JOURNAL OF RESEARCH IN SCIENCE TEACHING VOL. 36, NO. 4, PP. 411–430 (1999)

Question-Posing Capability as an Alternative Evaluation Method: Analysis of an Environmental Case Study

Yehudit J. Dori, Orit Herscovitz

Department of Education in Technology and Science, Technion, Israel Institute of Technology, Haifa 32000, Israel

Received 1 April 1997; revised 2 March 1998; accepted 16 June 1998

Abstract: An effective strategy for improving problem-solving ability is fostering students' question-posing capabilities through the use of real-world problems. This article describes research on scientific question-posing capabilities among 10th-grade students who were studying air quality in a cooperative way, using the jigsaw method. Case studies and analyses of daily problems and dilemmas were integrated within the module the Quality of Air around Us, which was designed and developed specially for this research. The students were required to pose questions and cope with real-life problems while practicing a variety of learning activities, such as reading press or scientific articles, analyzing tables and graphs, and creating posters and advertisements that related to the problem. The students' question-posing skills were evaluated by using pre- and postcase study questionnaires. We found the number, orientation, and complexity of questions students posed to be three indices of question-posing capability. Following study of the Quality of Air around Us module, a significant increase was observed in the factors of number and complexity of questions students posed. The difference between students at high and low academic levels in the extent of increase in both number and complexity of posed questions was significant. As for orientation, the percentage of solution- and opinion-oriented questions increased in the posttest, and fewer questions dealt with the problem and related hazards. This indicates an increase in students' awareness of the need for and feasibility of seeking practical solutions to a given problem, as well as considerable improvement of their ability to analyze a related case study. On the basis of these findings, we recommend incorporating analysis of question-posing capability as an alternative evaluation method. To this end, fostering of question posing into the case study–based teaching/learning approach is the preferred strategy, in particular when environmental aspects are involved. © 1999 John Wiley & Sons, Inc. J Res Sci Teach 36: 411–430, 1999

Theoretical Background

A major goal of science education is to emphasize the development of students' scientific understanding, thinking, and problem-solving abilities. Current science education emphasizes knowledge (facts and principles) and ignores higher-order thinking (Shepardson, 1993). Nakhleh (1993) pointed out that putting emphasis on algorithmic/mathematical problem solving does not correlate with solving conceptual problems.

Teaching scientific problem solving can be improved if we base our teaching on an adequate understanding of how good problem solving is achieved. Science taught in schools often

---

Correspondence to: Y.J. Dori

CCC 0022-4308/99/040411-20

differs from both actual science and everyday life, and scientific ways of thinking are inadequately taught. Good problem-solving performance includes, among other things, knowledge of how to describe problems effectively, procedures useful for making judicious decisions, and the search for a particular action from among many possible actions that leads to a desired goal (Reif, 1983; Reif & Larkin, 1991).

Thinking and Questioning

In his book *The practice of questioning,* Dillon (1990) cited Socrates from *Republic* VII:534: ". . . so you will make a law that they must devote themselves especially to the technique of asking and answering questions . . ." Dillon went on to say (p. 7) that children everywhere are schooled to become masters at answering questions and to remain novices at asking them. The current practice is for classroom questioning to mean teacher questioning. He distinguished recitation and discussion as two broad types of classroom interaction. These two processes differ in the predominant speaker, type of exchange, sequence, pace, kinds of questions asked, answers given, and evaluations supplied. In the recitation mode the teacher is the predominant speaker, and the interaction involves questions posed by the teacher and answered by the student. In the discussion process, the questions come not just from the teacher's side. Here, both questions and answers are at a higher thinking level, and the answer is not always definite or even known. The interactions are not solely between teacher and students, but also among the students. Students do not demonstrate, but construct, gain, or use knowledge about the matter in question.

Questions are an essential education tool for all disciplines in general and for science in particular. Questions can be rank ordered according to the level of thought required for answering it. The most common hierarchy is the Bloom taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956): knowledge, comprehension, application, analysis, synthesis, and evaluation. Later, other question hierarchies were suggested; the only two levels agreed upon are lower-order and higher-order questions. This categorization is context dependent and is influenced by the objectives or expected learning outcomes of the lesson in which the questions were asked (Barden, 1995). Thus, if a student had previously encountered a similar question, a higher-order question may turn into a lower-order question.

Ashmore, Frazer, and Casey (1979) and Shepardson (1993) regarded question asking as a component of thinking skills for learning tasks and as a stage in the problem-solving process. Ashmore et al. claimed that the range of problems spans from ones that have a unique answer, to the highest level of research work, where the answer must be generated by the solver's observations and/or experimentation. According to Shepardson, each student's task—question or statement—within an activity can be evaluated based on thinking skills needed to complete the task. His scheme is based on that of Marzano et al. (1988), which divides thinking skills into 21 types grouped into eight categories. In this scheme, the skill of formulating questions is categorized under information gathering. Shepardson concluded that science curricula that promote learning must be restructured to engage students in higher levels of thinking about scientific information.

In another research study (Shepardson & Pizzini, 1991), the cognitive level of questions was categorized as one of three types: input, processing, or output. The input-level questions require students to recall information, the processing-level questions cause students to draw relationships among data that have been recalled, and the output-level questions require students to go beyond the data in new ways to hypothesize, speculate, generalize, create, and evaluate. They found that questions in science textbooks, regardless of discipline, emphasize input-level questions and lack enough processing- and output-level questions. An extensive number of input-level questions establishes low-level cognitive purpose and restricts student attention to textual information. This lack of balance causes students to select textual information specific to the question while failing to attain relationships among concepts within the text (Holliday, 1981; Holliday, Whittaker, & Loose, 1984).

Shodell (1995) noted that although the essence of thinking is asking questions, most students perceive science as the study of facts. This is in part because teachers ask fact-demanding questions rather than questions that require thinking. He claimed that the central role of science education and science courses should be to develop in students an appreciation of posing questions. A student question-driven classroom may give students the ability to share the worldview of scientific disciplines and mutually reinforce creativity and higher-order thinking skills. Student questions are thus a valuable alternative to achieving meaningful learning.

Meaningful learning potentially encourages the students and/or teachers to ask the key relevant questions that expose the underlying problems and demonstrate critical thinking (Zoller, 1987, 1990). Hence, there is a strong linkage between Piaget's autonomous learner, who is capable of making well-reasoned intellectual and moral decisions (Piaget, 1965; Bybee & Sund, 1982), and a student who is capable of posing processing- and output-type problems. Therefore, the aim of encouraging students to ask questions and seek solutions to the problems they raise is to educate them to make knowledge-based decisions, take position, and be actively involved (Yager, 1991–1992). This problem exposure process is of the utmost importance if we are to raise the next generation of citizens to be involved in constructive, scientifically grounded activities for the benefit of themselves and their fellow humans.

Alternative Evaluation

Guba and Lincoln (1981) and Nevo (1983) defined evaluation as the integration of description and judgment, in which the description part emphasizes the objective part of the assessment, while the judgment part dwells on its subjective aspect. As Nevo (1994) noted, this definition is especially useful in discussing evaluation perceptions with education practitioners, as it has major implications for developing evaluation mechanisms. Evaluation requires a systematic process, and the application of evaluation skills potentially enhances the objectivity of the descriptive part of the evaluation. Making the judgment procedures explicit and open to scrutiny reduces the arbitrariness of the subjective part.

Lewy (1996) advocated the collection of assessment tasks, which include oral responses, writing essays, performing series of manipulations, and selecting from a list of alternative possibilities. The assessment should include some tasks that need to be performed independently and others that require student cooperation within a working group (Resnick, 1987). Kamen (1996) argued that student alternative assessment has two distinctive features: (a) It includes assessment tasks, all of which provide an alternative to the traditional multiple choice test; and (b) the tasks are geared toward assessing student performance of real-life situations.

Science–Technology–Society (STS), Environmental Education (EE), and Science–Technology–Environment–Society (STES)

Changes in goals for science instruction in schools have induced new orientations in science education worldwide. The two prevalent approaches in modern science education are STS (Bybee, 1985; Aikenhead, 1985; Yager, 1986; Yager & Penick, 1988; Solomon, 1993) and EE (UNESCO, 1986, 1987). Integration of societal and/or environmental issues into the instruction of science and technology subjects makes the latter meaningful and more comprehensible to students.

Keiny and Zoller (1991) summarized the main EE objectives. These include increasing environmental awareness, educating toward responsibility and concern for environmental problems, stimulating the learner's ability and willingness to make a personal contribution, and involving the learner in activities to improve the environment. The two complementary approaches combined gave rise to the fusion of STS and EE, which has yielded the STES paradigm (Zoller, 1991). The ability to integrate concepts from various domains within an interdisciplinary system approach and to act accordingly are key features of STES education. An underlying assumption of STES is that exposing the teachers to a systems approach and appropriate teaching strategies, as well as team projects, will induce teachers to acquire the desired conceptual change.

Integrating science, technology, environment, and society was a major recommendation of the Harari Committee (Harari, 1992) a national committee in Israel whose task was to point the way in which science education in Israel should be shaped in the 21st century. The Harari Committee's main objective was to increase the scientific literacy of high school graduates as well as the number of students electing to major in science. Linking science to social phenomena and to applied technology in everyday life was assumed to make science more relevant and meaningful to students. This was the motivation for the development of the Science, Technology, and Environment in Modern Society (STEMS) project, funded by the Israeli Ministry of Education, which is currently being pursued jointly by three universities in Israel (Dori, Herscovitz, Keiny, Loyd, Gross, & Zoller, 1996; Keiny, 1996; Zoller, 1995).

Teaching/Learning Methods and the Air Quality Project

An STS-oriented curriculum with a strong emphasis on the interrelationship of scientific concepts and real-life phenomena may better serve nonscience students, who have limited scientific knowledge which is nevertheless necessary to prepare them as future citizens (Ben-Zvi & Gai, 1994). Training teachers to incorporate environmental issues within science teaching is an important component of STES education. As such, it has been carried out as part of the STEMS project by several 1-week in-service teacher workshops (Dori, 1995a). The aim of the training was to expose the teachers to core STES-related elements and appropriate teaching methodologies. The teamwork of one science teacher group during the workshop resulted in the development of the Quality of Air around Us module (Herscovitz & Dori, 1996a). This module draws on two well-founded teaching/learning methods that have been integrated. One is the case study approach—an environmental-oriented method—and the other is a cooperative method—jigsaw.

The Case Study Method

Starting at business and medical schools, the case method has become a model for effective learning and gaining the attention of the student audience. Cases are usually real stories, examples for us to study and appreciate if not emulate. They can be close ended, demanding correct answers, or open ended, having multiple solutions because the data involve emotions, ethics, and/or politics. Examples of such open-ended cases include global warming, pollution control, human cloning, and the mission to Mars (Herreid, 1997).

We had prior experience teaching nursing students chemistry and biochemistry using paramedical case studies (Dori, 1994) and teaching elementary and junior high school students science using environment-industrial case studies (Dori, 1995b; Dori & Tal, in press; Tal, Dori, & Lazarowitz, 1997). This experience has shown that the case study method is effective for raising students' concept understanding and critical thinking, as well as their motivation. Following these findings and the fact that our target population consisted of science, environment, and nonscience majors, we elected to apply the case study method in this research. Based on this choice, the concepts conveyed in the module are presented in the form of case studies. The sources for case studies were press and popular science magazine articles, which were used to show the societal and environmental sides of science and its relevance to our daily life (Herried, 1994b). The aspects of the Quality of Air around Us module are diverse and include scientific, environmental, technological, social, economical, and health aspects.

As Herried (1994a, p. 228) noted,

> Although the case study method cannot cure all of the ills in the teaching of science, it is nevertheless ideal for the development of higher-order reasoning skills, which every science teacher claims they strive to instill in their students.

When case studies are used as a supplement to lectures, they serve to deepen student knowledge in societal, environmental, and ethical issues, as well as foster student appreciation of the interdependence of these topics (Challen & Brazdil, 1996). Encouraging students to engage in posing questions and seeking solutions to the problem raised and instructing them as to the optimal ways of retrieving information that is most relevant to the problem at hand are cornerstones of contemporary science education.

The Jigsaw Method

The teaching/learning technique upon which the module is based is the cooperative learning method. To foster students' responsibility for their own topic as well as their peers' study, we decided to use the jigsaw method as the methodological foundation (Aronson, Stephan, Sikes, Blaney, & Snapp, 1978; Herscovitz & Dori, 1996b; Lazarowitz, Hertz-Lazarowitz, Baird, & Bowlden, 1988; Lazarowitz, 1991; Lazarowitz, Hertz-Lazarowitz, & Baird, 1994; Lonning, 1993).

While studying air quality, the students worked in small groups. These groups were called jigsaw groups. Every student within each jigsaw group received one of five different topics. Later, members from different jigsaw groups, who studied the same topic, created five expert groups. Within each expert group, the students learned the topic in relative depth, performed experiments, exchanged ideas, and read case studies while posing questions to their expert peers. Later, the students decided, planned, and prepared materials for teaching their peers. This peer teaching took place when they returned to their original jigsaw group.

The Air Quality Project

The air quality project incorporates a cooperative teaching/learning strategy—the jigsaw method—with a case study–based STES interdisciplinary subject, air quality, and higher-order thinking skills. This method synthesis induces the generation of an autonomous learner through the judicious use of case studies and real-world problems. The concept of autonomy in the learning context dates back to Piaget (1965), as discussed by Kamii (1984), who noted that a child acquires knowledge just as he or she acquires moral values: by constructing it from within, not by internalizing it from the environment. Cooperation is another important factor in generating an autonomous learner, as exchanging points of view contributes positively to children's social, affective, moral, and political development. The autonomous learner is characterized by higher-order thinking skills and a developed capability of posing relevant and constructive questions that illuminate the problem under investigation.

The Quality of Air around Us module encompasses five independent topics: nitrogen oxides (NO_x), sulfur oxides (SO_x) and particles, carbon oxides and the greenhouse effect, ozone layer depletion, and odor: inconvenience or pollution?

The following authoring guidelines were used while developing the Quality of Air around Us module: integration of environmental and social aspects (the STES principle); exploring a familiar (local or global) event as an opening case study for each one of the five topics; using relevant scientific journals or newspaper articles; introducing basic concepts in ecology, such as a system and its component, equilibrium, and positive and negative feedback; and fostering higher-order cognitive skills, critical thinking, asking questions, judging values, solving problems, and using creativity.

The above guidelines are manifested in the module in a variety of ways, including the following instruction tools:

- Directed meaningful reading of information sources
- Posing questions related to a given text
- Analyzing data tables and graphs
- Conducting critical group discussions on given topics and summarizing the different positions
- Playing different roles to illustrate various aspects of a problem, and
- Fostering creative expressions that convey a certain message by creating slogans, commercial-like advertisements, and titles for texts.

To meet the abilities of the heterogeneous class members, the five topics in the Quality of Air around Us module were designed to be at three difficulty levels: high, intermediate, and low. An example of one of the two topics at the high difficulty level is Nitrogen Oxides (NO_x), which includes the following issues:

- Chemistry of NO_x and nitrogen cycle in nature
- Photochemical smog phenomena: formation and hazards
- Sources of air pollution from NO_x, and
- The use of catalytic converter to reduce air pollution from vehicles.

An example of a low difficulty level topic is Odor: Inconvenience or Pollution? which consists of the following issues:

- Experiments with the smell of various everyday materials to determine their level of delight or discomfort
- Air pollution effects associated with bad odors and related health hazards
- Scientific and societal aspects of accidents in chemical plants as a source of air pollution and bad odors, and
- Problems of legislation and law enforcement.

In each one of the five module topics, we included assignments dealing with the scientific, environmental, and societal aspects. While each student studied in depth the topic of his expert group, through peer teaching mandated by the jigsaw method, all students were exposed to the other four topics.

At the end of the module, students were exposed to a scenario about a nearby power plant, environmental problems it created, and possible technological and legislative solutions to these problems. The technological solutions involved gas scrubbers, chimneys, and particle removers such as cyclones, fabric filter baghouses, and electrostatic precipitators. The solutions involved laws, individual fines to violating firm executives, and air quality and emission standards.

Research Objectives and Plan

Guided by the project rationale, we targeted the module for 10th-grade high school students. The two objectives of our research were (a) to examine ways of using question-posing capability as an alternative evaluation tool, and (b) to investigate the effect of the case study teaching/learning approach on the question-posing capability of high school students at different academic levels. The research problem was whether and how students' question-posing capabilities, enhanced through the case study approach, can be used as an alternative evaluation method.

Population

The module was implemented experimentally during the 1995 and 1996 academic years in seven 10th-grade classes in Israel. The research population, presented in Table 1, included seven intact 10th-grade classes from five schools of different types in the northern part of Israel. Tenth-grade students in Israel are required to take at least one scientific course. Science major students choose one or more of the three courses—physics, chemistry, or biology. Following the Harari Committee recommendations (Harari, 1992), nonscience majors can choose the course entitled "Science and Technology for All." All the teachers who participated in the STEMS project and consented to teach the Quality of Air around Us module were the teachers of classes who took part in the research. Hence, there was no preselection of the research population, except for the teachers' willingness to teach the topic.

Based on a classification made by each school management, the student population was divided into three academic levels: high (H) (science majors), medium (M) (average students), and low (L) (students with some learning difficulties). Science majors (H-level students) studied the module for extra credit, while M and L students took it as the Science and Technology for All course. All classes are in general heterogeneous, but the average scientific and academic levels of the three types reflect their classification into the three types. The classification of courses into the three levels by the schools was verified through the scientific literary (second) part of the pretest. A Kruskal–Wallis test yielded a significant difference among the three levels, $\chi^2 = 43.9$; $p < .0001$.

Table 1
Research population background

Year	School type	Class type	Academic level	No. of students
Winter 1995/6	Urban	Scientific	High	20
		Scientific	High	18
	Agricultural I	Scientific	High	21
		Environmental	Medium	16
Spring 1995	Agricultural II	Environmental	Medium	13
	Rural I	Nonscientific	Low	19
	Rural II	Nonscientific	Low	20

Research Tools

A knowledge pretest administered to the research population before starting the study of the Quality of Air around Us module confirmed this classification. The students studied the module during 30 academic hour sessions over a period of 2 months.

The development and implementation processes of the Quality of Air around Us module were accompanied by a formative evaluation focused on question posing and case study analysis capabilities before studying the module and afterward. The pre- and posttest questionnaires that were administered to the high school students consisted of two parts. The first part included a half-page article describing a case study. The pretest case study dealt with rain forests, while the posttest concerned the threat of health hazard problems caused by the ozone layer in the Krayot, a highly populated urban area north of Haifa, Israel. The “bad” ozone, according to the local newspaper, is created by polluting chemical plants in this area. After reading the case study, the responders were required to carry out a task which was phrased as follows: “You are asked to design an assignment for your friends. Compose as many questions as you can about the case study you have read.”

The second part included 15 multiple choice questions. In the pretest, these questions were related to prior knowledge in chemistry and biology. In the posttest, three questions related to each of the five topics included in the air quality module.

The contents of the pretest and the posttest were validated regarding clarity, lack of ambiguity, and consistency of difficulty levels across the five topics. The validation was done by 4 doctoral students who were also experienced high school teachers from the areas of chemistry (2 teachers), biology, and environment. Additional validation was done by a researcher in our department. Following this validation, some items were removed while others were modified or added. The reliability of the second part of both tests was α Cronbach = .7.

Data Analysis

The first part of the pre- and posttest questionnaires, which included the case study, was followed by an assignment in which the student was asked to compose as many questions as he or she could ask about the case study. To assess the effect of the case study method on students’ question-posing capability, the results of both the pre- and posttest case studies of all the responders were analyzed according to (a) the number of questions posed by each student, (b) the orientation of each question, and (c) the complexity of each question.

First, the questions posed by students in the pre- and posttest questionnaire were counted. The questions were then analyzed according to three categories: orientation, relation to the case study, and complexity. Each category was classified into several binary attributes, which are listed in Table 2. A binary (Boolean) attribute can have one of two values: positive (“yes”) or negative (“no”), as exemplified below.

The three possible orientation attributes of a question are (a) phenomenon and/or problem description, (b) hazards related to the problem, and (c) treatment and/or solution. A positive value of the attribute “phenomena and/or problem description,” for example, means that the question composed by the student refers to a phenomenon described and/or a description of a problem raised in the case study.

If one of the relation category attributes was “The answer is partially in the article,” or “The answer is not in the article,” then at least one of the attributes in the complexity category was assigned a positive value. This ensured that a question was categorized as complex only if the answer to it was at least partially not in the case study, and hence required higher-order think-

Table 2
Classification of questions, by categories and attributes

Category	Attribute
Orientation	• Phenomena and/or problem description • Hazards • Treatment and/or solution
Relation to case study	• The answer is provided in the case study. • Part of the answer is provided in the case study. • The answer cannot be found in the case study.
Complexity	• Application and/or analysis • Interdisciplinary approach • Judgment and/or evaluation • Taking position and/or personal opinion

ing skills. Put another way, if at least one of the last two attributes in the relation category scored a positive value, then the question was classified as complex. Note that the classification referred to the case study rather than to the way the student phrased the question.

The attributes in the complexity category were influenced by Bloom's taxonomy (Bloom et al., 1956), thinking skills classification (Shepardson & Pizzini, 1991; Shepardson, 1993), and criteria for problem solving (Zoller, 1987).

To validate the consistency of the classification scheme, a random sample of 10% of the questions collected from the students was given to six expert judges: two Ph.D.s and four doctoral students. These science educators were asked to classify the questions, i.e., assign values to the question attributes. All six expert judges fully agreed (100%) on the value assigned to the attributes of orientation and relation to the case study categories. The agreement among the expert judges on attributes of the complexity category was 85%. Following this validation, a training session of 2 h on question classification and attribute values assignment was conducted. All student questions were then classified according to the pattern that emerged from the expert judges' classification.

To illustrate the considerations involved in assigning attribute values to questions, let us examine two questions. For the question, "What is the 'bad' ozone?" the attribute to which a positive value was assigned in the orientation category is problem/phenomenon description. In the relation (to the case study) category, the positive attribute is, "The answer is in the article," and in the complexity category, the value is zero since it requires only knowledge.

A more involved example is the question, "Is it possible in your opinion to stop the spread of pollution and eliminate the ozone depletion problem?" which was asked by several students in the posttest. This question scored a positive value in the following attributes: "possible solutions to the problem" of the orientation category, "the answer is partially in the article" of the relation category, and hence it also scored positively in the "analysis," "judgment and/or evaluation," and "taking position" of the complexity category.

Figure 1 was taken during the jigsaw interaction in one of the classes from an agriculture school that took part in the research. The text in the callout bubbles represents a sample of questions students posed at the postcase study questionnaire. Following is the analysis of attributes for questions recorded in Figure 1, by categories. Enclosed in parentheses next to each question categorization are the question scores, the calculation of which is explained in the sequel.

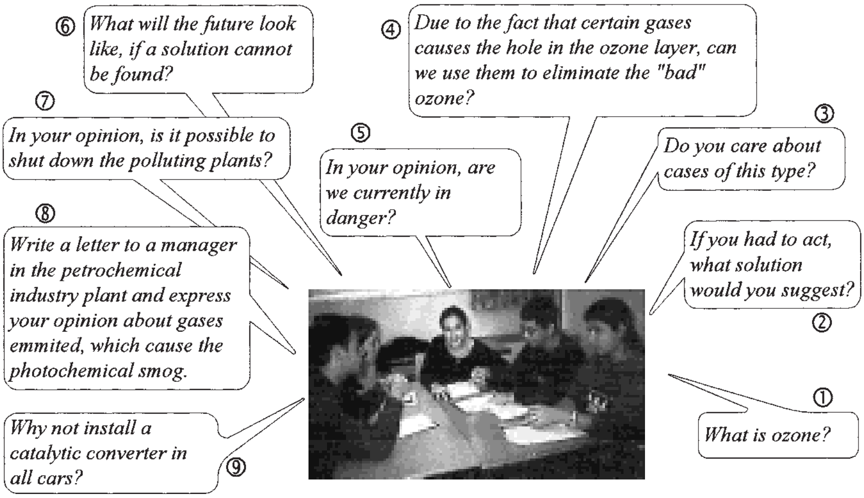

Figure 1. Examples of questions asked in the postcase study questionnaire.

Question 1: Orientation category, “Problem description”
Relation category, “The answer is in the article”
Complexity category, “Knowledge” ($p_i = 0$)
Question 2: Orientation category, “Possible solutions”
Relation category, “The answer is not in the article”
Complexity category, “Application and/or analysis” ($p_i = 1$)
Question 3: Orientation category, “Problem description”
Relation category, “The answer is not in the article”
Complexity category, “Taking position” and “Judgment and/or evaluation” ($p_i = 2$)
Question 4: Orientation category, “Possible solutions”
Relation category, “The answer is not in the article”
Complexity category, “Application and analysis,” “Judgment and/or evaluation,” and “Interdisciplinary approach” ($p_i = 3$)
Questions 5 and 6: Orientation category, “Problem description”
Relation category, “The answer is not in the article”
Complexity category, “Judgment and/or evaluation” and “Taking position” ($p_i = 2$)
Question 7: Orientation category, “Possible solutions”
Relation category, “The answer is not in the article”
Complexity category, “Taking position,” “Judgment and/or evaluation,” and “Interdisciplinary approach” ($p_i = 3$)
Question 8: Orientation category, “Problem description”
Relation category, “The answer is not in the article”
Complexity category, “Taking position” ($p_i = 1$)
Question 9: Orientation category, “Possible solutions”
Relation category, “The answer is not in the article”
Complexity category, “Application and/or analysis” and “Interdisciplinary approach” ($p_i = 2$)

Student Aggregate Question Complexity Calculations

The student aggregate question complexity scores regarding the case study were calculated for each student individually according to the values assigned to each question attributes as defined above.

$$C = \sum_{i=1}^{n} (p_i + 1) = n + \sum_{i=1}^{n} p_i \qquad (1)$$

C is the student aggregate question complexity score, n is the number of questions asked by the student, and $p_i \in \{0,1,2,3\}$ is the number of attributes which scored a positive value in the complexity category of question i. Thus, if a question had no positive attributes in the complexity category (i.e., it is a knowledge question), $p_i = 0$, it scored 1 point ($p_i + 1 = 1$), while if all three attribute values were positive, it scored 4 points. More concrete examples are listed in the Question Categorization subsection above. Referring to these examples, a student who came up with the five questions whose numbers (as in Figure 1) are 1 ($p_i = 0$), 2 ($p_i = 1$), 4 ($p_i = 3$), 6 ($p_i = 2$), and 9 ($p_i = 2$) scored a total of $C = 5 + (0 + 1 + 3 + 2 + 2) = 13$ points.

Results

The number of questions posed by students at the three academic levels is summarized in Table 3. As can be seen, the average number of questions asked by students at each level in the final case study questionnaire—the posttest—was more than twice the number asked in the pretest. The gain between the average number in questions students posed at the three academic levels from the pretest to the posttest increased along with the academic level. The increase was 2.33, 2.93, and 3.85 for the low, medium and high academic levels, respectively.

Not only did the number of questions increase, their variability, i.e., the number of different questions posed by all students, increased in a similar proportion: 235 different questions were asked by all the students in the posttest, compared with 116 questions in the pretest. The distribution of the number of different types of questions in the pretest among the three academic levels H, M, and L was 52, 38, and 26 compared with 115, 67, and 53 in the posttest, respectively. This is a respective increase of 121%, 76%, and 104%.

Table 3
Mean scores and standard deviations of number of questions students posed in pre- and postcase study questionnaires, by academic levels

Level	Test type	No. of students	$\bar{X}$	*SD*	Maximum questions
H	Pre	59	2.53	0.99	4
	Post	56	6.38	2.13	12
M	Pre	29	2.31	1.49	6
	Post	29	5.24	1.53	9
L	Pre	39	2.05	1.12	6
	Post	29	4.38	1.39	9

The null hypothesis was that there is no significant difference between the average number of questions posed by each student before and after the course. To test this hypothesis, Wilcoxon single rank tests were conducted. The results ($p < .0001$) indicated that the difference was significant.

To test whether the differences in average number of posed questions among the three academic levels were significant, we conducted a series of Kruskal–Wallis tests (χ^2 approximation). The results, summarized in Table 4, indicated that overall, there was a significant difference among the three levels regarding the average number of questions students asked in both the pretest and the posttest. When comparing pairs of levels, we found that the difference between the average number of questions asked by H- and L-level students was significant in both the pretest and the posttest. However, a significant difference between H and M and between M and L existed only in the posttest but not in the pretest. In other words, the students' entry-level question-posing capability within each pair of conjugate levels (H–M and M–L), measured by the average number of questions a student could pose, was not significantly different. However, after the treatment, i.e., the case-based course, the difference between each level within both pairs of conjugate levels was significant. This indicates that the treatment, i.e., exposure to and dealing with the air quality subject, improved the question-posing capability of students and may also serve as a diagnostic tool to differentiate among academic levels.

To test whether the difference in the extent of the increase in the average number of questions among the three academic levels in the pretest and posttest is significant, we conducted another series of Kruskal–Wallis tests (χ^2 approximation). The results indicated that overall, there was a significant difference in the extent of increase in the average number of questions among the three academic levels, ($\chi^2 = 7.29$; $p < .026$). This difference was entirely due to the significant difference between levels H and L, $\chi^2 = 6.81$; $p < .009$, as no significant difference was found between H and M levels, or between M and L levels. In other words, the improvement in the question-posing capability of H-level students, as measured by the average number of questions posed before and after the course, was significantly higher than that of L-level students, but the improvement in both H versus M levels and M versus L levels was the same.

Question Orientation

About half of the pretest student questions were concerned with the description of the hazards caused by the problem or phenomenon presented in the case study. As Table 5 shows, few-

Table 4
Kruskal–Wallis test (χ^2 approximation) for mean number of questions per student among levels

Between levels	Test type	χ^2	*df*	*p*
H, M, L	Pre	6.75	2	.0343
	Post	19.51	2	.0001
H, L	Pre	7.81	1	.0052
	Post	16.26	1	.0001
H, M	Pre	0.86	1	.3538
	Post	3.78	1	.0520
M, L	Pre	0.67	1	.4126
	Post	7.46	1	.0063

Table 5
Distribution of question orientation in pre- and postcase study questionnaires, by academic level

Level	Test type	No. of questions	Phenomenon	Hazards	Solution
H	Pre	150	34%	45%	21%
	Post	358	43%	24%	32%
M	Pre	68	26%	52%	22%
	Post	154	46%	24%	30%
L	Pre	80	41%	45%	14%
	Post	127	41%	24%	35%

er questions dealt with the description of the phenomenon itself, and yet fewer questions related to possible solutions to the problem. However, the corresponding proportions in the posttest were quite different, as more questions dealt with possible solutions and fewer with hazards related to the phenomenon. This indicated an increase in students' awareness of the need for and feasibility of seeking practical solutions to a given problem. The average number of questions involving value judgment and evaluation increased from 16% in the pretest to 36% in the posttest, and questions that involved taking a position or criticizing rose almost threefold from 8% in the pretest to 22% in the posttest.

Question Complexity

Table 6 shows a significant increase ($p < .0001$) in the mean question complexity scores in the posttest relative to the pretest at all three academic levels. Another observation is that the highest complexity scores in both pre- and posttests were similar at all three levels. The range of question complexity scores between pre- and posttests decreased from H to M and from M to L academic levels. This indicates that high achievers benefited most from this type of learning.

To test whether the difference in the extent of increase in the average complexity of questions among the three academic levels in the pretest and posttest is significant, we conducted yet another series of Kruskal–Wallis tests (χ^2 approximation). The result indicate that overall, there was a significant difference in the extent of increase in the average complexity of questions among the three academic levels, $\chi^2 = 11.08$; $p < .0004$. This difference was entirely due to the significant difference between levels H and L ($\chi^2 = 11.26$; $p < .0008$), as no significant

Table 6
Mean scores, range, and standard deviations of question complexity in pre- and postcase study questionnaires, by level

Level	Test type	*n*	$\bar{X}$	*SD*	Maximum	*p*
H	Pre	59	3.71	1.90	8	.0001
	Post	56	10.23	3.80	20	
M	Pre	29	3.59	2.80	10	.0001
	Post	29	8.79	3.50	18	
L	Pre	39	2.85	1.61	8	.0001
	Post	29	6.31	4.18	18	

difference was found between H and M levels or between M and L levels. This indicates that the improvement in question-posing capability of H-level students, as measures by the average complexity of questions posed before and after the course, was significantly higher than that of L-level students. The improvements of H versus M levels and M versus L levels was not significant. This pattern is consistent with our findings regarding the number of questions. The number of questions posed and their complexity are two complementary indices of question-posing capability.

Question Complexity Analysis

The questions were classified into three levels of complexity:

- Low-complexity questions: questions at the knowledge level, i.e., the answer to these questions is entirely in the article, such that only understanding the text is required to answer them correctly.
- Intermediate-complexity questions: questions with one positive complexity attribute.
- High-complexity questions: questions with more than one positive complexity attribute.

By comparing the percentage of students who asked low-complexity questions with those who asked high-complexity questions for each academic level, we can show that as a result of this project, students improved their question-posing capabilities between the pre- and postcase study.

The percentage of knowledge questions decreased by a factor of about 3 from the pretest to the posttest for both H students (from 36% to 11%) and M students (from 36% to 14%). For low-ability students, the percentage increased by about a third (from 35% to 45%). However, as Figure 2 shows, the percentage of high-complexity questions increased from the pretest to the posttest by a factor of about 6 (from 14% to 81%) for H students, 4 for M students (from 17% to 66%), and 12 for L students (from 3% to 36%).

While the number of high-complexity questions asked by H, M, and L students in the pretest—14, 17, and 3, respectively—did not show a clear trend (except for the large difference between H and M students versus L students), the corresponding numbers of posttest figures—

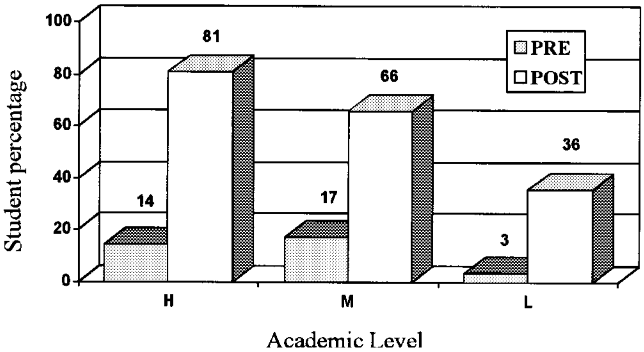

Figure 2. Percentage of students posing questions at high complexity level. Two attributes or more of the complexity category.

81, 66, and 36—were in accord with the predefined classification of the students, which, as noted, was carried out both by the school principals and through the use of the preliminary knowledge test. This outcome strongly indicates the potential of posing questions as a viable evaluation tool that offers an alternative to conventional evaluation methods.

Student Achievements: Pre- versus Posttest and Expert versus Other Topics

Student achievements for the air quality topic before and after students studied the air quality module were tested using the second (achievement) part of the pre- and posttests. The results for the H, M, and L academic levels were 56.0, 41.7, and 38.0, respectively, in the pretest ($n = 137$) and 80.4, 65.0, and 56.5, respectively, in the posttest ($n = 131$). The improvement was 44%, 56%, and 49%, respectively.

To analyze the source of difference in students' knowledge, an analysis of covariance was conducted with student achievements in the pretest serving as the covariate. The results showed that the treatment, i.e., study of the air quality module, $F_{123} = 18.3$; $p < .0001$, but not the prior knowledge, $F_{123} = 0.49$; not significant, affected students' knowledge. In addition, a significant difference, $F_{123} = 21.05$; $p < .0001$, was found among the three academic levels. This is in accord with the significant difference found in the increased complexity of questions among the three academic levels in the posttest compared to the pretest (see Table 6).

We investigated whether there was a difference between students' knowledge of their expert topic and knowledge of other topics. To this end, we defined success as correctly answering at least two thirds of the 15 questions of the achievement part of the posttest. Hence, success in one's expert topic was defined as correctly answering at least two of the three questions in that topic. Success in the four other topics combined was defined as correctly answering 8 of the remaining 12 questions.

Figure 3 shows that the achievements of all students were high in the expert topic. In the other topics, only students at the high academic level maintained the success rate of the expert topic in the other topics. Achievements of students at the medium and low levels declined proportionately. A significant difference was found between the success of H-level students and Levels M and L, $\chi^2 = 5.99$; $p < .014$.

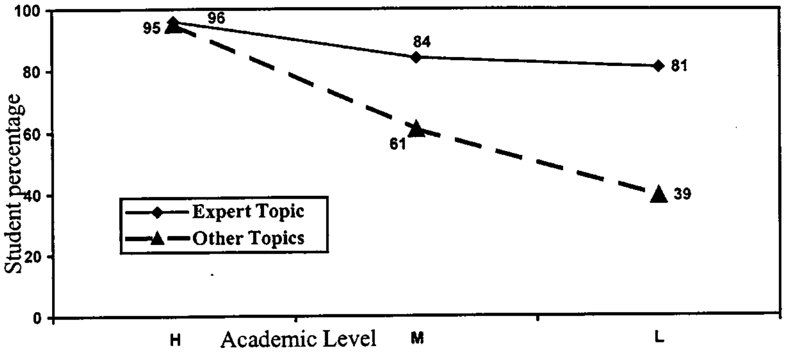

Figure 3. Student success frequencies in the expert topic and in other topics in the posttest, by academic level.

Summary of Findings and Their Limitations

We summarize our research findings below. A discussion of their limitations follows.

- A significant increase from the pretest to the posttest was observed in both the number and complexity of questions students posed. An analogous significant difference was found between student achievements.
- The difference between students at high and low academic levels regarding the extent of increase in number, complexity of posed questions, and achievements was significant. Those at the high academic level increased both the number and complexity of questions posed to a larger extent than the low academic level students.
- The percentage of solution- and opinion-oriented questions increased in the posttest, and fewer questions dealt with the problem and related hazards.
- Each academic level group increased its knowledge in the posttest compared to the pretest, by about 50%.
- High academic level students maintained the same level of knowledge in their expert topic as in the other topics. The knowledge level of medium and low academic level students declined significantly and proportionately in other topics compared to the expert topic.

All students were exposed to all five topics of the air quality module, either in depth or through peer teaching. All five groups were required to cope with scientific, technological, environmental, and societal aspects of their expert topic, as well as the other four topics. However, a limitation of the experiment may arise from the nature of topics studied by the high academic student group compared to the topics studied by the two other groups. It is possible that the nature of the topics studied by the high-level group lent itself to easier transfer to the novel topic given on the posttest than the topics studied by the middle- and low-level groups. This is true for the achievements but not for the questions students posed after reading the case study. There, we examined the complexity and orientation of questions, but not their content.

Discussion and Recommendations

Linking science to social phenomena and to applied technology in everyday life is assumed to make science more relevant and meaningful to students. Shepardson (1993) claimed that science curricula that promote learning must be restructured to engage students in higher levels of thinking about scientific information. The range of problems spans from the ones that have a unique answer to questions that must be generated by the observations and/or experimentation.

Question asking is a component of thinking skills for learning tasks and a stage in the problem-solving process. Questions should come from not only the teacher’s side, but also that of the student (Dillon, 1990). Shepardson and Pizzini (1991) characterized the cognitive level of questions as one of three types: input, processing, or output. They found that questions in science textbooks, regardless of discipline, emphasize input-level questions and establish a low-level cognitive purpose. In our research, processing-type questions—drawing relationships—correspond to the interdisciplinary approach and application and analysis attributes of the complexity question category. Output-type questions, which according to Shepardson and Pizzini imply speculation, generalization, creation, and evaluation, correspond in our research to taking a position and judging and/or evaluating complexity attributes.

Shodell (1995) claimed that the central role of science education and science courses should be to develop in students an appreciation for posing questions. Questions at a high complexity level were generated as a result of student–student interaction and construction of new knowledge while being exposed to new learning situations through case studies.

As part of Science, Technology, and Environment in Modern Society (STEMS) project, science teachers developed the module The Quality of Air around Us, which incorporates case studies that are studied through the jigsaw cooperative learning method.

The research problem was whether and how students' question-posing capabilities, enhanced through the case study approach, can be used as an alternative evaluation method.

We examined the effect of the air quality project on students' question-posing capabilities, whose three indices were the number, orientation, and complexity of questions a student posed before and after the treatment. To determine the questions' complexity as systematically and objectively as possible, we developed and applied a quantitative method for calculating the complexity of each question and the student aggregate question complexity. In a nutshell, the method calls first for determining whether the question requires only knowledge present in the case study. If not, each one of four complexity binary attributes is assigned a value, which is positive if the attribute characterizes the question and negative otherwise. The number of positive complexity values plus 1 is the question complexity score. Summing all the scores of questions a student asked yields the student aggregate question complexity.

The results show that the air quality project brought about a significant increase in students' question-posing capability. The increased capability is significant at all three academic levels and is reflected in the aspects of total number of questions, orientation, and complexity. Following Kamen (1996), we included assessment tasks as an alternative to the traditional test by combining a case study part and a multiple choice knowledge part. As Kamen suggested, our tasks were geared toward assessing student performance of real-life situations.

Through the study of the air quality topic, students gained a more complex view of real-world problems. These include societal and economical considerations of locating a chemical plant in a nearby area, the underlying scientific facts about gases the plant emits, and the option to equip it with an advanced filtering system. Bringing students to understand conflicts such as those presented in the case studies may encourage them to read a daily or scientific article critically and question the quality of the given information. The improvement in question-posing capability indicates an improvement in the students' verbal expressive power, and this finding is in accord with those of Lonning (1993) and Lazarowitz, Hertz-Lazarowitz, and Baird (1994).

The significance of the improvement in question-posing capability indicates that this capability can be used effectively as an alternative evaluation tool for assessing the extent to which students understand and analyze a topic and make a value judgment regarding a related case study. This would add another dimension to conventional student evaluation and improve the achievements of low academic level students, who usually get low scores in knowledge tests.

The contributions of the research concern several aspects. One is the use of student question-posing capability as a means for evaluating higher-order thinking skills. Another contribution related to the fact that students at all three academic levels significantly improved both their knowledge and their question-posing capability. The fact that high-level students improved more than the other two groups does not undermine the importance of our approach, since what really matters to us as science educators is the improvement of the student's scientific and technological literacy with respect to his or her starting point.

Fostering students' understanding of conflicts, such as those included in the Quality of Air around Us module, may lead 21st-century citizens to critically read articles and advertisements dealing with phenomena that require knowledge of science and technology. If they can question the quality of the given information and understand that some of the problems may have more than one possible solution, or may not have a solution at all, then our efforts to improve students' STES literacy are rewarded.

On the basis of these findings, we recommend incorporating an analysis of question-posing capability as an alternative evaluation method. To this end, fostering of question posing into the case study–based teaching/learning approach is the preferred strategy, in particular when environmental aspects are involved. More research is needed to determine and analyze the potential of students to respond to the questions they themselves raise as a result of engaging in the case study learning method. It is interesting to investigate whether a correlation exists between their question-posing capability and their ability to suggest possible solutions to the conflicts or problems that arise.

This research was supported by the Fund for the Promotion of Research at the Technion.

References

Aikenhead, G.S. (1985). Collective decision-making in the social context of science. Science Education, 69, 453−475.

Aronson, E., Stephan, C., Sikes, J., Blaney, N., & Snapp, M. (1978). The jigsaw classroom. Beverly Hills, CA: Sage.

Ashmore, A.D., Frazer, M.J., & Casey, R.J. (1979). Problem solving and problem solving networks in chemistry. Journal of Chemical Education, 56, 377−379.

Barden, M.B. (1995). Effective questioning and the ever-elusive higher-order questions. The American Biology Teacher, 57, 423−426.

Ben-Zvi, N., & Gai, R. (1994). Macro- and micro-chemical comprehension of real-world phenomena. Journal of Chemical Education, 71, 730−732.

Bloom, B.S., Engelhart, M.B., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain. New York: Longmans Green.

Bybee, R.W. (1985). The Sisyphean question in science education: What should a scientifically and technologically literate person know, value and do . . . as a citizen? In R.W. Bybee (Ed.), Science–Technology–Society 1985 NSTA Yearbook. Washington, DC: NSTA Publications.

Bybee, R.W., & Sund, R.B. (1982). Piaget for educators (2nd ed., pp. 161–184). Columbus: Charles E. Merrill Publishing Company.

Challen, P.R., & Brazdil, L.C. (1996). Case studies as a basis for discussing method teaching in introductory chemistry courses. The Chemical Educator, 1. http://journals.springer-ny.com/chedr.

Dillon, J.T. (1990). The practice of questioning. London: Routledge.

Dori, Y.J. (1994). Achievement and attitude evaluation of a case-based chemistry curriculum for nursing students. Studies in Educational Evaluation, 20, 337−348.

Dori, Y.J. (1995a). In-service training for leadership: Incorporating environmental aspects into chemistry teaching. In Proceeding of the 68th annual National Association for Research in Science Teaching Conference, San Francisco, CA.

Dori, Y.J. (1995b). Education towards awareness and understanding of the relationship between industry and environment. In R. Tal & H. Ackerman (Eds.), Industry and manufacturing: The central theme for 1996. Ministry of Education, Culture, and Sport, Center for Environmental Education in Kfar Vradim, Tefen Industrial Park and Iscar (pp. 20−29) (in Hebrew).

Dori, Y.J., Herscovitz, O., Keiny, S., Loyd, D., Gross, M., & Zoller, U. (1996). Science, technology and environment in modern society: A national project in Proceedings of the Second Jerusalem International Science and Technology Education Conference, Jerusalem, Israel.

Dori, Y.J., & Tal, R. (in press). Industry and environment projects: Informal science activities in a community school. Science Education.

Guba, E.G., & Lincoln, Y.S. (1981). Effective evaluation. San Francisco: Jossey-Bass.

Harari, H. (1992). Tomorrow ’98: Report of the Supreme Committee on Science and Technology Education. Ministry of Education, Culture, and Sport, Jerusalem, Israel, pp. 36−37 (in Hebrew).

Herried, C.F. (1994a). Case studies in science: A novel method of science education. Journal of College Science Teaching, 23, 221−229.

Herried, C.F. (1994b). Journal articles as case studies The New England Journal of Medicine in Breast Cancer: Promoting critical thinking in science. Journal of College Science Teaching, 23, 349−355.

Herried, C.F. (1997). What is a case? Bringing to science education the established tool of law and medicine. Journal of College Science Teaching, 26, 92−94.

Herscovitz, O., & Dori, Y.J. (1996a). The quality of air around us. Department of Education in Technology and Science, Technion IIT, Haifa, Israel (in Hebrew).

Herscovitz, O., & Dori, Y.J. (1996b). Cooperative learning: Application of the jigsaw method while studying acids and bases in chemistry laboratory. Halacha Lemaashe—from Theory to Practice in Curriculum Development, 11, 127−150 (in Hebrew).

Holliday, W.G. (1981). Selective attentional effects of textbook study questions on student learning in science. Journal of Research in Science Teaching, 18, 283−289.

Holliday, W.G., Whittaker, H.G., & Loose, K.D. (1984). Differential effects of verbal aptitude and study questions on comprehension of science concepts. Journal of Research in Science Teaching, 21, 140−143.

Hunter, B. (1993). Internetworking: Coordinating technology for systematic reform. Communications of the ACM, 36, 42−46.

Kamen, M. (1996). A teacher’s implementation of authentic assessment in an elementary science classroom. Journal of Research in Science Teaching, 33, 859−877.

Kamii, C. (1984). Autonomy: The aim of education envisioned by Piaget. Phi Delta Kappa, 65, 410−415.

Keiny, S. (1996). A community of learners: Promoting teachers to become learners. Teachers and Teaching: Theory and Practice, 2, 243−272.

Keiny, S., & Zoller, U. (Eds.). (1991). Conceptual issues in environmental education. New York: Peter Lang.

Lazarowitz, R. (1991). Learning biology cooperatively: An Israeli junior high school study. Cooperative Learning, 11, 19−21.

Lazarowitz, R., Hertz-Lazarowitz, R., & Baird, J.H. (1994). Learning science in a cooperative setting: Academic achievement and affective outcomes. Journal of Research in Science Teaching, 31, 1121−1131.

Lazarowitz, R., Hertz-Lazarowitz, R., Baird, J.H., & Bowlden, V. (1988). Academic achievement and on task behavior of high school biology students instructed in a cooperative small investigative group. Science Education, 72, 475−487.

Lewy, A. (1996). Postmodernism in the field of achievement testing. Studies in Educational Evaluation, 22, 223−244.

Lonning, R.A. (1993). Effect of cooperative learning strategies on student verbal interaction and achievement during conceptual change instruction in tenth grade general science. Journal of Research in Science Teaching, 30, 1087−1101.

Marzano, R.J., Brandt, R.S., Hughes, C.S., Jones, B.F., Presseisen, B.Z., Rankin, S.C., & Schor, C. (1988). Dimensions of thinking: A framework for curriculum and instruction. Alexandria, VA: Association for Supervision and Curriculum Development.

Nakhleh, N.B. (1993). Are our students conceptual thinkers or algorithmic problem solvers? Journal of Chemical Education, 70, 52−55.

Nevo, D. (1983). The conceptualization of educational evaluation: An analytical evaluation of the literature. Review of Educational Research, 53, 117–128.

Nevo, D. (1994). Combining internal and external evaluation: A case for school-based evaluation. Studies in Educational Evaluation, 20, 87–98.

Piaget, J. (1965). The moral judgement of the child. New York: Free Press.

Reif, F. (1983). How can chemists teach problem solving? Journal of Chemical Education, 60, 948–953.

Reif, F., & Larkin, J.H. (1991). Cognition in scientific and everyday domains: Comparison and learning implications. Journal of Research in Science Teaching, 28, 733–760.

Resnick, L.B. (1987). Learning in school and out. Educational Researcher, 16, 13–20.

Shepardson, D.P. (1993). Publisher-based science activities of the 1980's and thinking skills. School Science and Mathematics, 93, 264–268.

Shepardson, D.P., & Pizzini, E.L. (1991). Questioning levels of junior high school science textbooks and their implications for learning textual information. Science Education, 75, 673–682.

Shodell, M. (1995). The question-driven classroom: Student questions as course curriculum on biology. The American Biology Teacher, 57, 278–281.

Solomon, J. (1993). Teaching Science, Technology and Society (pp. 11–36). Buckingham, Philadelphia: Open University Press.

Tal, R., Dori, Y.J., & Lazarowitz, R. (1997). Industry–environment projects: An evaluation approach in environmental education. In Proceedings of the Second International Conference on Teacher Education: Stability, Evolution, and Revolution, Israel, pp. 175–187.

UNESCO-UNEP. (1986). Environmental Education: A process for pre-service teacher training curriculum development. UNESCO-UNEP-IEEP, Environmental Education Series, No. 25, Paris.

UNESCO-UNEP. (1987). Strategies for the training of teachers in environmental education. UNESCO-UNEP-IEEP, Environmental Education Series, No. 26, Paris.

Yager, R.E. (1986). Restructuring science teacher education programs as they move towards an S/T/S focus. In R.K. James (Ed.), Science, technology and society: Resources for science educators (pp. 46–55). Columbus, OH: SMEAC Information Reference Center, Ohio State University.

Yager, R.E. (1991–1992). Basic science of STS: Which is better for science learning? The case for STS as reform. NSTA Handbook of STS, 1, 8–13.

Yager, R.E., & Penick, J.E. (1988). Changes in perceived attitudes towards the goals for science instruction in schools. Journal of Research in Science Teaching, 25, 179–184.

Zoller, U. (1987). The fostering of question-asking capability. Journal of Chemical Education, 64, 510–512.

Zoller, U. (1990). The IEE: An STS approach. Journal of College Science Teaching, 19, 289–291.

Zoller, U. (1991). Teaching/learning styles, performance, and students' teaching evaluation in S/T/E/S-focused science teacher education: A quasi-quantitative probe of a case study. Journal of Research in Science Teaching, 28, 593–607.

Zoller, U. (1995). STEMS (Science, Technology, and Environment in Modern Society). In Proceeding of the 68th annual National Association for Research in Science Teaching Conference, San Francisco, CA.