/Statistics), 5.0 (Relevance of the Research), and 6.0 (Related Research). When pooled, a group effect was evident, $F(2,66)=13.13$, $p < .01$ (see Table 8). Fourth-year Psychology majors were more likely than first-year students and fourth-year English majors to generate specific justifications regarding good scientific practices, $F(1,66)=26.28$, $p < .01$. There was no significant difference between first-year students and fourth-year English majors. To understand this trend, the second level of analysis must be investigated.

Group Effects: Other Trends

For clarity, each subdomain will be discussed briefly, if the frequency of justification warrants discussion. Insignificant group effects will be discussed if interesting (i.e., where group differences were expected).

1.1 Adequacy of Research Design. There was a significant group effect, $F(2,66) = 4.37$, $p = .017$ (see Table 9). Fourth-year Psychology majors generated this type of justification significantly more often than first-year students, and fourth-year English majors, $F(1,66) = 8.53$, $p < .01$. No significant difference existed between the latter two groups. This trend parallels the results for Topic 3.1-Design (see Table 5). Thus, Psychology majors generated more questions about design than the other two groups and were able to explain the relevance of design in evaluating research. A significant group by article effect complicated this finding, however, $F(6,198)=2.19$, $p = .046$. Fourth-year Psychology majors generated this type of justification more often than first-year students and fourth-year English majors for all news briefs _except_ for the Meditation news brief, $F(6,198)=7.34$, $.01 < p < .001$. For the Meditation news brief, individuals in all three groups generated this type of justification with equal frequency. It is not clear why this trend occurred.

1.2 Theoretical Explanation. All three groups similarly and frequently generated this type of justification (see Table 9). This trend parallels the results for topic 2.0-Agent/Theory (see Table 3). All three groups similarly inquired about agent and theory and were similarly able to explain the relevance of having an adequate theoretical explanation. University training in Psychology had no effect on students ability to evaluate this issue for the four domains described in the news briefs. Including majors in other disciplines in the study may have resulted in a different group trend.

1.3 Control of Confounding Variables. All three groups similarly and frequently generated this type of justification (see Table 9). It was quite surprising that fourth-year Psychology majors did not generate more justifications than the other two groups regarding this issue. It seems that all three groups have at least some understanding of the importance of controlling for confounding factors.

1.5 Measurement. University training did influence students' ability to explain the relevance of using adequate measures, $F(2,66)=4.63$, $p = .013$ (see Table 9). Fourth-year Psychology majors generated this type of justification significantly more often than first-year students and fourth-year English majors, $F(1,66)=8.60$, $p < .01$. There was no significant difference between first-year students and fourth-year English majors.

1.6 Statistical Inferences. University training also affected students' ability to explain the significance of data and statistics in making inferences, $F(2,66)=4.98$, $p = .01$ (see Table 9). Fourth-year Psychology majors generated this type of justification significantly more often than first-year students and fourth-year English majors, $F(1,66)=9.67$, $p < .01$. There were no significant differences between first-year students and fourth-year English majors. These results differ from the findings found for Topic 4.0-Data/Statistics (see Table 3), whereby no group differences were found. Although all three groups similarly and frequently generated questions about data, fourth-year Psychology majors were better able to explain the relevance of data in establishing the plausibility of research findings. This case illustrates the value of asking students to justify the meaning of their questions. The data based on the questions alone would erroneously lead one to conclude that statistical training does not increase Psychology majors' sensitivity or understanding of the role of data and statistics underlying research.

1.7 Consensual Processes. It was expected that fourth-year Psychology majors would generate this type of justification because it was presumed that they have had some exposure to how scientific debates are resolved in the scientific community. This prediction was not supported. In fact, this type of justification was never generated by any of the groups (see Table 9). Again, as suggested under Topic 6.0 (Related Research), it is not clear whether students lacked knowledge about the communal nature of science and the role of consensus in establishing scientific knowledge or if the instructions precluded them from demonstrating their knowledge.

1.8 Boundary Conditions. The three groups of students who participated in this study did not have equal understanding of the implications of boundary conditions in interpreting research results, $F(2,66)=3.87$, $p = .026$ (see Table 9). Fourth-year Psychology majors generated this type of justification significantly more often than first-year students and fourth-year English majors, $F(1,66)=8.51$, $p < .01$. There were no significant differences between the latter two groups. Thus, although all three groups similarly and frequently generated questions about generalizability of effect, Psychology majors were better able to explain the relevance of establishing boundary conditions. Again, having students justify their questions reveals the knowledge used when students evaluate research.

Good Scientific Practices: Rudimentary. It was expected that fourth-year Psychology majors would be less likely than the other two groups to generate this type of justification. This prediction was not supported; all three groups frequently and similarly generated rudimentary justifications about good scientific practices.

Article Effects: Pooled

Because this knowledge domain was so heterogeneous, the article effects were expected to vary with the scientific issue being considered. Conversely, when all justifications regarding good scientific practices were pooled, no particular article effects were expected. There was a significant article effect, however, $F(3,198)=21.26$, $p < .01$ (see Table 9). Generally, this type of justification was given significantly less often for the Environment news brief than for the Pesticides, Dreams, and Meditation news briefs, $F(1,198)=59.48$, $p < .01$ There were no significant differences between the latter three news briefs. At this point, it is difficult to interpret this article effect. It is necessary to look at the article effects for each subdomain separately.

Article Effect: Other Trends

As seen in Table 11, the trend varied with the issue (i.e., subdomain) being considered. Furthermore, for each subdomain, students did not always generate fewer justifications after reading the Environment news brief. Also, the trends did not always parallel the findings seen with regard to conceptually related questions. Each subdomain will be discussed briefly only if the frequencies were adequate to warrant discussion. Insignificant results will be reported only if interesting.

1.2: Theoretical Explanation. The frequency with which students justified their questions in terms of adequate theoretical explanation varied across news briefs, $F(3,198)=4.47$, $p = .005$ (see Table 11). They gave this type justification after reading the Environment and Dreams news briefs significantly more often than after reading the Pesticides and Meditation news briefs, $F(1,198)=18.95$, $p < .01$. There were no significant differences between the Pesticides and Meditation news briefs nor between the Dreams and Environment news briefs. This trend is similar to trend seen for Topics 2.1 (Agent ID) and 2.2 (Agent Mechanism). For those topics, questions were more frequent after students read the Environment news brief. Some examples included:

Q: *How did they know that the crystals have this effect?*

J: *They would have to some scientific explanation for this effect or these results are bogus.*

Q: *What mechanism do the researchers think is behind this effect (decreased air pollution)?*

J: *The researchers should have a sound chemical explanation for this finding.*

For the Dreams research, students sometimes set up this type of justification with other types of questions, rather than with theoretical questions. Again, students' justifications may be more revealing of their knowledge than the question generation data.

1.3 Control of Confounding Factors. There was a significant article effect with regard to the number of justifications generated concerning the elimination of alternative explanations, $F(3,198)=18.77$, $p < .001$ (see Table 11). Students gave this type of justification significantly less often after reading the Environment news brief than after reading the Pesticides, Dreams, and Meditation news briefs, $F(1,198)=56.21$, $p < .01$. There were no significant differences between the latter three news briefs. Students were also less likely to generate this type of *question* (topic 3.1.4) after reading the Environment news brief.

1.5 Measurement. As in the question generation task, consideration of measurement issues varied across news briefs, $F(3,198)=10.85$, $p < .0001$ (see Table 11). Students generated this type of justification after reading the Dreams and Meditation news briefs significantly more often than after reading the Pesticides and Environment news briefs.

1.8 Boundary Conditions. There was a significant article effect, $F(3,198)=9.57$, $p < .0001$ (see Table 11). Students generated this type of justification significantly more often after reading the Meditation news brief than after reading the Pesticides, Dreams and Environment news briefs. Furthermore, students generated this type of question significantly more often after reading the Pesticides news brief than after reading the Dreams and Environment news briefs, $F(1,198)=7.91$, $p < .01$. There was no significant difference between the latter two news briefs. This pattern is similar to the pattern seen for Topic 5.1-Generalizability of Effect.

Sex Effect

With regard to justifications regarding KD1.6-Statistical Inferences there was a significant sex effect, $F(1,66)=4.09$, $p = .047$. This type of justification was generated by females (M=1.869, SD=1.670) more frequently than males (M=1.139, SD=1.641). This pattern parallels the sex effect with regard to the number of questions generated about data and statistics. Again, there is no logical explanation for this effect.

Knowledge Domain 2.0: Social Context of Science

Group Effects: Pooled

This knowledge domain is conceptually related to Topic 1.0-Social Context. It was expected that all students would have some knowledge about social context, and that all students would generate this type of justification. Group differences were evident, however, $F(2,66)=58.25$, $p < .01$ (see Table 8). As in the question generation task, fourth-year majors gave this type of justification significantly more often than first-year students, $F(1,66)=9.76$, $p < .05$. There was no significant difference between fourth-year majors. Fourth year students were equally capable of explaining the relevance of their questions regarding social context. Again, general university experience or extra-curricular experiences with science may have increased fourth-year majors awareness of the social context of science.

Group Effects: Other Trends

As in the question generation task, references to funding (knowledge domain 2.2) and communication outlet (knowledge domain 2.3) were rare. Justifications about expert opinion (knowledge domain 2.5) were never generated. It seems that both high school and university-level training does not sensitize students to such issues. Most justifications concerning social context were about the people associated with the research, in particular the researchers who conducted the research and the groups who promoted the research.

2.1 Researchers and Research Groups. Group differences for this knowledge domain differed from the overall trend, $F(2,66)=3.40$, $p = .039$. This type of justification was generated by fourth-year Psychology majors ($M=1.708$, $SD=1.805$) more often than fourth-year English majors ($M=1.042$, $SD=1.681$) and first-year students ($M=.542$, $SD=.977$). Furthermore, fourth-year English majors generated this type of justification significantly more often than first-year students, $F(1,66)=5.10$, $p < .01$. It is not clear why the pattern changed for this knowledge domain.

2.4 Promoters and Detractors. Group differences for this knowledge domain were similar to the overall trend, $F(2,66)=3.22$, $p = .046$. Fourth-year Psychology ($M=1.250$, $SD=1.595$) majors and English majors ($M=1.583$, $SD=1.816$) generated this type of justification significantly more often than first-year students ($M=.500$, $SD=.933$), $F(1,66)=5.91$, $p < .05$. There were no significant differences between fourth-year majors.

Article Effects: Pooled

In general, it was expected that students would generate this type of justification after reading news briefs that were given a lower plausibility rating (i.e., Dreams, Environment). This prediction was supported by a significant article effect,

$F(3,198)=3.42$, $p = .018$ (see Table 10). When pooled, students generated this type of justification more often after reading the Dreams news brief (M=.375, SD=.592) than after reading the Pesticides (M=.319, SD=.552), Meditation (M=.194, SD=.399), and Environment (M=.220, SD=.451) news briefs, $F(1,198)=9.94$, $p < .01$. There was no significant difference between the latter three news briefs.

2.4 Promoters and Detractors The article effect for this knowledge domain was significant, $F(3,198)=3.54$, $p = .016$. Students generated this type of justification more often after reading the Dreams and Environment news briefs than after reading the Pesticides and Meditation news briefs, $F(1,66)=6.11$, $p < .05$. Again, the frequency with which students consider this issue seems to be affected by plausibility of the articles. There were no significant differences between the Dreams and Environment news briefs or between the Pesticides and Meditation news briefs.

Knowledge Domain 3.0: Functions of Science in Society

Group Effects: Pooled

Significant group effects with regard to this type of justification were not expected. Students' appreciation of the importance of practicality was not presumed to be a group-related variable. This prediction was supported in this study; this type of justification was similarly generated by all three groups (see Table 8). Thus, all the students in this study had a similar level of appreciation for the relevance of scientific research.

Group Effects: Other Trends

No group differences existed at the second level of analysis; all three groups similarly and frequently justified their questions in terms of utility for application (Knowledge Domain 3.1) but never in terms of information value (Knowledge Domain 3.3).

Article Effects: Pooled

It was presumed that student's understanding of the functions of science in society is an individual difference variable. Therefore, significant article effects were not expected. This expectation was not supported, $F(3,198)=57.72$, $p < .0001$ (see Table 10). Students generated this type of justification more often after reading the Environment news brief than after reading the other three news briefs, $F(1,198)=160.86$, $p < .01$. Furthermore, students gave this justification significantly more often after reading the Pesticides news briefs than after reading the Dreams and Meditation news brief, $F(1,198)=11.09$, $p < .01$. There was no significant difference between the Dreams and the

Meditation news briefs. The treatment agents in the Environment and Pesticides news briefs are harmful substances, and thus have practical implications. The fact that Quipmanol's poisonous properties were made explicit in the Environment news brief may have cued students to generate more justifications regarding functionality and practicality.

Article Effects: Other Trends

With regard to knowledge domain 3.1 (Utility for Application), there was a significant article effect, $F(3,198)=71.69$, $p < .0001$. Students gave this type of justification significantly more often after reading the Environment (M=1.556, SD=1.310) news brief than after reading the Pesticides (M=.236, SD=.567), Dreams (M=.042, SD=.201), and Meditation (M=.097, SD=.342) news briefs, $F(1,198)=212.36$, $p < .01$. This effect is parallel to the article effect regarding the number of questions asked about practicality (Topic 5.2). Again, the poisonous nature of Quipmanol probably motivated students to evaluate the utility of this agent.

Knowledge Domains 4.0-Others, 5.0-Ambiguous, and 6.0-No Knowledge Domain

Group Effects

With regard to knowledge domains 4.0 (Other), no particular group effects were expected. As shown in

Table 8, justifications given a coding of 4.0 were extremely rare.

With regard to knowledge domain 5.0 (ambiguous) and 6.0 (no knowledge domain), it was expected that fourth-year Psychology majors would generate these types of justifications less often than the other two groups. As shown in Table 8, this prediction was not supported; all three groups similarly, though infrequently generated these types of justifications. Thus, regardless of education, students can usually articulate an identifiable reason for asking various types of questions.

Article Effects

There were no expectations as to how often these codings would be assigned as a function of article. As shown in Table 10, justifications receiving a coding of 4.0-Other or 5.0-Ambiguous were too infrequent to analyze. With regard to knowledge domain 6.0, there were no significant article effects. Thus, there is nothing inherent in the articles that would cause students to generate justifications other than those included in ESR.

Summary

The theoretical goal of this research was to determine the nature of the knowledge structures that are presumably involved in the evaluation of scientific media reports. Examining the justifications given for each question was a step towards this goal. They were useful in explicating the underlying meaning of associated questions, especially questions that were rudimentary, ambiguous or circuitous.

Students in this study occasionally failed to generate justifications for their questions and sometimes generated rudimentary justifications. For example, first-year students and English majors had more problems than Psychology majors in explaining the relevance of issues such a data and boundary conditions. In the main, however, most students were able to explain the relevance of the issues introduced in their questions. Therefore, they were able to evaluate a variety of issues underlying scientific research at more than a superficial level. Furthermore, they were able to apply their knowledge to everyday tasks (e.g., evaluating media reports of science research) in a variety of domains. Thus, one may say that they have achieved some degree of scientific literacy.

Examining the justifications also helped to reveal students' knowledge of certain scientific issues, knowledge that was not revealed in the question generation data. In particular, the justification generation data showed that fourth-year Psychology majors are indeed sensitized to statistical issues associated with research. Furthermore, fourth-year Psychology majors were more able to explain the relevance of boundary conditions (i.e., generalizability of effect) than first-year students and fourth-year English majors. These findings should lead one to be more optimistic about the positive impact that university science instruction has on the quality of people's everyday thinking.

In conclusion, it is important to use tasks that get to the heart of students' scientific knowledge when assessing their evaluative skills. Assessors who rely on superficial measures of students' knowledge will likely underestimate the level of understanding that students truly have. Such measures may also over estimate the level of understanding that students have with regard to some issues. Having students provide justifications for the questions they generate presents a more accurate picture of the level of understanding they have with regard to various scientific issues.

IV. GENERAL DISCUSSION

In the current study, one aspect of scientific literacy was investigated; how individuals evaluate the plausibility of scientific research as reported by the media. Of particular interest were the kinds of scientific issues that university students consider when evaluating the truth of research conclusions. Scientific news briefs were selected because they are representative of the materials that individuals encounter in their daily lives. Evaluation of such news briefs was looked at because such a task can potentially tap a variety of scientific knowledge domains.

In the introduction of this paper, several components of scientific literacy were described, including knowledge about the epistemological aspects of science (e.g., methods, theory, data and statistics,) and knowledge about the social aspects of science (e.g., credentials and motivations of the individual researcher, the research community, science-society interaction). In the current study, a scientifically literate person was defined as one who possesses knowledge of these components and can apply this knowledge when evaluating the plausibility of reported research conclusions. With regard to this definition, two main questions were addressed. The first question concerned whether training in science affects the breadth of scientific issues that students consider when evaluating media reports. In other words, does training in science sensitize students to a variety of scientific issues? This question was addressed by looking at the pattern of group effects for the Topics. The second question concerned whether or not scientific training has an influence on students' understanding of the implications of these issues. In other words, how does training in science affect the depth of understanding that students have with regard to the issues introduced in their questions? This question was addressed by looking at the pattern of group effects for each type of knowledge domain. A related question concerned the conditions (i.e., research domains) in which students apply their knowledge. This question was addressed by looking at the pattern of article effects for each topic and knowledge domain.

These questions represent the typical concerns that have emerged among science educators since science literacy was deemed a curricular goal. Being able to regurgitate a variety of facts on a multiple choice task is no longer considered to be indicative of scientific achievement. A successful product of the educational system in science must have knowledge on a wide range of scientific issues, and must be able to apply this knowledge when reasoning about a variety of problems that lie outside the classroom walls. Science education should empower students to think and act efficiently and responsibly in the world. In the following discussion, the three questions outlined above will be discussed, followed by a discussion on the future implications of the evaluation task and ESR.

Current Research Findings

Topics: Educational Effects

A rather interesting pattern of group effects emerged in the question generation data. When examining the frequencies of each topic separately, it seemed that science training at the high school and university-levels did not always make students more sensitive to certain scientific issues. For example, formal education in science did not seem to influence the frequency with which social context issues (e.g., training and motivation of the researchers and promoters) were evaluated. Instead, exposure to communities of scholars (e.g., English, Psychology) at university or extra-curricular exposure to science may be responsible for students' awareness of the social context of science.

With regard to the impact that high-school training in science has on students' scientific literacy skills, all three groups of students, including first-year students and English majors, generated a wide variety of questions about the epistemological aspects of scientific research. In other words, they demonstrated that they possessed at least some understanding about the importance of theory (e.g., mechanism), methodology (e.g., design, agent delivery, subjects, and measures), and data. Thus, high school training in science seems to have provided them with a breadth of scientific knowledge. All three groups were also able to apply their knowledge in the evaluation task, indicating that they have achieved at least some degree of scientific literacy.

Surprisingly, university-level training *appeared* to have had little impact on the frequency with which certain issues were evaluated. By virtue of their methodological and statistical training at the university level, it was expected that fourth-year Psychology majors would be more sensitive to such issues. Only for certain issues concerning design and measures did fourth-year Psychology students demonstrate increased awareness. They did not demonstrate greater awareness of some measurement issues (e.g., reliability, validity) or issues associated with agent delivery (e.g., setting) and subjects (e.g., number of subjects). They also did not demonstrate a greater understanding of the importance of boundary conditions (i.e., generalizability of effect) with regard to data interpretation. Especially surprising was the fact that fourth-year students did not demonstrate increased awareness of statistical issues that underlies research. Before concluding that university-level training has minimal impact on scientific literacy, however, data from the justification-generation task should be considered.

Knowledge Domains: Educational Effects

The data from the justification-generation task was quite revealing in two ways. First, it helped reveal the fact that compared to first-year students and fourth-year English majors, fourth-year Psychology majors had a better understanding of issues related to statistics and boundary conditions. Such understanding was not unmasked in the question-generation task. When knowledge is not revealed, it is difficult to assess accurately how it is influenced by scientific training. Having students provide justifications for their questions presents a more accurate picture of the level of understanding of various issues.

Second, the justification data showed that most students were generally able to articulate the relevance of the issues they introduced in their questions. Thus, their understanding was not always restricted to a superficial level of comprehension.

The justification-generation data revealed some other interesting findings. For example, Psychology students did not demonstrate any knowledge about some social issues (e.g., publications, funding), some methodological issues (e.g., research setting, consistency of methods, replicability) or consensual processes. It was expected that by virtue of their university training in science that they would evaluate these issues. Before concluding that university training does not increase students' understanding of these issues, however, it is important to determine whether or not the task used in the current study simply failed to elicit such knowledge. Despite the fact that the justification task presented a more optimistic picture of students' knowledge and the way it is influenced by education, it is possible that the task used in the current study was insensitive to some types of knowledge.

Article Effects

To this point, I have mentioned how scientific training affects two components of science literacy: breadth of scientific knowledge and depth of scientific knowledge. Another important component is the ability to apply this knowledge to a variety of research domains. This aspect of scientific literacy was addressed by examining the article effects on students' response patterns.

Students tapped a variety of knowledge structures when evaluating all the news briefs, demonstrating that they were able to generalize their knowledge to a variety of research domains. However, numerous article effects indicated that there were a variety of cues in the news briefs that seemed to have a selective effect on the issues students choose to evaluate. For example, in the Environment news brief, the fact that Quipmanol was described as being a poison seemed to sensitize people to issues of utility and practicality. These article effects illustrate the importance of determining how test conditions elicit students' knowledge. Failure to do so may cause assessors to misrepresent students' true knowledge.

Future Implications

The question/justification generation task, in concert with ESR, has the potential to reveal students' knowledge about a variety of scientific issues. This approach can also be implemented for other reasons, such as determining science teachers' views on various scientific issues and measuring student-responses across grade levels.

By investigating the responses that science teachers generate, it would be possible to determine how teachers view the issues that they teach their students. It would be especially interesting to compare teachers who vary in terms of experience with the STS curriculum and in terms of *success* (i.e., in terms of their students science achievement test scores). Teachers do not necessarily incorporate all aspects of the STS curriculum in their classroom instruction. Also, they may be inexperienced or indeed lack an understanding of some of the STS issues that they may teach their students. The approach used in the current study may help reveal where their priorities, strengths, and weaknesses lie.

Longitudinal studies, intended to investigate the response-patterns of students at different points in their science education, would help reveal the gains that they achieve across grade-levels. By using ESR, gains could be operationalized in terms of the breadth, specificity, and depth. In other words, gains would be demonstrated if, across grade levels, students consider more issues when evaluating materials (i.e., ask a broader range of questions), are more able to evaluate specific aspects of these issues (i.e., ask more specific questions) and are more able to articulate the relevance of the issues of issues introduced in their questions.

It should be mentioned that, although ESR and the question/justification task seem to have considerable potential for assessing the effectiveness of curricula, more research is required before such an approach is implemented. For example, it would be advisable that ESR be used to code student responses on other types of tasks to get a fuller picture of what students know about various scientific issues. As alluded to earlier, it is quite possible that the pattern of results seen in this study are specific to the goal and task that students were given. For example, in this study, students were instructed to focus on the research presented in the news briefs. They were also asked to evaluate the news briefs in relation to themselves (i.e., were told to imagine that the information presented in the news briefs were very important to *them*). Performing well on this task with this goal reflects an important aspect of scientific literacy. Because of the nature of the task, however, students may have been less inclined to ask about related research. Having students generate a list of evaluative questions that they think other researchers would generate, for example, may help reveal their knowledge about related research and the scientific community. Using expository forms that contain different types of cues (e.g., TV reports, refereed journals), would also be informative for the same reason.

Investigating the responses of scientists who differ qualitatively in terms of scientific expertise (e.g., Psychologists, Chemists, Statisticians), would also be informative. By investigating their responses, it would be possible to determine whether ESR is comprehensive enough to capture the variety and specificity with which scientific issues can be evaluated. Investigating such responses would also help reveal whether the tasks in the current study are sensitive to different types of knowledge. For example, if Statisticians failed to question the statistics used in the research described in the news briefs, one may conclude that the task does <u>not</u> elicit such knowledge.

Finally, because ESR has not yet been validated, it would also be advisable to compare response patterns to independent, external criteria (i.e., other measures of scientific literacy) before it is implemented.

In summary, operationalizing peoples' scientific knowledge and their inclination to use such knowledge for a variety of problems in their daily lives for purposes of assessment has not been easy. For this reason, relatively few researchers have tried to measure science literacy-in-action. Furthermore, relatively little research has been directed to effects of training and experience on evaluation skills. The approach adopted in the current study has the potential to address this gap in the research literature from a number of perspectives.

V. TABLES

Table 1

Percentage of Students Making Specific Requests for Additional Information-Pilot Study

	Diet	Dream	Pest-cides	Text-books	Across Briefs
Promoter Qual.	13	28	15	13	45
Researcher Qual.	15	20	8	8	33
Supporting Data	42	63	33	47	87
Theory and Mechanisms	65	48	65	33	83
Other Research	37	20	40	5	57
Dependent Measures	10	35	22	33	65
Subjects	37	45	18	45	68
Treatments	25	15	35	67	80
Control Groups	8	5	8	5	23
Statistics	3	5	3	8	13

Note. * indicates percentages of students who made a least one request for any one of the news briefs.

Table 2

News Brief Ratings Summed Across Groups

Ratings	Pest	Dreams	Med.	Envir.
Plausibility	5.67[a]	2.04[c]	5.64[a]	3.72[b]
Knowledge/ Experience	2.88[b]	2.24[b]	3.36[a]	2.57[b]
Relatedness	6.03[a]	2.43[c]	3.68[b]	6.08[a]
Interest	4.69[a]	3.36[b]	4.65[a]	4.91[a]

Note. Superscripts indicate means in a row that do not differ ($p<.05$); a > b > c.

N=72 per article.

Table 3

Mean Number of Questions (Summed Across Articles) as a Function of Topic and Group

Topic	First Year	Psych. Majors	English Majors	Mean Across Groups
(1.0) Social Context	1.38[b]	3.33[a]	2.89[a]	2.53
(2.0) Agent	3.54[a]	3.17[a]	3.25[a]	3.32
(3.0) Methods	5.91[b]	9.46[a]	5.46[b]	6.94
(4.0) Data/ Stats	1.83[a]	2.13[a]	1.33[a]	1.76
(5.0) Rele-vance	1.79[a]	1.88[a]	2.08[a]	1.92
(6.0) Related Research	.20	.33	.21	.25
(7.0) Other	.08	.08	.04	.07
(8.0) Ambig.	.87	.41	.50	.59
(9.0) Off-	.16	.00	.21	.12
Total of Means	15.80	20.92	15.96	

Note. Superscripts indicate means in a row that do not differ ($p<.05$); a > b.

n=24 per group.

Table 4

Mean Number of Questions (Summed Across Groups) as a Function of Topic and Article

	Topic	Pest-cides	Dreams	Medita-tion	Envir-ronment	Mean Across Articles
(1.0)	Social Context	.54[b]	.75[a]	.46[b]	.78[a]	.63
(2.0)	Agent	.81[a]	.93[a]	.79[a]	.79[a]	.83
(3.0)	Methods	1.87[a]	2.15[a]	2.13[a]	.79[b]	1.74
(4.0)	Data/ Stats	.50[a]	.57[a]	.33[a]	.36[a]	.44
(5.0)	Rele-vance	.43[b]	.22[c]	.42[b]	.85[a]	.48
(6.0)	Related Research	.07	.08	.04	.10	.07
(7.0)	Other	.01	.01	.00	.04	.03
(8.0)	Ambig.	.14	.15	.17	.14	.15
(9.0)	Off-task	.03	.07	.00	.03	.03
	Total of Means	4.40	4.93	4.34	3.88	

Note. Superscripts indicated means in a row that do not differ ($p<.05$); a > b > c.

N=72 per article.

Table 5

Mean Number of Methodological Questions (Summed Across Articles) as a Function of Topic and Group

Topic	First Year	Psych. Majors	English Majors	Mean Across Groups
(3.1) Design	1.25[b]	2.13[a]	.92[b]	1.44
(3.2) Agent Delivery	.67[a]	.71[a]	.58[a]	.66
(3.3) Subj.	1.71[a]	2.63[a]	2.04[a]	2.13
(3.4) Meas.	1.25[b]	2.38[a]	1.29[b]	1.64
Total of Means	4.88	7.85	4.83	

Note. Superscripts indicate means in a row that do not differ ($p<.05$); a > b.

n=24 per group.

Table 6

Mean Number of Methological Questions (Summed Across Groups) as a Function of Topic and Article

	Topic	Pest-cides	Dreams	Medita-tion	Envir-ronment	Mean Across Articles
(3.1)	Design	.63[a]	.42[b]	.22[c]	.17[c]	.36
(3.2)	Agent Deliv.	.38[a]	.08[b]	.14[b]	.06[b]	.17
(3.3)	Subj.	.38[c]	.67[b]	.93[a]	.15[b]	.53
(3.4)	Meas.	.18[b]	.64[a]	.64[a]	.18[b]	.41
	Total of Means	1.57	1.81	1.93	.56	

Note. Superscripts indicated means in a row that do not differ ($p<.05$); a > b > c.

N=72 per article.

Table 7

Mean Number of Methological Questions (Summed Across Groups) as a Function of Topic and Article

Topic	Pest-cides	Dreams	Medita-tion	Envir-ronment	Mean Across Articles
(3.3.1) Sub. ID	.01	.11	.03	.07	.06
(3.3.2) Qual. Charac.	.17[b]	.13[b]	.51[a]	.04[b]	.22
(3.3.3) Numb. of Sub.	.13	.17	.15	.03	.12
(3.3.4) Sample Select.	.01	.11	.08	.00	.05
(3.3.5) Sample Repre.	.00	.00	.06	.00	.02
(3.3.6) Expect Biases	.00	.14	.06	.00	.05
Total of Means	.32	.66	.89	.14	

Note. Superscripts indicated means in a row that do not differ ($p<.05$); a > b.

N=72 per article.

Table 8

Mean Number of Justifications (Summed Across Articles) as a Function of Knowledge Domain and Group

Knowledge Domain	First Year	Psych. Majors	English Majors	Mean Across Groups
(1.0) Scient. Prac.	9.68[b]	14.92[a]	9.33[b]	11.31
(2.0) Social Context	1.21[b]	2.96[a]	2.58[a]	2.26
(3.0) Func.	2.63[a]	2.08[a]	2.58[a]	2.53
(4.0) Other	.04	.04	.04	.04
(5.0) Ambig.	.29	.08	.21	.20
(6.0) No. Know.	2.96	1.33	1.83	2.04
Total of Means	16.81	21.41	16.57	

Note. Superscripts indicate means in a row that do not differ ($p<.05$); a > b.

n=24 per group.

Table 9

Mean Number of Justifications (Summed Across Articles) About Good Scientific Practices as a Function of Knowledge Domain and Group

Knowledge Domain	First Year	Psych. Majors	English Majors	Mean Across Groups
(1.1) Design	.21[b]	.83[a]	.21[b]	.36
(1.2) Theory	1.75[a]	2.20[a]	2.13[a]	2.02
(1.3) Control Confounds	2.63[a]	3.33[a]	2.54[a]	2.83
(1.4) Altern. Agents	.21	.17	.21	.19
(1.5) Meas.	.83[b]	1.88[a]	1.17[b]	1.30
(1.6) Stat. Infer.	1.00[b]	2.33[a]	1.21[b]	1.52
(1.7) Consen.	.00	.00	.00	.00
(1.8) Bound. Cond.	1.04[b]	1.96[a]	.88[b]	1.29
Total of Means	7.67	12.70	8.35	

Note. Supercripts indicate means in a row that do not differ ($p<.05$); a > b.

n=24 per group.

Table 10

Mean Number of Justications (Summed Across Groups) as a Function of Knowledge Domain and Article

	Topic	Pest-cides	Dreams	Medita-tion	Envir-ronment	Mean Across Articles
(1.0)	Scient. Prac.	2.96[a]	3.41[a]	3.13[a]	1.82[b]	2.83
(2.0)	Social Context	.53[b]	.75[a]	.49[b]	.49[b]	.57
(3.0)	Func.	.51[b]	.06[c]	.17[c]	1.69[a]	.63
(4.0)	Other	.00	.03	.01	.00	.01
(5.0)	Ambig.	.07	.06	.03	.04	.05
(6.0)	No Know.	.32	.65	.51	.56	.51
	Total of Means	4.39	4.96	4.34	4.70	

Note. Superscripts indicated means in a row that do not differ ($p<.05$); a > b > c.

N=72 per article.

Table 11

Mean Number of Justifications (Summed Across Group) About Good Scientific Practices as a Function of Knowledge Domain and Article

Knowledge Domain	Pesti-cides	Dreams	Medi-tation	Envi-ronment	Mean Across Article
(1.1) Design	.11	.14	.04	.07	.09
(1.2) Theory	.44[b]	.58[a]	.33[b]	.67[a]	.51
(1.3) Control Confounds	.97[a]	.78[a]	.94[a]	.14[b]	.71
(1.4) Altern. Agents	.04	.03	.08	.04	.05
(1.5) Meas.	.13[b]	.57[a]	.42[a]	.18[b]	.33
(1.6) Stat. Infer.	.46	.49	.39	.18	.38
(1.7) Consen.	.00	.00	.00	.00	.00
(1.8) Bound. Cond.	.39[b]	.18[c]	.58[a]	.14[c]	.33
Total of Means	2.54	2.77	2.78	1.42	

Note. Superscripts indicate means in a row that do not differ ($p<.05$); a >b > c.

N=72 per article.

VI. REFERENCES

Aikenhead, G.S. (1987). High school graduates beliefs about science-technology-society. III. Characteristics and limitations of scientific knowledge. Science Education, 71, 459-487.

Aikenhead, G.S. (1988). An analysis of four ways of assessing students beliefs about STS topics. Journal of Research in Science Teaching, 25, 607-627.

Aikenhead, G.S., Fleming, R.W., & Ryan, A.G. (1987). High school graduates' beliefs about science-technology-society. I. Methods and issues in monitoring student views. Science Education, 71, 145-161.

Aikenhead, G.S. & Ryan, A.G. (In Press). Assessing students' views on STS topics. I. The development of a new generation instrument. Journal of Research in Science Teaching.

Aikenhead, G.S., & Ryan, A.G. (1992) The development of a new instrument: "Views on Science-Technology-Society (VOSTS), Science Education, 76, 477-491.

Brouwer, W., & Singh, a. (1983). Historical approach to science teaching. The Physics Teacher, 21, 230-235.

Champagne, A.B., Newell, S.T. (1992). Directions for research and development: Alternative methods of assessing scientific literacy. Journal of Research in Science Teaching, 29, 841-860.

Cheung, P.W. & Holyoak, K.J. (1985). Pragmatic reasoning schemas. Cognitive Psychology, 17, 391-416.

Chiappetta, E.L., Fillman, D.A., & Sethna, G.H. (1991). A method to quantify majors themes of scientific literacy in science textbooks. Journal of Research in Science Textbooks, 28, 713-725.

DeBoer, G.E. (1991). Scientific literacy and the new progressivism. In G.E. DeBoer (Ed.), A History of Ideas in Science Education. (pp 173-189). Teachers College Press, Teachers College, Columbia University

Feyerabend, P.K. (1981). Philosophical papers. Cambridge: Cambridge University Press.

Fleming, R. (1988). Undergraduate science students views on the relationship between science, technology and society, International Journal of Science Education, 10, 449-463.

Fong, G.T., Krantz, D.H., & Nisbett, R.E. (1986). The effects of statistical training on thinking about everyday problems. Cognitive Psychology, 18, 253-292.

Fong, G.T., & Nisbett, R.E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. Journal of Experimental Psychology: General, 120, 34-45.

Gallagher, J. (1971). A broader base for science teaching. Science Education, 55, 329-338.

Gregory, R.L. (1981). Mind in science. Penguin Books.

Hodson, D. (1985). Philosophy of science, science and science education. Studies in Science Education, 12, 25-57.

Hodson, D. (1988). Toward a philosophically more valid science curriculum. Science Education, 72, 19-40.

Hofstein, A., & Yager, R. (1982). Societal issues as organizers for science education in the 80's. School Science and Mathematics, 82, 539-547.

Howe, A. (1971). A lost dimension in elementary school science. Science Education, 55, 143-146.

Koelsche, C. (1965). Scientific literacy as related to the media of mass communication. School Science and Mathematics, 65, 719-724.

Korpan, C.A., Bisanz, G.L., Dukewich, T.L., Robinson, K. M., Bisanz, J., Thibodeau, M.H., Hubbard, K. E., & Leighton, J.P. (1994). Assessing scientific literacy: A taxonomy for classifying questions and knowledge about scientific research (Report No. 94-1).

Korpan, C.A., Sadesky, G.S., Mabbott, D. Henderson, J.M., Bisanz, J. & Bisanz, G.L. (1992). News briefs about scientific research: What do inquiring minds want to know. Presented at the annual convention of the Canadian Psychological Association. Quebec City, PQ.

Kuhn, T. (1970). The structure of scientific revolutions. Chicago: University of Chicago Press

Langley, P., Simon, H.A., Bradshaw, G.L., & Zytkow, J.M., (1987). Scientific Discovery: Computational Explorations of the Creative Processes. MIT Press, Cambridge Mass.

Lederman, N.G., & O'Malley, M. (1990). Students' perception of the tentativeness in science: Development, use, and sources of changes. Science Education, 74, 225-239.

Lehman, D.R., Lempert, R.O., & Nisbett, R.E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. American Psychologist, 6, 431-442.

Lehman, D.R., & Nisbett, R.E., (1990). A longitudinal study of the effects of undergraduate training on reasoning, Developmental Psychology, 26, 952-960.

Martin, B., Kass, H., & Brouwe, W. (1990). Authentic science: A diversity of meanings. Science Education, 74, 541-554.

Medawar, P. (1984). The limits of science. Oxford University Press.

Mitroff, I. (1974). The subjective side of science.

Amsterdam: The Elesevier Scientific Publishing Company

Mullis, I.V.S., & Jenkins, L.B. (1988). The Science Report Card: Elements of Risk and Recovery. Educational Testing Service, Princeton, NJ.

Munby, (1982). What is scientific thinking?-A discussion paper. Science Council of Canada.

Nadeau, R. & Desautels, J. (1984). Epistemology and the Teaching of Science. Science Council of Canada, Canadian Government Publishing Centre. Toronto

Nisbett, R.E., Krantz, D.H., Jepson, C., & Kunda, A. (1983). The use of statistical heuristics in everyday inductive reasoning. Psychological Review, 90, 339-363.

NSTA. (1982). Science-technology-society: Science education for the 1980s (National Science Teachers' Association Position Statement) Washington DC:NSTA.

Popper, K. (1970). The logic of scientific discovery. Hutchinson of London, London.

Posner, G.J., Strike, K.A., Hewson, P.W., & Gertzog, W.A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. Science Education, 66, 211-227.

Rosenthal, D.B., (1989). Two approaches to science-technology-society (S-T-S) Education. Science

Education, 73, 581-589.

Russell, B. (1948). Human knowledge: Its scopes and limits. London.

Ryan, A.G. (1987). High-school graduates' belief about science-technology-society. IV. The characteristics of scientists. Science Education, 71, 489-510.

Ryan, A.G., & Aikenhead, G.S. (1992). Students preconceptions about the epistemology of science, Science Education, 76, 559-580.

Shortland, M. (1988). Advocating science: Literacy and public understanding. Impact of Science on Society, 38, 305-316.

Smith, N. (1974). The challenge of scientific literacy. The Science Teacher, 41, 34-35.

Snow, R.E. (1987). Core concepts for science and technology literacy. Bulletin of Science and Technology in Society, 7, 720-729.

Tweeney, R.D., & Walker, B.J. (1990). Science Education and the Cognitive Psychology of Science. In B.F. Jones and L. Idol (Eds.) Dimensions of thinking and cognitive instruction. Hillsdale, N.J

VII. APPENDICES

Appendix A: Topic Codes

TOPIC CODES

1.0 SOCIAL CONTEXT

1.1 PEOPLE

1.1.1 Researchers/Experimenters

1.1.1.1 Identification
1.1.1.2 Number
1.1.1.3 Testimonials
1.1.1.4 Training/Qualification/Credentials
1.1.1.5 Motivation/Expectations/Biases/Beliefs
1.1.1.6 Researchers/Experimenters - Other

1.1.2 Promoters and Detractors

1.1.2.1 Identification
1.1.2.2 Number
1.1.2.3 Testimonials
1.1.2.4 Training/Qualifications/Credentials
1.1.2.5 Motivation/Expectations/Biases
1.1.2.6 Promoters and Detractors - Other

1.1.3 Creators of Agent

1.1.3.1 Identification
1.1.3.2 Number
1.1.3.3 Testimonials
1.1.3.4 Training/Qualification/Credentials
1.1.3.5 Motivation/Biases
1.1.3.6 Creators of Agent - Other

1.1.4 Other Experts

1.1.4.1 Identification
1.1.4.2 Number
1.1.4.3 Testimonials
1.1.4.4 Training/Qualification/Credentials
1.1.4.5 Motivation/Biases
1.1.4.6 Other Experts - Other

1.1.5 People - Other

1.2 SOURCE OF INFORMATION

1.2.1 Publication Outlet
1.2.2 Author(s) or Broadcasters
1.2.3 Source of Information - Other

1.3 FUNDING ISSUES

1.4 LOCATION OF RESEARCH - IDENTIFICATION

1.5 SOURCE OF RESEARCH QUESTION - IDENTIFICATION

1.6 SOCIAL CONTEXT - OTHER

2.0 AGENT

2.1 AGENT IDENTIFICATION/PROPERTIES

2.2 AGENT MECHANISMS, EFFECTS/SIDE EFFECTS
Agent Mechanisms
Side Effects

2.3 ALTERNATIVE AGENTS

2.4 AGENT - OTHER

3.0 METHODS

3.1 DESIGN
3.1.1 Type of Design
3.1.2 Research Context
3.1.3 Duration of the Research
3.1.4 Control of Other Factors/Confounding Factors
General
Specific
3.1.5 Design - OTHER

3.2 AGENT DELIVERY
3.2.1 Nature of Agent Delivery
3.2.2 Setting
3.2.3 Calendar Time
3.2.4 Agent Delivery - Other

3.3 SUBJECTS
3.3.1 Identification
3.3.2 Qualitative Characteristics
3.3.3. Number of Subjects Tested
3.3.4 Sample Selection
3.3.5. Representativeness of Sample
3.3.6. Expectations/Motivations/Awareness/Beliefs
3.3.7. Subjects - Other

3.4 MEASURES
3.4.1 Identification of Measures
Basic
Technical-Operational Definition of Dependent Variable
3.4.2 Manipulation Check
3.4.3 Evaluation of the measures
3.4.4 Additional measures
3.4.5 Measures - Other

3.5 CONSISTENCY OF METHODS/PROCEDURES

3.6 REPLICABILITY OF METHODS/PROCEDURES

3.7 METHODS - OTHER

4.0 DATA/STATISTICS

4.1 DATA

4.1.1 Absolute Nature of the Data for the Dependent Variables
4.1.2 Comparative Nature of the Data for the Dependent Variable
4.1.3 Duration of Effect
4.1.4 Additional data
4.1.5 Data - Other

4.2 STATISTICS

4.2.1 Type/Nature of Statistical Analysis
4.2.2 Appropriateness of Statistics
4.2.3 Statistics - Other

5.0 RELEVANCE OF THE AGENT/RESEARCH ON THE AGENT

5.1 GENERALIZABILITY/SPECIFICITY OF AGENT EFFECT

5.2 PRACTICALITY/UTILITY/FUNCTION OF AGENT APPLICATION

5.3 RECENCY OF RESEARCH

5.4 IMPACT OF THE AGENT/RESEARCH FINDINGS/CONCLUSION

5.5 CONTRIBUTION OF THIS RESEARCH TO THE DOMAIN

5.6 AUDIENCE FAMILIARITY WITH THE RESEARCH FINDINGS

5.7 RELEVANCE - OTHER

6.0 RELATED RESEARCH (BEYOND THE STUDY DESCRIBED)

6.1 SIMILAR DOMAIN OF STUDY

6.2 METHOD OF STUDY

6.3 SUPPORTING DATA/DATA REPLICATION

6.4 CONSENSUS/NONCONSENSUS

6.5 RELATED RESEARCH - OTHER

7.0 OTHER

8.0 AMBIGUOUS BUT RELEVANT TO THE RESEARCH DESCRIBED

9.0 OFFTASK (IRRELEVANT TO THE STUDY)

Appendix B: Knowledge Domain Codes

KNOWLEDGE DOMAIN CODES

1.0 KNOWLEDGE ABOUT "GOOD" SCIENTIFIC PRACTICES AND CONVENTIONS

1.1 ADEQUACY OF RESEARCH DESIGN

1.2 THEORETICAL EXPLANATION

1.3 CONTROLLING FOR CONFOUNDING VARIABLES/ELIMINATION OF OTHER CAUSES
1.3.1 Subjects
1.3.2 Treatment
1.3.3 Alternative Explanations - Other

1.4 IDENTIFICATION OF AGENTS OR TREATMENTS WITH IDENTICAL EFFECT

1.5 MEASUREMENT
1.5.1 Reliability
1.5.2 Validity
1.5.3 Measurement - Other

1.6 STATISTICAL INFERENCES
1.6.1 Sample size
1.6.2 Effect size
1.6.3 Variability
1.6.4 Statistical Significance and Chance
1.6.5 Statistical Inferences - Other

1.7 CONSENSUAL PROCESSES
1.7.1 Replication by the Original Researchers
1.7.2 Independent Replication
1.7.3 Fit with other Evidence or Theories
1.7.4 Consensual Processes - Other

1.8 BOUNDARY CONDITIONS

1.9 OTHER

2.0 KNOWLEDGE ABOUT THE SOCIAL CONTEXT OF SCIENCE

2.1 RESEARCHERS OR RESEARCH GROUPS
2.1.1 Quality
2.1.2 Biases
2.1.3 Researchers - Other

2.2 SOURCE OF FUNDING
2.2.1 Quality
2.2.2 Biases
2.2.3 Source of Funding - Other

2.3 COMMUNICATION OUTLET
2.3.1 Quality
2.3.2 Biases
2.3.3 Communication Outlet - Other

2.4 PROMOTERS AND DETRACTORS
2.4.1 Quality
2.4.2 Biases
2.4.3 Promoters and Detractors - Other

2.5 EXPERT OPINION

2.6 OTHER

3.0 KNOWLEDGE ABOUT THE FUNCTIONS OF SCIENCE IN SOCIETY

3.1 UTILITY FOR APPLICATION

3.2 INFORMATION VALUE

3.3 OTHER

4.0 OTHER

5.0 AMBIGUOUS

6.0 NO IDENTIFIABLE KNOWLEDGE DOMAIN

Appendix C: News Briefs Used in the Pilot Study

News Briefs Used in the Pilot Study

DIET

Researchers have reported that colon-cancer patients who have a diet high in mono-casanates show remarkable recovery from chemotherapy. Recovery from chemotherapy has been an especially serious problem for many cancer patients. Members of the Cancer Support Group have hailed this finding and have concluded that <u>this diet is important for countering the negative effects of chemotherapy in colon cancer patients</u>.

DREAMS

Researchers have reported that people who wear ollinite crystals during sleep are more likely to dream accurately about a future event. Public interest in this type of phenomenon has grown in recent years. Members of a human potential group, Mind Matters, have hailed this new finding and have concluded that <u>wearing this type of crystal is important for helping a person to dream accurately about future events</u>.

PESTICIDES

Researchers have reported that robins that have been exposed to the insecticide Permaldrin are much less likely to mate than usual. Robins have declined drastically in numbers over the past several years. Members of Nature Unlimited, an environmental group, have hailed this new finding and have concluded that <u>the use of this insecticide is important for causing a decline in mating behaviors in robins.</u>

TEXTBOOKS

Researchers have reported that high school students in schools where the "Science for Life" textbook is used have higher scores than the provincial norms on science tests. The alarmingly poor performance of Canadian youth in science has become a source of great concern among educators, scientists, engineers, and parents. Members of the educational reform group Learning Now have hailed this new finding and have concluded that <u>this textbook is important for improving high school students' understanding of science.</u>

Appendix D: Ratings Statements

PRACTICING A CERTAIN TYPE OF MEDITATION COULD BE IMPORTANT FOR INCREASING THE SENSE OF WELL BEING IN SENIORS?

a. How likely do you think it is that the statement you just read is true?

Very Unlikely	1	2	3	4	5	6	7	Very Likely

Please explain briefly why you chose this rating.

b. How much experience with or knowledge of the general topic of the above statement do you have?

Little Experience or Knowledge	1	2	3	4	5	6	7	Much Experience or Knowledge

Please explain briefly the nature of your experience or knowledge.

c. How closely related is this statement to topics covered in natural and physical sciences such as Physics, Chemistry, or Biology?

Very Unrelated	1	2	3	4	5	6	7	Very Related

Please explain briefly why you chose this rating.

d. How interested are you in the general topic of the above statement?

Very Uninterested	1	2	3	4	5	6	7	Very Interested

Please explain the nature of your interest.

Appendix E: News Briefs Used in Current Study

Dreams

People have long been interested in what the future holds. Researchers have reported that people who wear Ollinite crystals during sleep are more likely to have dreams that predict the future. Members of Mind Matters have hailed this finding and have concluded that wearing this crystal is important for increasing the frequency of dreams about future events.

Pesticides

People are concerned that declines in wildlife populations will result in extinction for some species. Researchers have reported that robins that have been exposed to the insecticide Permaldrin are less likely to mate than usual. Members of Nature unlimited have hailed this finding and have concluded that using this insecticide is an important factor in causing a decline in robins' mating behavior.

Meditation

People in western countries have long been fascinated by traditional Eastern religious practices. Researchers have reported that senior citizens who practice Mai Handu meditation show an increased sense of well being. members of Lifestyles for Seniors have hailed this finding and have concluded that practices this meditation is important for increasing the sense of well being inn seniors.

Environment

People are concerned about the environmental effects of automobile emissions. Researchers have reported that vehicles that burn gasoline containing the poisonous chemical Quipmanol will reduce existing levels of air pollution. Members of Autos for the Future have hailed this finding and have concluded that fueling vehicles with gasoline containing this poisonous chemical is important for decreasing current levels of air pollution.

Appendix F
Example of Question and Justification Generation Task

Example of Question and Justification Generation Task

People have long been interested in what the future holds. Researchers have reported that people who wear ollinite crystals during sleep are more likely to have dreams that predict the future. Members of Mind Matters have hailed this finding and have concluded that <u>wearing this crystal is important for increasing the frequency of dreams about future events</u>.

Suppose that this conclusion is very important to you and that you must determine whether it is true. Please generate a list of as many questions as you can that you would want to have answered before you decided whether the conclusion made by mind matters is true.

Also, for each question you list, please indicate how you think the answer to that question would help you to evaluate the conclusion in the news brief.

What is the first question you would want answered?

How would an answer to this question help you to decide whether the underlined conclusion in the news brief is true?

What is the next question you would want answered?

How would an answer to this question help you to decide whether the underlined conclusion in the news brief is true?

What is the next question you would want answered?

How would an answer to this question help you to decide whether the underlined conclusion in the news brief is true?

What is the next question you would want answered?

How would an answer to this question help you to decide whether the underlined conclusion in the news brief is true?

What is the next question you would want answered?

How would an answer to this question help you to decide whether the underlined conclusion in the news brief is true?

What is the next question you would want answered?

How would an answer to this question help you to decide whether the underlined conclusion in the news brief is true?

What is the next question you would want answered?

How would an answer to this question help you to decide whether the underlined conclusion in the news brief is true?

What is the next question you would want answered?

How would an answer to this question help you to decide whether the underlined conclusion in the news brief is true?

What is the next question you would want answered?

How would an answer to this question help you to decide whether the underlined conclusion in the news brief is true?

IF YOU HAVE ADDITIONAL QUESTIONS PLEASE LIST THEM ON THE FOLLOWING BLANK PAGE. FOLLOWING EACH QUESTION, EXPLAIN HOW AN ANSWER WOULD HELP YOU DECIDE WHETHER THE UNDERLINED CONCLUSION IN THE NEWS BRIEF IS TRUE.

Appendix G: Questionnaire 1--Background Information

Questionnaire 1--Background Information

Sex M F

Birth Date______________

Year graduated from High School__________

If applicable, please check which course you are currently taking:
______________psychology 104 ______________psychology 105

Please check which best applies to you:

_____1st year university student

_____2nd year university student

_____3rd year university student

_____4th year university student

_____graduated from university but taking some undergraduate courses

Please answer the following questions:

1. What faculty are you in (e.g., Arts, Science, Engineering, etc)?

2. What is you major area of concentration (e.g., Psychology, Math, undecided)?

3. What is you minor area of concentration, if any (e.g., Psychology, Math, English, undecided)?

4a. We are interested in your background in science as a high school student. Please indicate the number of semester-long courses you had in each of the subject in high school listed below. (A two-semester course should be counted as two courses).

Biology 1 2 3 4 5 6 6+

Physics	1	2	3	4	5	6	6+
Chemistry	1	2	3	4	5	6	6+
Psychology	1	2	3	4	5	6	6+
Mathematics	1	2	3	4	5	6	6+
Computers	1	2	3	4	5	6	6+

4b. Please list any other semester-long courses in science you had in high school, including other Natural or Physical Sciences, Social Sciences or General Science courses (give both the name and the total number of semesters you completed, e.g., Sociology: 2 courses).

5. In what province did you receive most of your high school education? (If you were educated out of the country, please indicate the country).

6a. We are interested in you background in science as a university students as well. Please indicate the total number of semesters you are taking or have taken in each of the subjects listed below in university. Give both the total number of courses plus the name of the subject. Also, please indicate when the course was a full year course.
See examples below.

** Note: If you are unsure of where to list a course, simply list it under "other science courses" at the bottom.

NATURAL OR PHYSICAL SCIENCES: (e.g., botany, zoology, physics, chemistry, genetics, biology)

Total Number of Courses______________

(e.g., Zoology: 1 course; Chemistry: 2 half-term courses and 1 full-year course)

SOCIAL SCIENCES (e.g., economics, political science, anthropology, sociology, psychology)

Total Number of Courses______________

MATHEMATICS

Total Number of Courses______________

STATISTICS

Total Number of Courses______________

COMPUTER PROGRAMMING OR COMPUTER SCIENCES

Total Number of Courses_____________

OTHER SCIENCE OR SCIENCE-RELATED COURSES

Total Number of Courses_______________

Appendix H: Questionniare 2--Knowledge about the Limitations of Scientific Knowledge

Imagine the following scenario:

One year ago, a nuclear reactor was built near the area in which you live. Public pressure has forced the managers of the reactor to hire two independent scientists. These scientists are responsible for monitoring radiation levels as well as the effects of radiation on plant and animal life. Each scientist has just released a report containing data and conclusions that contradict one another.

The following five questions refer to the nuclear reactor passage. Please respond TRUE or FALSE to the following statements and explain your responses.

1. If I want to know which report is correct, all I need to do is to look at the data. (T / F)

2. The data that the scientist collects are "facts"; they are precise, reliable and represent the way things truly are. (T / F)

3. I would expect scientists to be able to provide unequivocal information when I ask questions about the risks involved in having a nuclear near my home such as "Is it safe to stay outdoors for several hours at a time?" (T / F)

4. The statement "the levels of radiation have no effect on the plant and animal life" means the same thing as "the level of radiation has an insignificant effect on the plant and life." (T / F)

5. Scientists should be able to precisely predict future radiation levels and the amount of damage to plant and animal life. (T / F)

Appendix I
Questionnaire 3--Interest in Science-Related Activities

1. How interested are you in pursuing a science-oriented career?

   Very Little Interest 1 2 3 4 5 6 7 Very much Interest

   Please elaborate.

2. How interested are you in reading about science-related topics?

   Very Little Interest 1 2 3 4 5 6 7 Very much Interest

   Over the past year, approximately what percentage of your leisure reading was science-oriented?

3. How interested are you in watching TV programs about science-related topics?

   Very Little Interest 1 2 3 4 5 6 7 Very much Interest

   Over the past year, approximately what percentage of your TV watching involved science-related programs?

4. How interested are you in participating in a science oriented hobby, such as electronics, bird watching, astronomy, etc. ?

   Very Little Interest 1 2 3 4 5 6 7 Very much Interest
   If applicable, please explain your interest.

Appendix J: Questionnaire 4--Belief in the Paranormal

Questionnaire 4--Belief in the Paranormal

1. The soul continues to exist though the body may die.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

2. Some individuals are able to levitate objects though mental forces.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

3. Black magic really exists.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

4. Black cats bring bad luck.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

5. Your mind or soul can leave your body and travel (astral projection).

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

6. The abominable snowman of Tibet exists.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

7. Dreams can provide information about the future.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

8. There is a devil.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

9. Psychokinesis, the movement of objects through psychic powers, does occur.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

10. Witches do exist.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

11. If you break a mirror, you will have bad luck.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

12. During altered states, such as sleep or trances, the spirit can leave the body.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

13. The Loch Ness monster of Scotland exists.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

14. Some people have the ability to predict the future.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

15. I believe in God.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

16. A person's thoughts can influence the movement of a physical object.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

17. Voodoo is a real methods to use paranormal powers.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

18. The number "13" is unlucky.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

19. Reincarnation does occur.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

20. Big Foot exists.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

21. The idea of predicting the future is foolish.

1	2	3	4	5
strongly disagree		undecided or don't know		strongly agree

22. There is a heaven and hell.

1 strongly disagree	2	3 undecided or don't know	4	5 strongly agree

23. Mind reading is not possible.

1 strongly disagree	2	3 undecided or don't know	4	5 strongly agree

24. There are actual cases of Voodoo death.

1 strongly disagree	2	3 undecided or don't know	4	5 strongly agree

25. It is possible to communicate with the dead.

1 strongly disagree	2	3 undecided or don't know	4	5 strongly agree