THE JOURNAL OF THE LEARNING SCIENCES, *12*(2), 145–181

Higher Order Thinking Skills and Low-Achieving Students: Are They Mutually Exclusive?

Anat Zohar
School of Education
Hebrew University

Yehudit J. Dori
Department of Education in Technology and Science
Technion, Israel Institute of Technology
and
Center for Educational Computing Initiatives
Massachusetts Institute of Technology

Fostering students' higher order thinking skills is considered an important educational goal. Although learning theories see the development of students' thinking as an important goal for all students, teachers often believe that stimulating higher order thinking is appropriate only for high-achieving students. According to this view, low-achieving students are, by and large, unable to deal with tasks that require higher order thinking skills and should thus be spared the frustration generated by such tasks. Because this view may cause teachers to treat students in a nonegalitarian way, it is important to find out whether or not it is supported by empirical evidence. The goal of this study is to examine this issue in light of four different studies, by asking the following question: Do low-achieving students gain from teaching and learning processes that are designed to foster higher order thinking skills? Each of the4 studies addressed a different project whose goal was to teach higher order thinking in science classrooms. Following a brief general description of each project, we provide an analysis of its effects on students with low and high achievements. The findings show that by the end of each of the 4 programs, students with high academic achievements gained higher thinking scores than their peers with low academic achievements. However, students of both subgroups made considerable progress with respect to their initial score. In one of the 4 studies the net gain of low achievers was significantly higher than for high achievers. Our findings strongly suggest that teachers should encourage students of all academic levels to engage in tasks that involve higher order thinking skills.

Correspondence and requests for reprints should be sent to Anat Zohar, School of Education, Hebrew University, Jerusalem, Israel 91905. E-mail: Msazohar@mscc.huji.ac.il

Fostering higher order thinking among students of all ages is considered an important educational goal. As explained in what follows, however, teachers often believe that this important goal is not intended for all students. A common belief among teachers is that tasks that require higher order thinking are appropriate only for high-achieving students, whereas low-achieving students, who can barely master the basic facts, are unable to deal with such tasks (Zohar, Degani, & Vaaknin, 2001).

Writing this article was motivated by incidents we encountered repeatedly as part of our fieldwork in teachers' professional development workshops. These workshops were designed to prepare teachers for instruction of higher order thinking skills in the context of science modules, prepared as part of a large-scale educational reform. Teachers' attitudes toward instruction of higher order thinking skills were in general favorable, expressing the view that it is an important and valuable educational goal. Nevertheless, many teachers often qualified this attitude by expressing views such as the following:

- Some kids simply can't do it. … You cannot ignore the variability among children.
- I also think that it's [i.e., higher order thinking] inappropriate for weak students. I would very much like it to be for the weak students, but I have a feeling it will work well only with the strong ones. … You can trust them, they are interested and curious. The weaker ones, we have to give them a lot of support and carry them on our shoulder to get some results.

These excerpts indeed express the belief that instruction of higher order thinking is an appropriate goal mainly for high-achieving students and that low-achieving students, who have trouble with mastering even basic facts, are unable to deal with tasks that require thinking skills. This belief may have serious educational implications because it undermines the goal of helping lower achieving students in closing gaps, thereby denying them equal educational opportunities. It is therefore important to examine empirical evidence regarding this issue to find out whether this belief is founded. The purpose of this article is to shed light on this issue by presenting four studies and discussing their implications for teaching thinking skills to low-achieving students.

THEORETICAL UNDERPINNING

General Background

The educational goal of fostering students' thinking has been the focus of numerous books and research articles (e.g., Adey, 1999; Adey & Shayer, 1994; Brown & Campione, 1990; Bruer, 1993; Burden & Williams, 1998; Carmichael, 1981; Chance, 1986; De Bono, 1985; Feurstein, Rand, & Rynders, 1988; Greeno &

Goldman, 1998; Halpern, 1992; Lipman, 1985; Nickerson, Perkins, & Smith, 1985; Perkins, 1992; Perkins & Grotzer, 1997; Resnick, 1987; Resnick & Klopfer, 1989; Schoenfeld, 1989, 1992; Tishman, Perkins, & Jay, 1995). Each of the programs described in these sources has its own definition of thinking and of skills. In fact, the different definitions of thinking and the number of available options can be confusing (Marzano et al., 1988). Referring to this confusion, Resnick (1987) wrote that thinking skills resist precise forms of definition; yet, higher order thinking skills can be recognized when they occur. Some of the characteristics of higher order thinking, according to Resnick, are the following: it is nonalgorithmic, it tends to be complex, it often yields multiple solutions, and it involves the application of multiple criteria, uncertainty, and self-regulation. The term *higher order thinking skills* may also be used to delineate cognitive activities that are beyond the stage of understanding and lower level application according to Bloom's taxonomy (Bloom, 1956). We object to the hierarchies of educational goals implied by Bloom's work, but we find that it specifies cognitive levels that are clear, succinct, and still useful. Based on Bloom's taxonomy, memorization and recall of information are classified as lower order thinking whereas analyzing, synthesizing, and evaluating are classified as higher order. Additional examples of cognitive activities that are classified as higher order include constructing arguments, asking research questions, making comparisons, solving nonalgorithmic complex problems, dealing with controversies, and identifying hidden assumptions. Most of the classical scientific inquiry skills, such as formulating hypotheses, planning experiments, or drawing conclusions are also classified as higher order thinking skills. It is justified to group such varied cognitive activities into the same category of "higher order thinking" activities because despite the fact that they are so different from each other, they all follow the characteristics of higher order thinking according to Resnick. In addition, all of them would also be classified into stages that are beyond recall of information and comprehension according to Bloom's taxonomy.

Teaching and Learning for Understanding: Higher Order Thinking and Low-Achieving Students

In the early part of the 20th century, education focused on the acquisition of basic literacy skills: reading, writing, and calculating. Most schools did not teach to think and read critically or to solve complex problems. Textbooks were loaded with facts that students were expected to memorize and most tests assessed students' ability to remember these facts. The main role of teachers was perceived as that of transmitting information to students (Bransford, Brown, & Cocking, 2000). Traditional learning theories were based on Behaviorism, which advocated learning as linear and sequential. Learning objectives were sequenced to progress from simple, lower order cognitive tasks to more complex ones. Complex understanding was thought to occur only by the accumulation of basic, prerequisite learning (e.g., Bloom, 1956; Gagne, 1974). It was commonly believed that only after students have mastered a new subject at the level of information recall may they progress to engaging in that subject at higher cognitive levels. These proposed hierarchies of learning forms implied that problem solving and other activities recognized as thinking occupy the top of these hierarchies. Although such theories helped keep alive the idea that there was more to education than acquiring bodies of facts, they isolated thinking and problem solving from the main "basic" or "fundamental" activities of learning. Thinking and reasoning became not the heart of education, but hoped-for summits that most students never reached (Resnick & Klopfer, 1989). Consequently, low-achieving students were often chronically engaged in lower order cognitive assignments because they never mastered the simplest level of knowledge. In contrast, higher achieving students, having mastered the basic skills, were viewed as prepared to handle more complex learning tasks (Shepard, 1991)

Contrary to this view, more recent educational approaches consider aspects of "high" literacy as essential for tackling the complexities of contemporary life. As information and knowledge are growing at a far more rapid rate than ever before in the history of humankind, the meaning of "knowing" has shifted from being able to remember and repeat information to being able to find and use it effectively. Developments in cognitive science do not deny that facts are important for thinking and problem solving, but show clearly that "usable knowledge" is not the same as a mere list of disconnected facts. Being able to use knowledge to solve new types of problems means that one must *understand* that knowledge. Thus, new teaching and learning practices emphasize learning with understanding (Bransford et al., 2000). Such learning is tightly related to thinking and reasoning. This idea was formulated by Perkins and Unger (1999) in the following way: "Understanding a topic is a matter of being able to think and act creatively and competently with what one knows about the topic. … The ability to perform in a flexible, thought-demanding way is a constant requirement" (p. 97).

An important implication of this view is that the mental processes we have customarily associated with thinking are not restricted to some advanced stage of learning. Instead, thinking skills are intimately involved in successful learning of even elementary levels of reading, mathematics, and all other school subjects. If acquiring knowledge is defined as learning with understanding, learning simply cannot take place without thinking. Understanding is seen as being constructed while learners engage in thinking and inquiry in contexts that make sense to them. Learning inherently involves components of inference, judgment, and active mental construction. Thus, the traditional view that the basics can be taught as routine skills, with thinking and reasoning to follow later as an optional activity that may or may not take place, can no longer guide the educational practice. Instead, thinking must be applied to all learning and to all learners (Bransford et al., 2000; Bruer, 1993; Perkins, 1992; Resnick & Klopfer, 1989; Resnick & Resnick, 1992; Perkins

& Unger, 1999). This view, namely that teaching for higher order thinking is important for the learning of all students in all academic tracks, is emphasized by several additional researchers (Levine, 1993; Newmann, 1990; Peterson, 1988; Pogrow, 1988, 1996; White & Frederiksen, 1998). Resnick (1987) referred to this idea in an eloquent way by saying that fostering students' thinking is one of the most ancient goals of education, dating back to the days of Plato in ancient Greece. During many generations, this goal was intended only for a small, restricted group of elite students; the vast majority of students did not have the privilege of enjoying an educational tradition that fostered their thinking. Therefore, said Resnick, there is nothing new in including the teaching of higher order thinking and problem solving in the curriculum of *some* students. Including this goal in the curriculum of *all* students is, however, an educational innovation. A similar idea is also expressed in the Science Technology Society (STS) approach.

STS Approach: Higher Order Thinking and Low-Achieving Students

Curriculum reform in science education over the past five decades has been characterized by different interpretations of the role of science in the curriculum. During the 1950s and 1960s, disciplinary knowledge was at the focus of the science curriculum. As of the 1970s, emphasis was placed on knowledge relevant to fulfilling personal and societal needs (Bybee, 1993; National Research Council, 1996; Wallace & Louden, 1998). These reform movements included the interrelated and partially overlapping movements of STS (Bybee & Ben-Zvi, 1998), environmental education (Zoller, 1986/1987), and Science for All (Fensham, 1985). Based on the idea of incorporating societal, cultural, environmental, political, and ethical aspects into the science curriculum, the STS approach aims to endow students with scientific literacy (Aikenhead & Ryan, 1992; Bybee, 1987; Dori & Herscovitz, 1999; Yager & Hofstein, 1986; Yager & Penick, 1988; Yager & Tamir, 1993).

STS education does not only encourage scientific literacy in a societal perspective (Driver & Leach, 1993; Pedretti & Hodson, 1995), but it also aims at improving students' higher order thinking skills (Dori & Tal, 2000; Tal, Dori, & Lazarowitz, 2000). To be literate consumers of scientific knowledge, students need to know how to read popular scientific articles written by lay people in a critical manner and how to solve complex problems that involve science, technology, and society in an effective way. They also need to know how to apply value judgments to technological innovations, to question the quality of available information, and to understand that some problems may have more than one possible solution or may not even have a solution at all (Dori & Herscovitz, 1999; Tal, Dori, Keiny, & Zoller, 2001). These skills, all of which are necessary for scientific literacy according to the STS approach, are all components of higher order thinking.

STS curricula are expected to make science meaningful to *all* students. One of the goals of applying such curricula is to increase the number of students who would gain scientific literacy, as compared to the relatively few students who find science meaningful when taught by traditional curricula. The rationale for this approach is that it is important that all citizens, not just an elite of scientists, be science literate. Thus, the entire student population should be challenged to develop their higher order thinking skills, not only high-achieving students. This idea is reflected very clearly in the name of the movement *Science for All* (Fensham, 1985). STS principles of Science for All advocate teaching science to students at all thinking levels, not just to a select elite. Thus, the purpose of STS education is to teach higher order thinking and problem-solving skills to all students, high-achieving as well as low-achieving ones. Unfortunately, current educational practices often neglect this goal.

The Discrepancy Between Theory and Practice

Practitioners often disregard many of the theoretical recommendations and suggestions described earlier. Specifically, they often overlook the recommendations put forward by many theoreticians to teach higher order thinking to all students. Raudenbush, Rowan, and Cheong (1993) reported a number of studies showing that teachers in classes of high-achieving students are substantially more likely to emphasize higher order thinking processes than teachers in classes of low-achieving students. Raudenbush et al. suggested the hypothesis that the higher the academic track of a class, the more likely a teacher is to report an emphasis on teaching for higher order thinking in that class. If this hypothesis is correct, one can assume that the same teacher will teach differently in higher and lower academic tracks, leading to considerable "within-teachers" variability. Raudenbush et al. asked teachers in 16 schools to identify their instructional goals for each of their classes and then constructed an instrument to capture higher order thinking in Mathematics, Science, Social Studies, and English. The results showed that the same teacher tended to emphasize higher order thinking when teaching students of higher academic achievements more than when teaching students of lower academic achievements. Raudenbush et al. also cited additional studies, showing that teaching for higher order thinking in high school occurs far more often in accelerated tracks than in low-track classes (Metz, 1978; Oakes, 1990; Page, 1990).

Another study, pointing in a similar direction, addressed teachers' beliefs in this field. Teachers' beliefs regarding low-achieving students and instruction of higher order thinking were addressed by a study using clinical interviews (Zohar et al., 2001). According to this research, only 20% of the interviewed teachers believed that the goal of teaching higher order thinking is equally appropriate for low- and high-achieving students, whereas 45% believed that it is appropriate only for high-achieving students. The most common explanation teachers gave for this distinction between low- and high-achieving students is that thinking-based learning creates difficulties and confusion for weak students, alienating them from the lesson. Indeed, teachers' beliefs that it is inappropriate for low-achieving students to engage in higher order thinking seems to be a major factor in dissuading them from using this method. Some teachers indicated that thinking-based learning might induce frustration in weak students, leading to affective difficulties. The findings also suggest that teachers' beliefs in this context are related to their general theory of instruction. Viewing learning as hierarchical in terms of students' academic level was found to be related to a traditional view of learning, seeing learning as progressing from simple, lower order cognitive skills to more complex ones.

These beliefs may have far-reaching consequences, as they may lead teachers to deprive low-achieving students from tasks requiring higher order thinking, which are crucial for their development. Thus, teachers' beliefs might become a self-fulfilling prophecy, as they are likely to influence them to expose high-achieving students to tasks requiring higher order thinking skills more often than they would expose low-achieving students to such tasks. Consequently, the gap between low- and high-achieving students would become wider.

EDUCATIONAL CONTEXT: AN OVERVIEW OF SCIENCE EDUCATION REFORM IN ISRAEL

A significant decline in the number of high school students electing science courses in Israel, along with insufficient scientific literacy among nonscience majors and relatively low scores in international assessments, motivated the Israeli Ministry of Education to call for a reform in science teaching. The Harari National Committee resulted in an elaborate report—*Tomorrow '98*—that included 43 recommendations for new programs, special projects, changes, and improvements in the areas of curriculum development and implementation, pedagogy of science, and professional development of science teachers (Harari, 1992). Following the report, the Ministry of Education financed several reform projects whose goal was to implement the Harari committee recommendations. The projects reported in this article were all part of these reform initiatives. The Tomorrow '98 report considered the need to make science an integral part of the education for all citizens, suggesting two specific recommendations. One concerned high school students who do not opt to specialize in any of the science disciplines (biology, chemistry, or physics). A new curriculum was proposed for these students—*Science and Technology for All.* A second recommendation concerned middle school students for whom another new curriculum was proposed—*Science and Technology for Middle School*. These programs consisted of interdisciplinary modules integrating Science, Technology, Environment, and personal topics with the framework of teaching *Science for All.*

Hofstein, Aikenhead, and Riquarts (1988) identified several problems concerning the implementation of such STS-type programs, including the following:

- The interdisciplinary nature of the content and unfamiliarity of the teachers with a subject matter in which they were not originally trained.
- Unfamiliarity of teachers with required teaching strategies.
- Inappropriate professional development techniques and procedures for pre- and inservice training.

Several of the Tomorrow '98 projects attempted to overcome these obstacles by involving teachers in the process of curriculum development, as well as in the development of instructional techniques and relevant assessment methods. It was hypothesized that by involving teachers in the process of "bottom-up" as opposed to "top-down" curricular procedures, one would reduce the level of anxiety that often exists among teachers who are expected to teach unfamiliar subject matter (Dori & Hofstein, 2000; Dori, Tal, & Tsaushu, 2003). Sabar and Shafriri (1982) claimed that teachers' participation in curriculum development gives the teacher greater autonomy and internalization. Development of learning and curricular materials by teachers has been recognized in the last two decades as an important and effective method for understanding the curriculum potential (Ben-Peretz, 1985) and for professional development of teachers (Tal et al., 2001).

In this article we present two modules in which teachers were actively involved in development and implementation. These modules are *Quality of the Air Around Us* (Dori & Herscovitz, 1999) and *Biotechnology, Environment, and Related Issues* (Dori et al., 2003). These STS-oriented modules, which followed the recommendations of the Harari committee, cater to the requirement of fostering higher order thinking skills. While developing and applying these modules, special care was taken to involve not only high-achieving students but also low-achieving ones to ensure that *all* students would develop higher order thinking skills to the best of their abilities.

Another recommendation of the Harari (1992) committee addressed in this article concerns the need to foster students' higher order thinking and problem-solving skills:

> In many places in the world today there are programs designed to improve the individual's creative thinking, inventive thinking, logical thinking, etc. … This issue is worthy of exploration. The intention is to investigate the feasibility of including such programs in our schools. (p. 47)

One of the practical consequences of this recommendation was the funding of the Thinking in Science Classrooms (TSC) project (Zohar, 1996; Zohar & Nemet, 2000, 2002; Zohar, Weinberger, & Tamir, 1994). As two of the units described in this article (the *Genetic Argumentation* unit and the *Critical and Scientific Thinking* unit) are part of the TSC project, some background information about the project is called for.

Of the many approaches to teaching higher order thinking, the TSC project undertakes the infusion approach, arguing that thinking takes place within various curricular areas. Although learning is embedded in rich conceptual frameworks, the thinking principles (also referred to as *thinking skills* or *strategies*) are made explicit and become a focused goal of instruction and thus a common target of classroom discourse (Burden & Williams, 1998; Ennis, 1989).

The project's outcome is a set of learning activities that were specifically designed to foster inquiry, higher order thinking, and scientific argumentation in multiple science topics. The learning activities match topics from the junior high school science curriculum. Four books of learning activities were published and inservice professional development courses took place all over the country.

It is particularly important to clarify the meaning of the term *skill* in the context of the TSC project. In the higher order thinking literature, the term *thinking skills* often refers to general entities that are disconnected to the rich conceptual frameworks of academic subjects. However, the emphasis on thinking skills in the TSC project integrates skill learning into studies of particular topics in science. In TSC lesson instruction revolves around tasks and problems that students are asked to solve. For example, students may be asked to argue about bioethical dilemmas in human genetics, to criticize an article about the diminishing ozone layer, or to engage in open inquiry about vitamins. The cognitive demands for solving these tasks consist of multiple thinking skills. After engaging these thinking skills on a procedural level (i.e., completing the tasks and solving the problems), students engage in a metacognitive activity regarding these skills. Through guided discussions and activity sheets, students reflect on the thinking skills they have been using; make generalizations and rules regarding these skills; and verbalize how, when, and why each specific skill is being used. Teachers are also advised to engage in transfer activities, directing students to additional circumstances (both in other school subjects and in everyday life) where the same thinking pattern (or skill) may be employed.

Thus, thinking skills are embedded in rich science contents and are also addressed as explicit educational goals. One of the assumptions the project is based on is that teaching of higher order thinking must be systematic. Practicing a skill once or twice a year through problem solving may offer students an exceptionally interesting lesson, but will not be very useful in fostering their thinking. The methodology used in the TSC project is to repeat the same skill time and again in different scientific contexts and to apply it to various types of problems. Accordingly, several different types of learning activities were developed: learning activities that follow lab experiments, Invitations to Inquiry (Schwab, 1963), critical assessment of newspaper clips, investigation of microworlds, fostering argumentation skills, and open-ended inquiry learning activities. Evaluation studies have shown that students who studied with the TSC learning activities gained significantly higher scores on reasoning tasks and on science knowledge tests than students from comparison groups who studied in the traditional way (Zohar, 1996, 1999; Zohar & Nemet, 2000, 2002; Zohar et al., 1994).

THE RESEARCH QUESTION AND ITS EDUCATIONAL SIGNIFICANCE

Weighing the theoretical and practical implications discussed in the previous section, two contradictory views become apparent. The first, embraced by both the "learning for understanding" conceptions of learning and by the STS approach, asserts that all students should be the target of teaching higher order thinking. The second view, embraced by many practitioners, holds that low-achieving students are, by and large, not really capable of higher order thinking.

To be considerate of low-achieving students' limitations and to avoid frustrating them, many teachers maintain that these students should be spared the challenges involved in tasks requiring higher order thinking. Such a belief, however, may be questioned in the face of our empirical evidence, which addresses the following research question:

> Do low-achieving students gain from teaching and learning processes that are designed to foster higher order cognitive skills, and if so, to what extent?

It should be noted that seeking an answer to this question was not the primary goal of the studies described in this article. Rather, these studies were conducted to assess the effects of four different research projects that were designed and implemented as part of the Harari science education reform in Israel. All four teaching units or modules described in this sequel share a common goal: to develop students' higher order thinking skills as an essential component of science learning.

Each of the two authors was the director of two of these projects. After the assessment of our projects had been finalized, we shared our findings. We had worked on the different projects independently of each other. The students who participated in these projects were of different ages. Some of them studied in homogeneous classes, whereas others studied in heterogeneous ones. Nevertheless, we were struck by the observation that despite the considerable diversity in students' backgrounds and learning environments, our findings regarding the effect of these projects on low-achieving students were similar. In the course of our fieldwork, both of us also encountered numerous teachers who questioned the value of using these modules with low-achieving students (as demonstrated by the excerpts in the opening section of this article). The realization that our findings might be significant in addressing the gap between theory and practice described earlier led us to collaborate in communicating these findings through a joint article.

The somewhat unusual chronicle of events that led to this article is reflected in its special layout. In what follows, each of the four studies is described in a separate section. For each study, we first provide general information, including the main research questions or objectives, a brief description of the research setting, the participants who were evaluated, the primary means of assessment, and the main findings. We then elaborate on the specific findings regarding the effects of that particular study on low- and high-achieving students. We wrap up with general conclusions regarding our findings and with recommendations for further research. A brief summary that serves as an advance organizer is presented in Table 1.

STUDY 1: FOSTERING QUESTION-POSING CAPABILITIES THROUGH A CASE-BASED TEACHING/LEARNING METHOD IN THE AIR QUALITY MODULE

Research Setting

Tenth grade students in Israel are required to take at least one science course. Following the Harari Committee recommendations (Harari, 1992), nonscience majors often choose a course titled “Science and Technology for All.” The module assessed in this study was developed as part of a Science, Technology, and Environment in Modern Society (STEMS) project, which was part of the effort to develop the Science for All[1] curriculum. The module titled *The Quality of Air Around Us* was developed by a group of science teachers mentored by an academic advisor (Dori & Herscovitz, 1999).

The goal was to expose students to controversial issues, to develop their ability to pose questions, and to teach them how to read scientific articles in a critical manner. Question posing is a fundamental cognitive component that guides human reasoning. Particular classes of questions invite mental construction of causal chains, justifications, and goal–plan–action hierarchies (Graesser, Baggett, & Williams, 1996).

Case studies have been effectively used in medical, business, and law schools (Dori, 1994; Herried, 1994). The Air Quality module consisted of five case studies taken from sources such as daily newspaper articles and popular science magazines that were applied using the Jigsaw cooperative learning method. The module was divided into five topics dealing with nitrogen, carbon and sulfur oxides, green

[1]In Israel, the term *Science and Technology for All* is used as an extension of the term *Science for All* in the science education literature.

TABLE 1
Overview of the Four Studies

Module Title and Reference	*Module Subject Matter*	*Module Duration*	*Participants*	*Main Research Objective*	*Higher Order Thinking Skill*	*Main Research Findings*
The Quality of Air Around Us (Dori & Herscovitz, 1999)	Chemistry, environment, and social aspects	30–40 hr	Seven 10th-grade homogeneous[a] classes from five schools: urban, rural, and agricultural (N = 127)	To investigate whether students' question-posing capabilities can serve as an alternative assessment method	Question posing capability	High and low academic level students improved the number and complexity of questions they posed
The Genetic Revolution: Discussion of Moral Dilemmas (Zohar & Nemet, 2000, 2002)	Human genetics	10–12 hr	Five experimental 9th-grade classes and four comparison classes from two middle- class heterogeneous[b] schools (N = 186)	To assess students' progress in argumentation skills and genetic knowledge; to assess transfer of argumentation skills from the context of genetic to everyday life	Argumentation skills	High and low academic level students advanced their reasoning skills

Biotechnology, Environment, and related Issues (Dori, Tal, & Tsaushu, 2003)	Biotechnology with moral and environmental dilemmas	30–40 hr	Eight 10th- to 12th-grade classes from six different high schools: Arab, Jewish secular and religious, urban and small community (*N* = 201)	To examine the effect of the Biotechnology module on students' knowledge and higher order thinking skills	Question posing, argumentation, and system thinking skills	High and low academic level students significantly improved their scores in the higher order thinking skills category
Fostering Critical and Scientific Thinking (Zohar & Tamir, 1993; Zohar, Weinberger, & Tamir, 1994)	Water balance in living organisms	24 hr	Ten experimental 7th-grade classes and 11 comparison classes from four schools that were heterogeneous in terms of socioeconomic background (*N* = 464)	To assess the effect of the Thinking in Science Classrooms unit on students' reasoning skills and biological knowledge	Testing hypotheses, identifying relevant information, recognizing logical fallacies, and differentiating between experimental results and conclusions	Students from all academic levels improved their reasoning skills.

[a]Homogeneous class means that students in that class have similar interests in science and come from a similar socioeconomic background. [b]Heterogeneous class means that students in that class have dissimilar interests in science and come from different socioeconomic background.

house effect, ozone layer depletion, and industrial odors as warning signs. Students were exposed to environmental problems created by a nearby power plant and their possible technological and legislative solutions. Students' assignments included case studies demonstrating social and environmental aspects of science and their relevance to daily life. After reading the case studies students were requested to analyze data, solve complex problems, pose questions, conduct critical group discussions, play different roles, and write creative titles and passages with regard to controversial issues. While they were exposed to new learning situations through case studies, students interacted with each other, thereby constructing new knowledge and posing questions at various complexity levels.

Research objectives were as follows:

- To examine ways of using students' question-posing capabilities as an alternative assessment method.
- To investigate the effect of the case study teaching and learning approach on question-posing capabilities of high school students at different academic level.

The research population included seven 10th-grade classes from five different types of schools in the northern part of Israel. All the teachers who participated in the STEMS project and consented to teach the *Quality of Air Around Us* module taught classes that became part of the research population. Hence there was no preselection of the research population, except for the teachers' willingness to teach the topic.

Based on a classification made by the management of each school, the student population was divided into three academic levels: high (H), science majors; intermediate (I), average students; and low (L), students with some learning difficulties. Science majors (H level students) were required to take one or more of three courses—physics, chemistry, or biology. These students took the module for extra credit, whereas intermediate and low-level students took it as the Science and Technology for All required course. Although all classes were to a certain extent heterogeneous, their average scientific and academic levels reflected their classification into the three academic levels. The school's classification of students into the three levels was verified through a part of the pretest addressing scientific literacy.

Assessment Method

To assess the effect of the case study method on students' question-posing capability, the results of pre- and posttest case studies were analyzed. Case studies were part of both the pre- and the posttests. The pretest provided data for both the instruction and research. For instruction, it served as a baseline for the teachers, who used it to classify students by academic levels and to assign them into the various Jigsaw groups. The posttest was used to assess students' performance and to grade them. Comparing the results of the pretest case studies with those of the posttest was used for measuring students' improvement in question-posing capability as a result of the learning process.

To illustrate our method of analyzing question-posing capability, consider the following set of four questions asked by student A:

1. What is ozone?
2. Write a letter to a manager in the petrochemical industry plant and express your opinion about gases emitted that cause the photochemical smog.
3. In your opinion, are we currently in danger?
4. Due to the fact that certain gases cause the hole in the ozone layer, can we use them to eliminate the "bad" ozone?

We counted the number of questions posed by each student (in our example it is four) and compared the percentage of questions asked before and after the treatment (see Table 2).

Next, we categorized each question by its orientation. The three question orientation attributes are phenomenon or problem description, hazards related to the problem, and treatment or solution. The researchers' views are that proposing solutions point to a higher level of understanding the problem than describing it, and that finding treatments or solutions is more positive and productive than just identifying hazards. Focus was placed not on recognizing the problem or on identifying the hazards, but on attempts to find solutions. The orientation of the first three questions (which student A posed in the aforementioned example) is "problem description," whereas that of the fourth (last) question is "possible solutions." The trends of change in the questions' orientation are presented in Figure 1.

A more thorough analysis was based on the complexity of each question, which was used in the computation of the aggregate score (see Table 3). To determine the questions' complexity systematically and objectively, we developed and applied a quantitative method for calculating the complexity of an individual question and of a set of questions. The complexity level of a set of questions asked by an individual student is the student's aggregate score. The coding scheme of the complexity was influenced by thinking skills classification (Shepardson, 1993; Shepardson & Pizzini, 1991) and criteria for question asking (Graesser & Person, 1994) as well as problem solving (Zoller, 1987).

In a nutshell, the method first calls for determining whether or not answering the question requires only knowledge that is presented in the case study. Questions whose answers required knowledge only received a complexity score of zero. Questions requiring application, analysis, value judgment, or expression of an opinion regarding controversial issues were assigned a higher score. It should be noted that the number of questions accounted for by the first component (the number of questions posed by each student) is different than the number of questions

TABLE 2
Mean Scores, Standard Deviations, and Maximum Number of Questions Students Posed in the Pre- and Posttests by Academic Levels in Study 1

	Pretest				*Posttest*				
Academic Level	*N*	*M*	*SD*	*Maximum Number of Questions*	*N*		*SD*	*Maximum Number of Questions*	p^a
High	59	2.53	.99	6	56	6.38	2.13	12	.001
Low	39	2.05	1.12	4	29	4.38	1.39	9	.001

[a]As computed by Kruskal-Wallis Test (χ^2 approximation) for mean number of questions per student among the levels.

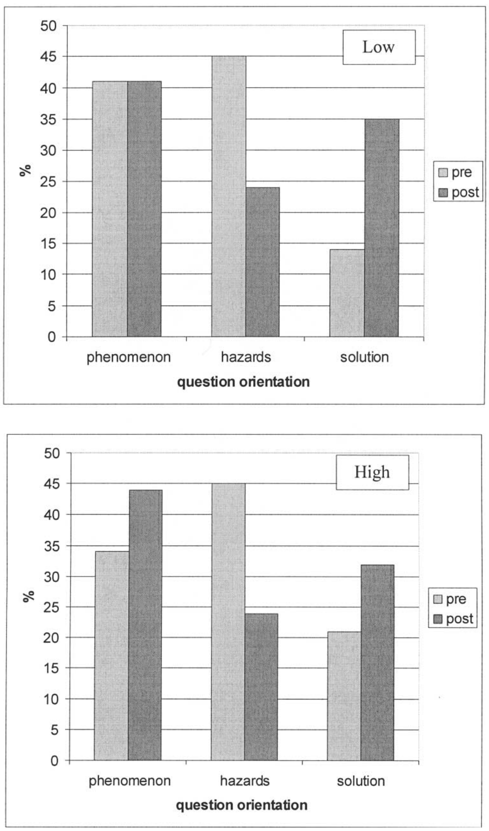

FIGURE 1 Trends in question orientation from the pretest to the posttest for low-achieving students (top) and high-achieving students (bottom) in Study 1.

TABLE 3
Mean Scores, Standard Deviations, and Significance of Question Complexity in the Pre and Postcase Study Questionnaires by Levels in Study 1

	Pretest				Posttest				
Academic Level	*N*	*M*	*SD*	*Maximum Complexity Aggregate Score*	*N*	*M*	*SD*	*Maximum Complexity Aggregate Score*	p^a
High	59	3.71	1.90	8	56	10.23	3.80	20	.001
Low	39	2.85	1.61	8	29	6.31	4.18	18	.001

[a]As computed by a Kruskal–Wallis Test (χ^2 approximation) for mean complexity aggregate score per student among the levels.

accounted for the student's aggregate score. Student A asked four questions, of which the first question required only knowledge (because the answer was provided in the case study). That student's aggregate score is the sum of complexity scores of only the three questions (2, 3, and 4) because only these questions require higher order thinking skills. The categorization and the question's scores (as computed by the formula in Dori & Herscovitz, 1999) were as follows:

2. Write a letter to a manager in the petrochemical industry plant and express your opinion about gases emitted that cause the photochemical smog. Complexity category = "Expressing opinion." Score = 1 point.
3. In your opinion, are we currently in danger? Complexity category = "Judgment and/or evaluation" and "Expressing opinion." Score = 2 points.
4. Due to the fact that certain gases cause the hole in the ozone layer, can we use them to eliminate the "bad" ozone? Complexity category = "Application and analysis," "Judgment and/or evaluation," and "Interdisciplinary approach." Score = 3 points.

The student's aggregate question complexity score was obtained by summing over the complexity scores of the questions that the student asked. In this case only the aforementioned three questions were accounted for by computing the aggregate complexity score, which was therefore 6.

Findings

The results indicated that overall, students increased their scores in the posttest compared to the pretest. Students' performance improved significantly between the pretest and posttest with respect to all the three components that were analyzed (i.e., number of questions posed, question orientation, and question complexity).

The total number of questions posed by students increased from 298 in the pretest to 639 in the posttest ($p < .0001$).

Regarding question orientation, we found that in the pretest half of the students were primarily concerned with hazards related to the problem presented in the case study. Only about one fifth of the questions students posed related to a possible solution or to formulating an argument. Examining trend changes in question orientation, we found that the percentages of solution- and argument-oriented questions increased from 19% in the pretest to 33% in the posttest. Fewer questions in the posttest (24%) than in the pretest (45%) dealt with hazards related to the problem. This indicated an increase in students' awareness of the need for and feasibility of seeking practical solutions to a given problem rather than being fixated on inquiring about risks.

Regarding question complexity we found that the mean question complexity increased from 3.88 in the pretest to 8.87 in the posttest ($p < .0001$). Through the study of the Air Quality unit, students gained a more complex view of the real world problems that were addressed in this module. Taken together, these findings show considerable gains in students' question-posing capabilities following instruction.

Low- Versus High-Achieving Students in the Air Quality Module

As explained earlier, students were engaged in studying the five topics of the Air Quality module using the Jigsaw method. Comparing students' achievements in the five topics that comprised the Air Quality module, we found out that high academic level students maintained the same level of knowledge and understanding in their expert topic as in the other topics (which they had learned from their peers). However, the knowledge level of intermediate and low academic level students declined in the topics they had studied from peers when compared to the expert topic (which they had learned on their own and taught their peers). More details appear in Dori and Herscovitz (1999).

In comparing academic levels, we found a significant difference in the extent of increase in the average number of questions among the three levels. Because this difference was entirely due to the significant difference between levels H and L, from now on we shall focus on these two levels, putting the results of the intermediate level aside. The results presented in Table 2 show the increase in students' question-posing capability in the Air Quality module. The increased capability was significant for both academic levels as reflected in the mean number of questions posed by each student.

Figure 1 shows the distribution of question orientation in the pre- and posttests. We found that the distribution was similar for H and L levels. The low-achieving students increased the solution orientation of their questions from 14% to 35%. This is a higher increase than that of the high-achieving students.

Studying the complexity of the questions posed by students, we found that both high- and low-achieving students improved significantly in the posttest as compared with the pretest (see Table 3). Taken together, these findings show that students from both H and L levels improved their question-posing capabilities following their study of the Air Quality Module.

Our findings are in line with those of Graesser and Person (1994), who found that students' achievements were positively correlated with the quality of the questions students posed.

STUDY 2: FOSTERING STUDENTS' ARGUMENTATION SKILLS THROUGH BIOETHICAL DILEMMAS IN GENETICS

Research Setting

The TSC project was funded in Israel as part of the science reform that followed the recommendations of the Harari Committee (Harari, 1992). The unit *The Genetic Revolution—Discussions of Moral Dilemmas* (or Genetic Revolution for short) is part of the TSC project. Learning activities in the TSC project were designed to foster higher order thinking skills. In this unit, scientific argumentation skills were integrated into the regular junior high school science curriculum. This particular unit was designed according to two sets of goals. One set of goals consists of a list of several topics in human genetics (e.g., genetic counseling, information about genetic traits, gene therapy and genetic cloning). The other set of goals consists of fostering argumentation skills (e.g., formulating an argument and justifying it or formulating a counter-argument and justifying it). This 12-hr unit, designed for ninth grade, includes 10 moral dilemmas about issues involving modern technologies in genetics.

Biological knowledge is addressed in two ways: first, each dilemma begins with a short written introduction presenting information about concepts in genetics. Second, students must make use of their biological knowledge when they are thinking about the problems presented in the dilemmas. The value of grounding decisions upon reliable knowledge is explicitly emphasized time and again throughout the unit. Argumentation skills are also addressed in two ways: first, they are addressed in a lesson that is entirely devoted to explicit instruction about argumentation. Arguments are defined and their structure is explained. Criteria distinguishing between good and bad arguments are discussed. Second, argumentation skills are addressed in each of the dilemmas when, in the specific context of each dilemma, students are asked to apply them (Zohar & Nemet, 2000, 2002).

The general research objective of this study was to investigate the learning that took place following the implementation of the Genetic Revolution unit and its effects on both biological knowledge and argumentation skills. More specifically, our goal was to answer the following questions:

1. How do students initially (i.e., before instruction) apply specific biological knowledge to argument construction and what is their initial ability to formulate arguments?
2. How does instruction of the Genetic Revolution unit affect students' biological knowledge and argumentation skills as compared to traditional instruction that covers the same biological content?
3. Can students who have acquired argumentation skills in the context of the Genetic Revolution unit transfer these skills to a new context (moral dilemmas taken from everyday life)?

Participants in this study were ninth grade students in two middle-class heterogeneous schools in Israel. The research design included an experimental group that received treatment and a comparison group that was taught the same topics in human genetics for the same amount of time using traditional instruction (Zohar & Nemet, 2000, 2002). The experimental group consisted of five classes and the comparison group consisted of four classes.

Assessment Method

Students' reasoning abilities were assessed before, during, and after instruction by several means, including an analysis of audio-tapes from group discussions and a series of written tests. Several of the written tests are relevant for the purpose of this study:

- Argumentation tests in genetics that revolved around two dilemmas related to genetics. The Cystic Fibrosis dilemma was used as a pretest and the Huntington dilemma was used as a posttest.
- Argumentation transfer tests that consisted of two moral dilemmas taken from everyday life (e.g., "Should students report a classmate who cheated in a test?").

The purpose of the latter tasks was to assess transfer of argumentation skills from the context of genetics to the context of everyday life. One of these tasks was assigned as a pretest and the other as a posttest. It should be noted that although the topics of these tests differed from each other, they were all identical in terms of their logical structure in the sense that students were required to construct the same type of arguments in response to the questions presented in each of the tasks.

In students' responses to the written pre- and posttests, we analyzed students' ability to formulate arguments, alternative arguments, and rebuttals and to justify them. Alternative arguments are arguments that contradict one's original opinion. Rebuttals are arguments that refute the alternative arguments. The criterion for argument formulation was whether or not the written responses included a conclusion with at least one relevant justification. Responses that included a conclusion with no justifications (e.g., "I think they should perform an abortion") or conclusions with pseudo-justifications (e.g., "I think they should perform an abortion because this pregnancy must be terminated") were not accepted as arguments. Justifications were scored according to their number and structure. The score range for the number of justifications was 0 (*no justification*), 1 (*one valid justification*), and 2 (*two or more valid justifications*). The score range for argument structure was 0 (*no valid justification*), 1 (*a simple structure, consisting of a conclusion supported by at least one reason*), and 2 (*a composite structure*). For each argument, counter-argument, or rebuttal, the scores thus ranged between 0 to 4 (because each was scored for both the number of justifications and the argument structure). Because each dilemma consisted of all three components (i.e., arguments, counter-arguments, and rebuttals), the score for each dilemma ranged between 0 to 12.

Findings

The analysis of the written tests revealed that prior to instruction, most students could formulate simple, unsophisticated arguments. Following instruction, an improvement was found in students' argumentation abilities. The genetics argumentation pretests showed that both experimental and control groups had similar scores, indicating an initial similar level of both groups. However, only students in the experimental group improved their scores in the posttest compared to their scores in the pretest. Their gains were found to be statistically significant. Similarly, the transfer tests showed that only the experimental group students were able to transfer the reasoning abilities taught in the context of bioethical dilemmas in genetics to the context of moral dilemmas taken from everyday life.

To assess the effect of the Genetic Revolution unit on students' knowledge in genetics, students were asked to answer a multiple choice test that consisted of 20 items. The results showed that students in the experimental group scored significantly higher than students in the control group in the knowledge test ($M = 72.9$, $SD = 6.0$ and $M = 59.4$, $SD = 4.1$, respectively; $t = 3.94$, $p < .001$). These results indicated that the Genetic Revolution unit is more effective for teaching genetics than the traditional mode of instruction.

Qualitative analysis of two excerpts from group discussions—one from an early discussion and another from a later one—revealed an improvement in the quality of students' argumentation. In the second discussion, students were more careful in expressing claims, in taking more care to make their claims explicit, and in justifying them, as compared to the first discussion. In the first discussion, students tended to talk briefly, but in the second they tended to talk for longer periods of time, suggesting an increase in the complexity of their discourse. Indeed, an additional analysis of transcripts showed that students' discourse in the second discussion was richer in ideas than in the first discussion (Zohar & Nemet, 2000, 2002).

Low- Versus High-Achieving Students and the Genetic Revolution Unit

The results summarized in the previous section show that, in general, students benefited from the Genetic Revolution unit. These results, however, do not indicate the type of students who made progress as a result of instruction. Theoretically, it may well be that only part of the student population (only higher achieving students or only lower achieving students) contributed to the significant gains.

To address this issue, we divided our student population into three subgroups according to their biology grade in the school term that preceded our treatment. Students whose biology grade was between 45 and 70 were grouped into the low-achieving groups; students whose grade was above 70 and below 90 were grouped into the medium-achieving group; and students whose grade was between 90 and 100 were grouped into the high-achieving group. Table 4 presents the results of the analysis for the genetic argumentation tests and Table 5 presents the results of the analysis for the transfer tests (relating to everyday life dilemmas).

The data presented in Tables 4 and 5 show that all three subgroups improved their posttest scores with respect to their pretest scores in both the genetic argumentation and transfer argumentation tests. In all three subgroups, the differences between pre- and posttests were statistically significant with medium to very large Effect Sizes (ES). Thus, these results show that both lower and higher achieving students advanced their reasoning skills following the implementation of the Genetic Revolution—Discussion of Moral Dilemmas unit.

STUDY 3: ENHANCING HIGHER ORDER THINKING SKILLS THROUGH CASE STUDIES IN BIOTECHNOLOGY

Research Setting

This study described and evaluated the *Biotechnology, Environment, and Related Issues* module developed by a group of six teachers from different science disciplines, a coordinator, and an academic advisor. The module addressed various aspects of developments in biotechnology such as new inventions in agriculture, the production of essential materials, and the transformation of genetic characteristics. Students' learning involved scientific and technological aspects through an evaluation of their impacts on and the relationships with society and the environment.

TABLE 4
Scores of Genetic Argumentation Tests in Study 2

		Pretest		*Posttest*				
Academic Level	*N*	*M*	*SD*	*M*	*SD*	*t*	*p*	*Effect Sizes*
All students	71	6.6	2.0	8.7	2.1	7.5	<.001	0.98
Low achievers	18	5.9	2.3	7.5	2.4	4.4	<.001	0.68
Intermediate achievers	26	6.5	2.1	8.5	1.7	3.8	<.001	1.05
High achievers	27	7.3	1.4	9.5	1.9	4.9	<.001	1.32

TABLE 5
Scores of Genetic Transfer Tests
(Relating to Everyday Dilemmas) in Study 2

		Pretest		*Posttest*				
Academic Level	*N*	*M*	*SD*	*M*	*SD*	*t*	*p*	*Effect Sizes*
All students	69	5.2	2.4	8.3	1.9	11.6	<.001	1.55
Low achievers	18	4.6	2.3	7.8	2.3	6.0	<.001	1.42
Intermediate achievers	24	5.6	2.1	8.0	1.6	4.9	<.001	1.30
High achievers	27	5.5	2.5	9.0	1.7	8.0	<.001	1.66

The teaching approach emphasized the development of a variety of thinking skills: posing questions, presenting arguments, and system thinking. The aim in teaching the Biotechnology module was to provide students with the ability to understand various topics concerning STS issues. The unique characteristic of the module is the system approach: case studies were combined with built-in moral dilemmas for both learning and assessment. The core of the module consists of moral questions and controversies concerning the environment, raised by biotechnology research and its applications. Discussing such controversies constitutes a major issue in the module and inspires debates among students and teachers. As mentioned earlier, the case study method was found to be suitable in other STS programs as well (Dori & Herscovitz, 1999; Dori & Tal, 2000; Herried, 1994). Both real stories and fictitious ones were used in the present unit.

The research objective was to examine the effect of the Biotechnology module on students' knowledge and higher order thinking skills.

The research population consisted of nonscience majors in eight classes of grades 10 to 12 from six different high schools. The 201 students represented heterogeneous populations—Arab, Jewish secular and religious schools, and urban and small community schools. The students were classified into three academic levels by the mean scores of a pretest. Students who scored less than 20% in the pretest were classified as low-achieving students. Those who scored between 21%

and 38% were considered intermediate and the ones who got more than 38% were classified as high-achieving students.

Assessment Method

Assessment of the Biotechnology module addressed knowledge and understanding of key scientific issues as well as higher order thinking skills. Such skills were measured in terms of students' ability to identify and analyze environmental, social, and moral dilemmas, as well as their ability to present arguments regarding controversial issues. Assessment tools included pre- and posttests consisting of case studies with built-in dilemmas. Through these tests, we investigated student performance regarding two categories: (a) knowledge and understanding of key scientific issues and (b) higher order thinking skills. The latter category included posing questions, presenting arguments, and system thinking. Each category was analyzed both separately (scoring 100%) and as part of a total score. The total score for each student was computed as a weighted average, with a weight of .3 assigned to the student's knowledge and understanding and a weight of .7 assigned to her or his higher order thinking skills. The benefit of applying these scoring schemes is that they yield a separate score for each category (low and high thinking) whereas the total score provides an overall picture of each individual student, each class, and the entire research population.

Prior to this project, matriculation examinations were developed exclusively for science majors. For the first time in Israel, teachers in Studies 1 and 3 (described in this article) were involved in the development of matriculation examinations geared toward nonscience majors (Dori & Hofstein, 2000; Dori et al., 2003). The examinations for the nonscience majors consisted of tests, projects, critical reading of scientific articles, cooperative assignments, and mini research. Teachers felt that their involvement in developing and matching each pedagogical method with adequate assessment tools benefited both the students and themselves. In this study, we focus on analyzing the results of the pre- and posttests administered as part of the project.

Findings

To investigate the effect of the STS-oriented Biotechnology module on students' learning outcomes, we compared between the pre- and posttests of the entire population. The results revealed a statistically significant improvement in the total scores of the entire student population ($t = 22.8$, $p < .0001$). This was due to improvement in both students' knowledge and understanding and in their higher order thinking skills (Dori et al., 2003).

Low- Versus High-Achieving Students in the Biotechnology Module

Comparing each academic level separately, we found a pattern similar to that of the entire research population.

As Table 6 shows, although both low- and high-achieving students significantly improved their total scores from the pre- to the posttests, the improvement (net gain) of the low achievers (54.1) was significantly higher than that of the high achievers (29.9). We further analyzed the two categories—knowledge and understanding and higher order thinking skills—by the two academic levels. In the knowledge and understanding category, the posttest scores of the low academic level students ($M = 80.5$; $SD = 16.9$) were even *higher* than the scores of their high academic level peers ($M = 74.5$; $SD = 15.2$).

In the higher order thinking skills category (i.e., posing questions, presenting arguments, and system thinking), students of both academic levels improved their scores significantly (see Table 7). The posttest scores of students from the low academic level group ($M = 64.6$; $SD = 31.9$) were *lower* than the scores of their peers from the high academic level group ($M = 73.0$; $SD = 28.1$). However, the *net gain* of higher order thinking skills for low-achieving students (55.7) was significantly higher than that of their high-achieving peers (31.6).

Score analysis by the two categories (knowledge and understanding and higher order thinking skills) revealed that the gap between the two academic groups in the pretest was very wide: 42.1 for the high achievers versus 15.2 for the low achievers. In the posttest, this gap nearly disappeared completely: 73.0 for the high achievers and 69.3 for the low achievers. This study establishes that an almost threefold gap in the pretest score between low and high academic achievers can be narrowed or even eliminated.

A possible explanation for the success of Studies 1 and 3 is teachers' involvement in the developing of curriculum and assessment tools. This has proved to be an effective strategy for elevating teachers' awareness of the pedagogical potential of the STS approach in general and of related assessment modes in particular (Dori & Tal, 2000; Tal, Dori, & Lazarowitz, 2000). These studies show that with the application of appropriate curriculum and instruction, students of all academic levels

TABLE 6
Mean, Standard Deviation, and *t* Tests of Students' Total Scores in the Pre- and Posttests by Academic Levels in Study 3

	Pretest			*Posttest*					
Academic Level	*N*	*M*	*SD*	*N*	*M*	*SD*	*Net Gain*	*t*	*p*
High	69	42.1	17.7	48	73.0	23.0	29.9	10.5	.0001
Low	78	15.2	11.3	68	69.3	24.6	54.1	17.6	.0001

TABLE 7
Net Gain, Standard Deviations, and *t* Tests of Higher Order Thinking Skill Scores by Academic Levels in Study 3

				Posttest Versus Pretest		*High Versus Low*	
Academic Level	*N*	*Net Gain*	*SD*	*t*	*p*	*t*	*p*
High	69	31.6	3.9	8.2	0.0001	–5.24	.0001
Low	78	55.7	4.2	13.2	0.0001		

profit. The curriculum development of the two modules evolved in a bottom-up fashion. Several experimental teachers noted that an important lesson had been that involvement of teachers in the development and assessment processes had positively affected the ability of these teachers to implement the approach that fosters higher order thinking skills through STS in their classes (Dori & Herscovitz, 1999; Dori et al., 2003).

STUDY 4: TEACHING CRITICAL AND SCIENTIFIC THINKING

Research Setting

The first stage in the TSC project was aimed at teaching critical and scientific thinking through carefully designed learning activities. Precisely because this was the first stage in the TSC project, it was the most traditional part of the project in terms of both the pedagogical means and assessment methods employed. The unit consists of learning activities that follow lab experiments, Invitations to Inquiry (Schwab, 1963), and critical assessment of newspaper clips (including advertisements). The TSC learning activities that follow lab experiments are based on "hands on" experimentation tasks emphasizing various elements of scientific reasoning and critical thinking that pertain to these tasks. Invitations to Inquiry present narratives describing real, historic, classical, or fictitious experiments; the story is divided into several segments that are presented to the student one at a time. Students are asked to "step into the shoes of the scientist" and solve various problems derived from each segment (Schwab, 1963). The third type of learning activity is critical assessment of newspaper clips or advertisements. Students were asked to read newspaper clips that relate to issues they had learned in science and to address a series of questions that lead to critical thinking regarding their content. The activities were taught in ways that match the spirit of critical thinking, including group and class discussions, problem solving, analysis of experiments, and handling data. Assessment of these activities consisted of multiple choice tests.

One of the units in that part of the project addressed the biological topic of water balance in living organisms. Seven thinking skills were selected as goals for this unit: identifying explicit and tacit assumptions, avoiding tautologies, isolating variables, testing hypotheses, identifying relevant information, recognizing logical fallacies, and differentiating between experimental results and conclusions. These skills were integrated into the relevant biological topics (through the TSC learning activities). Each of the seven skills was repeated between six and nine times throughout the unit (Zohar & Tamir, 1993; Zohar et al., 1994).

An evaluation study of that unit consisted of two groups (experimental and comparison) that studied the same biological topic and used the same textbook (addressing the issue of water balance in living organisms).

The research objectives were to find out whether (and to what extent) the teaching strategies used in the unit can:

- Contribute to the development of critical and scientific thinking in various biological topics.
- Contribute to the transfer of critical and scientific thinking skills to other (nonbiology) disciplines.
- Affect students' knowledge of the biological topics addressed in the unit.

A total of 21 seventh grade classes participated in this study divided between a comparison and an experimental group (10 classes were assigned to the experimental group and 11 classes were assigned to the comparison group). Student population was heterogeneous in terms of socioeconomical background. The topic was taught for about 24 periods in both groups. The comparison group studied the topic in a traditional manner, whereas the experimental group engaged in the TSC learning activities in addition to using the textbook.

Assessment Method

The effect of the program was assessed in three areas: students' reasoning skills, students' knowledge of biology, and teachers' feedback to the unit (Zohar et al., 1994). We used the following instruments:

1. Two parallel forms of a General Critical Thinking (GCT) test, which consists of 14 multiple choice items (with an option to justify the chosen response) that assess pre- and postperformance in the seven thinking skills listed earlier in the context of everyday reasoning. In developing the test, items and ideas from several sources were included (e.g., Jungwirth, 1985, 1987). One of the forms was used as a pretest and the other was used as a posttest.

2. The Biology Critical Thinking test (BCT), which is similar to the GCT test in its logical pattern, but addresses biological topics. An example of one item is presented in Figure 2. Cronbach's alpha reliability (internal consistency)

Students grew two bean plants, one in a temperature of 10°C and another in a temperature of 30°C. All other conditions were identical. After 3 weeks the plant grown in the 30°C temperature was bigger than the plant grown in the 10°C. What can we conclude?

1.	A temperature of 30°C is more suitable than 10°C for growing beans.
2.	The results are as expected because in general plants grow better in a warmer environment.
3.	Some bean varieties grow better in 30°C and other in 10°C.
4.	None of the above.

Explain: ____

FIGURE 2 An example of an item from the BCT test—Study 4.

indexes were .62 and .63 for the GCT and BCT, respectively. In critical thinking tests reliabilities tend to be relatively low, ranging from about .65 to .75 (Norris & Ennis, 1989). One way to increase the reliability is to increase the number of items. A combined score of the two tests was found to be significantly more reliable (α = .77).

3. A knowledge test that consists of 20 multiple choice items.
4. A follow-up of teachers' feedback to the unit through teachers' weekly reports and interviews at the end of the school year.

Findings

Achievements in the pretest were similar for the experimental and comparison groups, indicating that the initial reasoning level of students of both groups was the same. Comparing pretest to posttest scores, we found that students in the experimental group significantly improved their thinking skills relative to both their own initial level and to the level of students in the comparison group. Improved thinking skills were observed in tasks addressing a new biological context and nonbiological everyday topics, indicating transfer across domains. Students from the experimental group also scored significantly higher than the comparison group students on the knowledge test, suggesting that "learning facts" as one educational goal and "learning to think" as another need not conflict, but rather can support each other. Finally, the data from the teachers' weekly reports and interviews showed that teaching this unit decreased the frequency of teacher-centered teaching and enhanced a more active, student-centered learning (Zohar et al., 1994).

Low Versus High Achievers and the Critical and Scientific Thinking Module

As teachers were applying the learning activities in their classrooms, they noticed that students from all academic levels got involved in the animated group discussions that developed among small groups of students and between students and the teacher. Some of the teachers reported that students who had never before participated in class discussions raised their hands voluntarily for the first time during the class discussions that took place as part of the TSC learning activities.

To assess the effect of the critical and scientific thinking unit on students of different academic levels, students from both the experimental and comparison groups were divided into subgroups according to their biology grade (i.e., the final grade they had received in biology on the term before the study began). Then, the mean gain in students' scores in the GCT test was calculated separately for each subgroup of students (who received the same biology final grade). The results are presented in Table 8.

As expected, the mean posttest scores of students who were low achievers in biology were lower than those of students who were high achievers in biology. The better the students were in terms of their achievement in biology, the higher was their posttest score in the GCT test. Yet, the students of all subgroups made considerable progress with respect to their initial scores. Thus, it may be concluded that the unit contributed to developing the thinking skills of students from all academic levels.

TABLE 8
Gains of Students' of Different Academic Levels (According to Their Biology Grade) in the General Critical Thinking Test in Study 4

	Experimental Group[a]				*Comparison Group*[b]			
Biology Grade	*N*	*Pretest Score*	*Posttest Score*	*Gain*	*N*	*Pretest Score*	*Posttest Score*	*Gain*
4	—	—	—	—	4	30.4	32.7	2.3
5	11	35.7	66.2	30.5	14	45.9	37.2	–8.7
6	25	32.6	65.4	32.8	34	39.3	43.9	4.6
7	42	40.2	72.6	32.4	51	39.4	40.9	1.5
8	61	44.1	81.6	37.5	64	38.6	45.6	7.0
9	56	44.6	85.8	41.2	56	40.8	52.1	11.3
10	10	58.6	92.9	34.3	36	49.2	55.4	6.2

[a]*N* = 205. [b]*N* = 259.

CONCLUSIONS AND DISCUSSION

While participating in programs that focus primarily on fostering the thinking of disadvantaged students or students with low academic achievements, teachers have engaged these students in intensive thinking activities (e.g., Feuerstein, Rand, Hoffman, & Miller, 1980). However, research shows that when teachers participate in programs that are targeted toward a more general student population (i.e., schools or classes that are heterogeneous in terms of students' socioeconomic background and academic abilities), they often tend to engage low-achieving students in thinking activities less than the high-achieving ones (Raudenbush et al., 1993; Zohar et al., 2001). This tendency is likely to be motivated by good intentions: Teachers see higher order thinking tasks as difficult and highly demanding. Therefore, they refrain from assigning higher order thinking tasks to students whom, the teachers believe, will find such tasks hard and frustrating. Despite good intentions, this creates a vicious cycle: Precisely those students whose thinking skills need the most care and teacher attention get less attention from teachers than their high-achieving peers. Exposing teachers to empirical findings regarding this particular issue may contribute to changing their beliefs and habits.

The four studies described in this article shared the same general educational objective—fostering students' higher order thinking skills in the context of science and technology education. Each of the programs was unique in terms of its science content, specific reasoning goals, student population, and instructional and assessment means. Nevertheless, a similar pattern of findings recurred in all four studies. Students with both high and low academic achievements gained significantly from the educational interventions. Contrary to many practitioners' beliefs (Zohar et al., 2001) and to the findings of some previous studies (e.g., Welch, Klopfer, Aikenhead, & Robinson, 1981), our empirical evidence shows that instruction of higher order thinking skills is appropriate for students with high and low academic achievements alike.

These findings confirm the theories of teaching and learning for understanding, as well as the theoretical background for the STS approach that were described in the Theoretical Background section. These theories and the STS approach advocate that thinking is for *all* students (e.g., Bransford, Brown, & Cocking, 2000; Bruer, 1993; Dori & Herscovitz, 1999; Fensham, 1985; Perkins, 1992; Perkins & Unger, 1999; Resnick, 1987; Resnick & Klopfer, 1989; Resnick & Resnick, 1992; Yager & Tamir, 1993). As Solomon (1993) indicated in her book, *Teaching Science, Technology and Society*, "*All people* need some science education so that they can think, speak and act on those matters, related to science, which may affect their quality of living" (emphasis added; p. 15).

Solomon emphasized the need to foster all students' higher order thinking skills. Our research supports these theoretical views by showing empirically that it is indeed feasible to attain these goals not only for a selective section of the student population, but for *all* students.

Our studies show that by the end of the interventions students with high academic achievements gained higher reasoning scores than their peers with low academic achievements. This pattern was repeated in all four studies. This fact does not undermine the importance of our findings, because we are by no means suggesting that our treatments are guaranteed to close the gaps between low and high academic achievers. Our point is that by emphasizing the development of all students' thinking skills, the scientific and technological literacy of students at all academic levels may significantly improve relative to each student's initial starting point. In some cases the gap between low and high achievers can be narrowed.

In one of the studies, Case Studies in Biotechnology (Dori et al., 2003), the comparison between scores of low- and high-achieving students addressed knowledge and understanding of scientific concepts, in addition to scientific reasoning. Interestingly, by the end of the program, students who were initially classified as low academic achievers scored higher than students who were initially classified as high academic achievers in the knowledge and understanding category. Informal classroom observations and conversations with teachers indicated that the teachers who taught this unit tended to emphasize more higher order activities with students whom they considered academically "stronger," while emphasizing more drilling and recall of information with students whom they considered "weaker." These observations are in agreement with the findings of Raudenbush et al. (1993) and call for two remarks.

First, it may well be that all four studies were biased: Although all students supposedly went through the same program, in fact the "hidden curriculum" made teachers engage high-achieving students in more intensive higher order thinking than low-achieving students. Thus, if teachers would be educated to assign higher order thinking tasks equally to students at all levels, the "lower achievers" could make even greater progress in their thinking skills than our studies have shown. Second, this finding suggests that the emphasis science teachers place on teaching higher order thinking skills to high-achieving students may have caused these teachers to neglect the teaching of scientific concepts. Possibly, the traditional "lower achievers" may do better on a knowledge test because they were taught the science content more thoroughly. Ideally, teachers and students alike should target both of these learning objectives, rather than emphasize one at the expense of the other.

In all four studies reported in this article, research, development, and practice are interwoven, in line with the recommendations of Schoenfeld (1999). Aiming at teaching for understanding and higher order thinking skills, while using the methods described in these studies, we have reached both low and high academic level students and prepared them to function in the increasingly sophisticated environment of the world today, and more so tomorrow.

RESEARCH LIMITATIONS

Because this collection of four distinct studies was not originally designed to address the research goal we have been investigating here, we were somewhat limited in what we could analyze. Each study had different settings, research objectives, and variables, as well as assessment means. For example, we could not compare the knowledge and understanding category for each academic level in three of the four studies because we had not collected the data needed for such a comparison. We also could not compare specific thinking skills in all four studies because the different programs aimed at enhancing different thinking skills (e.g., question posing, formulating an argument and justifying it, system thinking, and critical thinking). Each program dealt with a subset of these skills.

Another research limitation is the sizes of the low and high academic level groups. In three out of the four studies (Study 3 was an exception), the number of students in the two groups was not balanced. There were more high academic level students than low ones because initially they were the majority in the programs within which we conducted our studies. This fact might have disadvantaged the low-achieving students. Claiming that higher order thinking is appropriate for all students does not imply that all students should be taught higher order thinking using the same methods. The study of teachers' beliefs about low-achieving students and higher order thinking (Zohar et al., 2001) showed that many of the teachers who believed that higher order thinking is appropriate for low achievers were not oblivious to their learning difficulties. While assessing these students' abilities realistically, the teachers did not consider these difficulties appropriate reasons for giving up on higher order thinking goals altogether. Instead, they were searching for ways to work toward these teaching goals by adapting special pedagogical means that included breaking up a complex task into simpler components, leading students through a sequence of steps necessary to solve a problem, giving clues, adding more examples, modeling ways for solving problems, and letting students work in groups of mixed abilities so that peers can learn from each other. However, teachers indicated that in heterogeneous classes they are often unable to teach in differentiated ways, targeting different instructional means for different types of students. Our hypothesis is that because many of the students in three of the studies were at intermediate or high academic levels, teachers were more attentive to their needs than to those of low academic achievers. In the Biotechnology project (Study 3), teachers paid special attention to learning difficulties of the low achievers. They therefore applied appropriate instructional methods that were especially beneficial for low-achieving students, inducing even larger gains in knowledge and understanding as well as higher order thinking skills.

The lack of balance in our studies between low and high academic achievers was further exacerbated due to higher attrition of low academic level students. Although this imbalance did not interfere with drawing statistically significant results, it may be argued that not all the low-achieving students in the classes we studied were equally represented in our findings. We therefore have to limit our findings by saying that a considerable group of low-achieving students gained significantly from our interventions. Further research is required to find out the extent by which subgroups of the low-achieving students are affected by projects such as the ones we have described.

On the other hand, the very fact that these four studies were so diverse is a source of strength for our conclusion, as we get four independent indications for the same phenomenon, namely that low academic level students benefit from engaging in education for higher order thinking as much as their peers that exhibit high academic achievements.

EDUCATIONAL IMPLICATIONS

Our findings bear educational significance for teacher development in the context of projects and programs that involve higher order thinking. The compelling empirical evidence shows that low-achieving students and higher order thinking are not mutually exclusive. This conclusion should be made an important element in the process of changing teachers' beliefs and practices in this field. Obviously, simply stating this conclusion is unlikely to be enough. We suggest structuring professional development regarding the issues discussed here around three main themes:

1. *Theoretical considerations*, explaining why our current views about the nature of teaching and learning and of the STS approach require that all students will be taught to think, as described in the theoretical background earlier.
2. *Empirical evidence*, such as the ones described in our findings, showing gains in thinking abilities of students from all academic levels.
3. *Practical tools* for helping students to accomplish tasks requiring higher order thinking even when these tasks may seem to be too difficult initially.

This final point is extremely important. Clearly, teachers are often correct in their belief that some tasks may be too difficult for some of their students, causing failure and frustration. However, instead of letting this belief lead to the prevalent conclusion that thinking tasks are just inappropriate for large sections of the student population, staff development programs may equip teachers with tools for helping students construct better abilities. Such practical tools may consist of the pedagogical means listed earlier as part of recommendations described in the study about teachers' beliefs. In addition, they may include the following means: modeling of thinking procedures, using metacognitive processes, peer learning, scaffolding and involving the teachers in the development of STS modules, and assessment instruments for their own classes (e.g., Dori & Herscovitz, 1999; Dori & Tal, 2000; White & Fredriksen, 1998, 2000; Zohar & Nemet, 2000, 2002). Incorporating these themes into professional development programs will be a step forward toward a more equitable education for *all* students.

ACKNOWLEDGMENTS

We would like to thank our former graduate students Orit Herscovitz, Flora Nemet, Revital Tali Tal, Masha Tsaushu, and Yehudit Weinberger for their contribution to the research described in this article. Thanks also to all the teachers and students who took part in the projects described.

REFERENCES

Adey, P. (1999). The science of thinking, and science for thinking: A description of Cognitive Acceleration through Science Education (CASE) [INNODATA Monographs–2]. Geneva, Switzerland: International Bureau of Education, UNESCO.

Adey, P., & Shayer, M. J. (1994). *Really raising standards*. London: Routledge.

Aikenhead, G. S., & Ryan, A. G. (1992). The development of a new instrument: "Views on Science-Technology-Society" (VOSTS). *Science Education*, *76*, 474–491.

Ben-Peretz, M. (1985). Curriculum potential. In T. Husen & N. T. Postlethwaite (Eds.), *The international encyclopedia of education* (Vol. 2c, pp. 1246–1248). Oxford, England: Pergamon.

Bloom, B. (1956). *Taxonomy of educational objectives: Handbook-I. Cognitive Domain*. New York: David McKay.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Research Council, National Academy Press.

Brown, A. L., & Campione, J. C. (1990). Communities of learning and thinking, or a context by any other name. *Contributions to Human Development, 21,* 108–126.

Bruer, J. T. (1993). *Schools for thought*. Cambridge, MA: MIT Press.

Burden, R., & Williams, M. (1998). *Thinking through the curriculum*. London and New York: Routledge.

Bybee, R. Y. (1987). Science education and the Science-Technology-Society (S-T-S) Theme. *Science Education, 71*, 667–680.

Bybee, R. Y. (1993). *Reforming science education, social perspectives, and personal reflections*. New York and London: Teachers College Press.

Bybee, R. Y., & Ben-Zvi, N. (1998). Curriculum change in science: Transforming goals to practice. In B. Fraser & K. Tobin (Eds.), *International handbook of science education* (pp. 487–498). Dordrecht, The Netherlands: Kluwer Academic.

Carmichael, J. W. (1981). *Project SOAR (Stress on Analytical Reasoning) instructor's manual.* New Orleans: Xavier University of Louisiana.

Chance, P. (1986). *Thinking in the classroom: A survey of programs*. New York: Teachers College Press.

De Bono, E. (1985). The Cort thinking program. In J. W. Segal, S. F. Chipman, & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 1, pp. 389–416). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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., & Herscovitz, O. (1999). Question posing capability as an alternative evaluation method: Analysis of an environmental case study. *Journal of Research in Science Teaching, 36*, 411–430.

Dori, Y. J., & Hofstein, A. (2000). The development, implementation and initial research findings of "Science and Technology for All" in Israel. Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education. (ERIC Document Reproduction Service No. ED439955, SE 063454)

Dori, Y. J., & Tal, R. T. (2000). Formal and informal collaborative projects: Engaging in industry with environmental awareness. *Science Education, 84*(1), 95–113.

Dori, Y. J., Tal, R. T., & Tsaushu, M. (2003). Teaching biotechnology through case studies—Can we improve higher order thinking skills of non-science Majors? Manuscript submitted to *Science Education*.

Driver, R., & Leach, J. (1993). A constructivist view of learning: Children's conceptions and the nature of science. In R. E. Yager (Ed.), *What research says to the science teacher: The science, technology, society movement* (Vol. 7, pp. 103–112). Washington, DC: National Science Teachers' Association.

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

Fensham, P. (1985). Science for all. *Journal of Curriculum Studies*, *17*, 415–435.

Feuerstein, R., Rand, Y., Hoffman, M. B., & Miller, R. (1980). *Instrumental enrichment and intervention program for cognitive modifiabilty.* Baltimore: University Park Press.

Feurstein, R., Rand, Y., & Rynders, J. E. (1988). *Don't accept me as I am*. New York and London: Plenum.

Gagne, R. M. (1974). *The conditions of learning* (2nd ed.). New York: Holt, Rinehart & Winston.

Greeno, G. G., & Goldman, S. V. (Eds.). (1998). *Thinking practices in mathematics and science learning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Graesser, A. C., Baggett, W., & Williams, K. (1996). Question-driven explanatory reasoning. *Applied Cognitive Psychology, 10*, 17–31.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal, 31*, 104–137.

Greeno, G. G., & Goldman, S. V. (Eds.). (1998). *Thinking practices in mathematics and science learning*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Halpern, D. F. (1992). *Enhancing thinking skills in the sciences and mathematics*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Harari, H. (1992). *"Tomorrow 98."* Report of the Supreme Committee for Science and Technological Education, Israel Ministry of Education, Jerusalem (in Hebrew).

Herried, C. F. (1994). Case studies in science: A novel model of science education. *Journal of College Science Teaching, 23*, 349–355.

Hofstein, A., Aikenhead, G., & Riquarts, K. (1988). Discussion over STS at the 4th IOSTE symposium. *International Journal of Science Education, 10*, 357–366.

Jungwirth, E. (1985). Science teaching & pupil-avoidance of logical fallacies. *South African Journal of Education, 5*(2), 55–60.

Jungwirth, E. (1987). Avoidance of logical fallacies: A neglected aspect of science education and science-teacher education. *Research in Science and Technological Education, 5*(1), 43–58.

Levine, D. U. (1993). Reforms that can work. *American School Board Journal, 180*(6), 31–34.

Lipman, M. (1985). Thinking skills fostered by philosophy for children. In J. W. Segal, S. F. Chipman, & R. Glaser (Eds.), *Thinking and learning skills* (Vol. 1). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

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

Metz, M. H. (1978). *Classrooms and corridors: The crisis of authority in desegregated secondary schools*. Berkeley: University of California Press.

National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.

Newmann, F. M. (1990). Higher order thinking in teaching social studies: A rationale for the assessment of classroom thoughtfulness. *Journal of Curricular Studies, 22*(3), 53–75.

Nickerson, R., Perkins, D., & Smith, E. (1985). *The teaching of thinking*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Norris, S. P., & Ennis, R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications.

Oakes, J. (1990). *Multiplying inequalities: The effects of race, social class and tracking on opportunities to learn math and science*. Santa Monica, CA: Rand.

Page, R. N. (1990). The lower track curriculum in a college preparatory high school. *Curriculum Inquiry, 20*, 249–281.

Pedretti, E., & Hodson, D. (1995). From rhetoric to action: Implementing STS education through action research. *Journal of Research in Science Teaching*, *32*, 463–485.

Perkins, D. N. (1992). *Smart schools—From training memories to training minds*. New York: Free Press.

Perkins, D. N., & Grotzer, T. A. (1997). Teaching intelligence. *American Psychologist, 52*, 1125–1133.

Perkins, D. N., & Unger, C. (1999). Teaching and learning for understanding. In C. M. Reigeluth (Ed.), *Instructional design theories and models* (pp. 91–114). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Peterson, P. L. (1988). Teaching for higher order thinking in mathematics: The challenge for the next decade. In D. A. Grows & T. J. Cooney (Eds.), *Perspectives on research on effective mathematical learning* (Vol. 1, pp. 2–26). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Pogrow, S. (1988). Teaching thinking to at-risk elementary students. *Educational Leadership, 45*, 79–85.

Pogrow, S. (1996). HOTS: Helping low achievers in grades 4–7. *Principal, 76*(2), 34–35.

Raudenbush, S. W., Rowan, B., & Cheong, Y. F. (1993). Higher order instructional goals in secondary schools: Class, teacher, and school influences. *American Educational Research Journal, 30*, 523–555.

Resnick, L. (1987). *Education and learning to think*. Washington, DC: National Academy Press.

Resnick, L., & Klopfer, L. (1989). Toward the thinking curriculum: An overview. In L. Resnick & L. Klopfer (Eds.), *Toward the thinking curriculum: current cognitive research: Yearbook of the Association for Supervision and Curriculum Development*. Association for Supervision and Curriculum Development.

Resnick, L. B., & Resnick, D. P. (1992). Assessing the thinking curriculum: New tools for educational reform. In B. R. Gifford & M. C. O'Connor (Eds.), *Changing assessments: Alternative views of aptitude, achievement and instruction* (pp. 37–75). Boston: Kluwer.

Sabar, N., & Shafriri, N. (1982). On the need for teachers training in curriculum development. *Curriculum Inquiry, 7,* 307–315.

Schoenfeld, A. (1992). Learning to think mathematically. In D. A. Grouws (Ed.), *Handbook of research in mathematics teaching and learning* (pp. 334–370). New York: Macmillan.

Schoenfeld, A. (1999). Looking toward the 21st century: Challenges of educational theory and practice. *Educational Researcher, 28*(7), 4–14.

Schwab, J. J. (Ed.). (1963). *Biology teacher's handbook*. New York: Wiley.

Shepard, L. (1991). Psychometricians' beliefs about learning. *Educational Researcher, 20*(7), 2–9.

Shepardson, D. P. (1993). Publisher-based science activities of the 1980s 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.

Tal, R. T., Dori, Y. J., Keiny, S., & Zoller, U. (2001). Assessing conceptual change of teachers involved in STES education and curriculum development—The STEMS Project Approach. *International Journal of Science Education, 23*, 247–261.

Tal, R. T., Dori Y. J., & Lazarowitz, R. (2000). A project-based alternative assessment system. *Studies in Educational Evaluation, 26*, 171–191.

Tishman, S., Perkins, D., & Jay, E. (1995). *The thinking classroom*. Boston: Allyn & Bacon.

Wallace, J., & Louden, W. (1998). Curriculum change in science: Riding the waves of reform. In B. J. Fraser & K. G. Tobin (Eds.), *International handbook of science education* (pp. 471–487). Dordrecht, The Netherlands: Kluwer.

Welch, W. Y., Klopfer, L. E., Aikenhead, G. L., & Robinson, J. T. (1981). The role of inquiry in science education: Analysis and foundations. *Science Education*, *65*, 33–50.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling and metacognition: Making science accessible to all students. *Cognition and Instruction, 16*(1), 3–118.

White, B. Y., & Fredriksen, J. R. (2000). Metacognitive facilitation: an approach to making science inquiry accessible to all. In J. Minstrell & E. H. van Zee (Eds.), *Inquiring into inquiry learning and teaching in science* (pp. 331–370). Washington, DC: American Association for the Advancement of Science (AAAS).

Yager, R. E., & Tamir, P. (1993). STS approach: Reasons, intentions, accomplishments, and outcomes. *Science Education, 77*, 637–658.

Zohar, A. (1996). Transfer and retention of reasoning skills taught in biological contexts. *Research in Science and Technological Education*, *14*, 205–209.

Zohar, A. (1999). Teachers' metacognitive knowledge and instruction of higher order thinking. *Teaching and Teachers' Education, 15*, 413–429.

Zohar, A., Degani, A., & Vaaknin, E. (2001). Teachers' beliefs about low achieving students and higher order thinking. *Teaching and Teachers' Education, 17*, 469–485.

Zohar, A., & Nemet, F. (2000). Fostering students' argumentation skills through bio-ethical dilemmas in genetics. In B. Anderson, U. Harmes, G. Hellden, & M. L. Sjobeck (Eds.), *Researcher in didaktik of biology* (pp.181–190). Proceedings of the second conference of researchers in didaktik of biology 1998, November 18–22. Goteborg, Germany: University of Goteborg,

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching, 39*, 35–62.

Zohar, A., & Tamir, P. (1993). Incorporating critical thinking within a regular high school biology curriculum. *School Science and Mathematics, 93*(3), 136–140.

Zohar, A., Weinberger, Y., & Tamir, P. (1994). The effect of the biology critical thinking project on the development of critical thinking. *Journal of Research in Science Teaching, 31*, 183–196.

Zoller, U. (1986/1987). The Israeli environmental education project: A new model of interdisciplinary student-oriented curriculum. *Journal of Environmental Education, 18*(2), 25–31.

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