Encouraging and Analyzing Student Questions in a Large Physics Course: Meaningful Patterns for Instructors

Kathleen A. Harper,[1] Eugenia Etkina,[2] Yuhfen Lin[1]

[1]*Department of Physics, The Ohio State University, 174 W. 18th Avenue, Columbus, Ohio 43210*

[2]*Graduate School of Education, Rutgers, The State University of New Jersey, 10 Seminary Place, New Brunswick, New Jersey 08901*

Received 11 February 2002; Accepted 12 November 2002

Abstract: In a large introductory physics course, structured weekly journals (weekly reports) regularly encouraged students to ask questions about the material. The resulting questions were collected for one quarter and coded based on difficulty and topic. Students also took several conceptual tests during the quarter. The reports contained more questions than typically observed in a college classroom, but the number of questions asked was not correlated to any measure of conceptual performance. Relationships among different types of questions and performance on these tests were explored. Deeper-level questions that focus on concepts, coherence of knowledge, and limitations were related to the variance in student conceptual achievement. © 2003 Wiley Periodicals, Inc. J Res Sci Teach 40: 776–791, 2003

Throughout education, the vast majority of questions asked in a classroom setting are asked by teachers, not students (Good, Slavings, Harel, & Emerson, 1987; Graesser & Person, 1994), and the few questions that students do ask typically are unsophisticated (Graesser, McMahen, & Johnson, 1994). Indeed, ‘‘children everywhere are schooled to become masters at answering questions and to remain novices at asking them’’ (Dillon, 1990, p. 7).

In science, generating questions is arguably the most important part of the process, but often courses for beginning scientists and engineers focus primarily on answers (Marbach-Ad & Sokolove, 2000; Orr, 1999). There are many reasons why student question-asking should become more fundamental in science courses. Encouraging and emphasizing question-asking better exposes students to the fundamental inquiry nature of science (Marbach-Ad & Sokolove, 2000). Students who ask questions retain material better than those who do not (Davey & McBride, 1986; King, 1989; Marbach-Ad & Sokolove, 2000). Links exist between improved question-asking

Contract grant sponsor: NSF; Contract grant number: DUE-0088906.
Correspondence to: K.A. Harper; E-mail: harper.217@osu.edu
DOI 10.1002/tea.10111
Published online in Wiley InterScience (www.interscience.wiley.com).

© 2003 Wiley Periodicals, Inc.

ability and improved problem solving (Dori & Herscovitz, 1999; King, 1991). Furthermore, promoting student question-asking aids in developing independent learners (Marbach-Ad & Sokolove, 2000). The instructor also gains insight from listening to or reading the students' questions; he or she can identify common misconceptions and difficulties the class is experiencing, answer some individual questions, and modify subsequent instruction to address issues raised by them (Etkina, 2000; Etkina & Harper, 2002; Heady, 1993). In addition, student questions may allow the instructor, as well as other students, to view the material from a new perspective, possibly resulting in enhanced understanding (Marbach-Ad & Sokolove, 2000).

Why do people ask questions? Asking a meaningful question implies not only a need for information, but also enough knowledge structure to formulate it and understand the answer (Miyake & Norman, 1979). The process of asking questions allows one to verbalize his or her current understanding of a topic and connect that bit of knowledge with other ideas (Marbach-Ad & Sokolove, 2000). One must have some basic understanding of or belief about a topic before he or she can ask a question regarding it. Therefore, in the classroom it is important that the entire question and answer process of instruction be tied closely to students' knowledge state or they will not benefit (Macmillan & Garrison, 1983). Teaching difficulties may primarily be due to the teacher's failure to realize and to help students answer the questions they really have (Dillon, 1990; Macmillan & Garrison, 1983). If an instructor encourages and reacts to questions, students realize they are an important element of learning science. An investigation of two differently taught biology classes found that in both cases, giving students many opportunities to ask questions improved their study skills. In addition, providing these students with a taxonomy of questions resulted in improved questions from them; they were "more insightful, thoughtful, and content-related, . . . not easily answered by consulting the textbook" (Marbach-Ad & Sokolove, 2000, p. 854).

> Since asking questions is fundamental in science, it should not be surprising that, a major feature of science curriculum reform is educators' growing interest in the modes, processes, and procedures of student assessment and examinations that promote, develop and measure high-order cognitive skills. . . . A wide consensus exists among researchers that the development of students' abilities to ask questions, reason, problem solve and think critically should become a central focus of current science education reform. (Zoller, Tsaparlis, Fatsow, & Lubezky, 1997, p. 99)

This article discusses a method for fostering student questions and an analysis of the questions that resulted.

One tool that may aid students in developing their question-asking ability is the weekly report (Etkina, 2000). A weekly report is a structured journal written by students each week in which they answer three questions: (a) What did I learn this week and how did I learn it? (b) What questions remained unclear? (c) If I were the professor, what questions would I ask my students to find out if they understood the material? Question b specifically encourages students to think about the gaps in their current knowledge state and ask questions to fill in these gaps. The instructor uses these questions to modify the subsequent instruction to address students' needs. These modifications may include specially designed exercises, activities, or additional questions. Although students may not submit all their questions, this helps close the feedback loop between student and instructor (Etkina & Harper, 2002). Whereas Questions a and c also help students reflect on their learning, this research focused on the second question and its potential as an assessment tool. This study describes and assesses questions asked by students in responding to this second prompt, addressing the following questions:

1. Do questions that students ask reveal the same difficulties in learning as more traditional assessment instruments such as tests and interviews?
2. What types of questions do students ask, and are there relationships between different types of questions?
3. Is there a relationship between the types of questions asked in the weekly reports and a student's conceptual understanding?

A great deal of work has been done by different researchers studying student questions. King reported that training students to ask specific questions on lecture comprehension improved their performance (King, 1989). The tutoring work of Graesser and Person describes the results of encouraging students to ask questions (Graesser & Person, 1994). At the same time, most work done at the collegiate level dealt with questions that students periodically had about reading assignments. Our study is different because it solicited questions about any aspect of the course and the students were provided with the opportunity to ask questions, but were not required to do so. Indeed, 30% of the reports contained no questions.

Course Description

Weekly reports were implemented in the Freshmen Engineering Honors introductory mechanics course at Ohio State University. The course lasted for one 10-week quarter and each week consisted of three large-room meetings (48 minutes each), two recitations (48 minutes each), and a laboratory (108 minutes). In Fall 2000, about 200 students, not all of them subjects in the study, were divided between two sections for large-room meetings. Students were in the same class of approximately 28 students for recitation and lab and worked in 4-person collaborative groups assigned by a stratified random process (Johnson, Johnson, & Smith, 1991).

The quarter was broken into cycles of approximately 2 weeks' duration, each emphasizing a different conceptual area: kinematics, linear dynamics, two-dimensional dynamics, energy, momentum, and statics. An instructional approach based on the scientific process was used for each conceptual area (Etkina & Van Heuvelen, 2001). Students constructed their understanding of physics concepts and relationships between the physical quantities through observations of carefully selected phenomena without explanation from the professor. They recorded their observations and generated qualitative explanations. Then they used these explanations to predict the outcome of a new experiment. If the experimental outcomes did not match the predictions, the explanations were revisited and modified by the students. If their predictions were correct, more observational experiments followed. The initial experiments were presented again so the students could identify physical quantities to describe phenomena and find quantitative relationships between them if possible. These relationships were tested again in new experiments.[1]

Activities in recitations and labs were structured to help students strengthen their conceptual understanding, analytical skills, and experimental skills (Table 1). Students learned to use different representations of physics ideas such as pictorial representations, motion diagrams, graphs, vector diagrams, free-body diagrams, and energy bar charts before they solved numerical problems (Van Heuvelen, 1991; Van Heuvelen & Zou, 2001). In problem-solving sessions they solved context-rich problems (problems that are ill-defined, are based on real-life situations, and may require estimations) (Heller & Hollabaugh, 1992) and jeopardy problems (problems in which students need to formulate a question for which the answer is provided in a form of equation or diagram) (Van Heuvelen & Maloney, 1999).

Each week students submitted weekly reports in which they reflected on the application of knowledge in a lab based on a previous cycle and on construction of the new cycle's knowledge in

Table 1
Summary of learning cycle used in FEH physics course

Part of Cycle	When and Where It Took Place
Making observations, developing qualitative (conceptual) explanations (models), designing testing experiments	First large-room meeting of the week (lecture) (48 min)
Developing conceptual understanding, qualitative problem solving	First small-room meeting (recitation) (48 min)
Inventing physical quantities, developing quantitative models (laws), designing more testing experiments	Second large-room meeting (lecture) (48 min)
Quantitative problem solving, context rich problem solving	Second small-room meeting (recitation) (48 min)
Applications of the law, complex problem solving	Third large-room meeting (lecture) (48 min)
Testing, application, design	Lab (2 h)

lectures and recitations. Students submitted these reports through an on-line system developed at The Ohio State University. They were due by Monday night, and two graders read and responded to the reports by Friday. They answered students' questions from Part b and provided feedback on the quality of responses to Questions a and c. Half of the reports each week received a numerical grade. An average report was about one page long and contained one question in Part b. A few reports were as long as six pages and some asked as many as five questions. The professors addressed common questions and problems (which emerged with about a 30 minute/week review of the reports) through activities in large room meetings or in recitations. In addition, they used some of the questions from Part c on the exams.

Analysis

Analysis of Questions during the Semester

The first part of the study took place during the fall quarter of 2000 when the graders read the reports and responded to students' questions. Every week they selected the most typical questions and made a list of the questions for the professors teaching large-room meetings. The professors then used the questions to design lecture and recitation activities to address student difficulties. Analysis of the content of the questions revealed that students asked a wide variety of questions, varying in both complexity and objectives. The questions yield interesting information about students' knowledge states (LaFrance, 1992). This will be further discussed in the Results section.

Question Coding

Analysis of questions from the perspective of difficulties of teaching and learning physics during the quarter led to the development of a coding scheme. Student questions were collected for the final 8 weeks of the quarter and coded. (Data from the first 2 weeks were discarded owing to difficulties with getting the on-line collection system up and running. As we found no significant differences from one week to the next, this does not affect our results.) The coding scheme, described in detail below, was developed. Then, two authors independently coded a small sample of the questions (about 100) and agreed on the coding for about 90% of this sample before one author coded the remainder of the questions. The coder did not know the students in the class. Each question was double-coded, depending first on the level of difficulty (Etkina, 2000) and second on its topic. If a student submitted more than one question during a particular week, each question was coded separately. Questions dealing with administrative issues represented $<1\%$ of the total number of questions and were not included in this study. We define a question as an inquiry (van der Meij, 1994). It does not have to be phrased interrogatively to be seeking new information (e.g., ‘‘I was wondering what exactly the difference between momentum and kinetic energy is.’’). Table 2 shows the taxonomy of the questions. All sample questions throughout the article are from actual student reports.

Coding of the questions according to their level of difficulty is similar to the taxonomy developed by Marbach-Ad and Sokolove for use in an introductory biology course (Marbach-Ad & Sokolove, 2000), as well as that of Graesser, Person, and Huber for an analysis of verbal questions (Graesser, Person, & Huber, 1992). In addition, this scheme fits well with Bloom’s taxonomy (Bloom, 1956). Questions of minimal difficulty are looking for factual knowledge, low-level questions seek better comprehension, medium-level questions deal with application or analysis, and high-level questions contain elements of synthesis or evaluation.

There are some similarities between the goal coding, or coding questions according to their topics, developed here and the coding devised by others for coding verbal questions in Grade Kindergarten through 12 classes (Good et al., 1987). The differences emerge from the different populations observed and from the different collection methods. Where Good et al. found it necessary to keep track of noncontent questions for their research, these were not relevant to our study.

Conceptual Measures

Students were given the Force Concept Inventory (FCI) (Hestenes, Wells, & Swackhamer, 1992) as a pre- and posttest, as well as the Mechanics Baseline Test (MBT) (Hestenes & Wells, 1992) as a posttest. The MBT appeared as part of the final exam for the course, whereas the FCI was given in recitation (for no grade). As part of the summative evaluation of the course, scores from these tests were used to indicate students’ conceptual understanding. The FCI is a purely conceptual test, whereas the MBT combines concepts with some basic problem solving. The normalized FCI gain, or Hake factor, was calculated to indicate the conceptual gain of the students during the course (Hake, 1998). The Hake factor is calculated as a percentage, or a fraction of 1. It shows the fraction of potential gain achieved by a student.

$$g = \frac{\text{Posttest score} - \text{Pretest score}}{1 - \text{Pretest score}}$$

This class had an average FCI pretest score of 53%, posttest of 79%, and gain of 0.53. This puts the class solidly into Hake’s interactive engagement zone, which starts at $g = .3$. Traditionally taught classes typically have gains of about 0.23 (Hake, 1998). Similarly, their MBT score of 73% compared favorably with the national average MBT posttest score for university physics of about 60%.[2]

Initial analysis showed that 24 students, or roughly 13% of the sample, never asked questions. These students were excluded from the study. Because this group was representative of the whole class, as measured by the distribution of exam scores and final course grades, this did not affect our conclusions. This actually yields an early result: It is not possible to predict the performance of a student based only on the fact that he or she asks no questions in weekly reports. In addition,

Table 2
Codes and examples of students' questions corresponding to code indications

Type of Coding	Code Indication	Description	Example
Level of difficulty	Minimal	The question indicates that a student has a definitional problem, essentially asking, "What is it?"	"What are the units for power?"
	Low	The student has a problem understanding the material presented by the instructor, and is asking, "What were you trying to tell me?"	"What was the purpose of the lab experiment dealing with free fall and using a ruler?"
	Medium	The student has a procedural difficulty, asking, "How do you do something?"	"I am still confused on what to choose as a system when you are doing a free-body diagram."
	High	The student wants to know why something happens or how new knowledge is connected to other areas	"Why is it that the moon, although it is constantly applying a gravitational force to the earth, doesn't pull the earth off its orbit around the sun?"
Topic	Equation	The student is concerned with the meaning of or use of an equation	"In the power equation how is work equal to change in energy?"
	Concept	The student is concerned with the meaning of a concept or the extension of the concept into a different area and tries to make sense out of it	"What does a negative work mean?"
	Application	The student tries to relate the concepts to her everyday experience or other physics concepts	"What practical every day problem can be solved using the shell theorem?" or "How is Newton's third law related to what we did this week?"
	Knowing	The student is interested in how a particular piece of knowledge was constructed by scientists	"When Newton first did his work involving the moon and its rotational velocity, how did he know the distance from the center of the earth to the center of the moon?" [sic]
	Experiment	The student did not understand the purpose, technical details or the explanation of an experiment	"I did not understand the happy and sad ball experiment."
	Limitations	The student is interested in the range of applicability of a concept, constant, etc.	"How far away from the earth do you have to get before there is a significant difference between the acceleration due to gravity and g [9.8 m/s^2]?"

comparison of posttest FCI scores with MBT scores and pretest FCI scores indicated that several students chose not to take the post-FCI seriously. They, too, were removed from the study. This brought the total student sample size to 158.

Results

This section first discusses the types of questions asked by students in the weekly reports, as well as their usefulness to the instructor. Then it describes the distribution of questions and relationships among the different question coding categories. Finally, it explores the relation of questioning to the FCI and MBT results.

Type of Questions

Students' questions reflected difficulties students experience learning physics that previously have been uncovered through different assessment techniques, such as interviews and conceptual tests. The information is summarized in Table 3. In addition to demonstrating conceptual difficulties similar to the ones reported in the literature, students' questions also showed their struggle with such aspects of physics as coherence (Hammer, 1989, 1994) and applicability, or connection of physics to the outside world (Redish, Saul, & Steinberg, 1998). An example of a

Table 3
Conceptual difficulties noted in weekly reports

Student Question	Difficulty	Reported Before by
"I don't understand when the word *energy* is used and when the word *work* is used."	Language and comprehension	Arons, 1997
"How does acceleration relate to velocity when the velocity is not constant?"	Confusion between velocity and acceleration	Halloun & Hestenes, 1985a; Trowbridge & McDermott, 1981
"I am still unclear on exactly what momentum is. What kind of force is it?"	Confusion of physical quantities, particularly force, with others	Kass & Lambert, 1983
"In the lab, the wheels of the monster truck were moving in an opposite direction to the frictional force. If this were the case and the truck wheels were not moving in the opposite direction I would understand"	Student does not yet have a clear understanding of Newton's third law	Brown, 1989
"Isn't there an inertial force that resists motion?" or "Why does [a projectile] always have a negative vertical acceleration due to gravity when it needs a positive acceleration to travel up?"	Impetus idea	Clement, 1982; Gamble, 1989; Halloun & Hestenes, 1985b; McCloskey, 1983
"How can a force be exerted by nothing? How is gravity causing everything to accelerate?"	Difficulty with a field concept	Minstrell & Stimpson, 1986

coherence question is, ''Why knowing power is important? Can work-energy equations be applied to solve most motion problems or applying Newton's second law is still a big part?'' [sic]. An example of applicability is the question, ''Is there such [a] thing as a perfect elastic collision in the real world?'' Some students ask questions that combine their search for both coherence and applicability: ''Why do bumpers use rubber if inelastic collisions are safer?''

Figure 1 summarizes the entire quarter's distribution of questions. A total of 1274 questions were coded. As the chart shows, 470 questions (roughly 37% of the total questions asked) were about concepts, more than about any other topic. Within this area, questions were distributed fairly equally among the four difficulty levels. The next most common type, equation, contained 433 (about 34% of the questions). For this type, the average level of difficulty is low (1.9); nearly 44% of these questions are of minimal difficulty. Seventeen percent, or 212, of the questions were categorized as coherence, but two thirds of these are of medium to high difficulty. Only 57 questions (<4.5%) are about limitations, but roughly 80% of these are medium- or high-difficulty questions.

Relationships among Question Types

We were interested in whether students who asked particular types of questions tended to ask other sorts of questions. In addition, we wanted to see whether students who asked some kinds of questions did not tend to ask others; thus, we performed a correlational analysis between question types. Investigating the relations among the different question topic areas yielded some interesting results. There are small correlations between questions about concepts and questions addressing the justification of knowledge ($r = 0.177$, $p < .026$), limitations ($r = 0.191$, $p < .016$), and coherence ($r = 0.364$, $p < .001$). One possible explanation for this is that students who tend to focus on concepts want to know where they come from, when they can be applied, and how they fit with other pieces of their knowledge. An additional correlation exists between coherence and

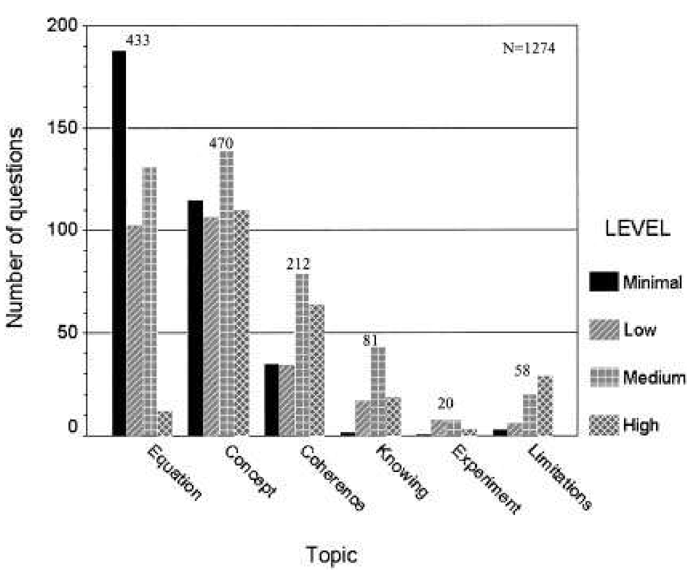

Figure 1. Distribution of questions by topic and difficulty level.

justification of knowing questions ($r = 0.227$, $p < .004$). This probably is because both types of questions seek to make sense of one's current factual knowledge.

Relations among Questions and Conceptual Performance

We then analyzed the entire sample's questions and test scores. Results are shown in Table 4. Perhaps the most striking result is that the raw number of questions asked has no significant correlation with any measure of conceptual achievement. However, the average level of difficulty is correlated weakly with the MBT score ($r = 0.288$, $p < .0002$). The percentage of minimal level questions is correlated negatively with the MBT results ($r = -0.220$, $p < .005$), whereas students who asked high-level questions tended to receive better scores on the MBT ($r = 0.261$, $p < .001$). Students who asked only simple questions tended not to score well on this conceptual test. This resonates with previous results (Graesser & Person, 1994).

Not only the difficulty level, but also the topic of the questions is related to student conceptual achievement, as might be expected. Students who ask a high percentage of questions about equations do not do as well on the posttests ($r = -0.173$, $p < .03$ for post-FCI, $r = -0.169$, $p < .034$ for MBT). This finding further confirms that although many students feel they must focus on equations to understand physics (Hammer, 1989), those who do so usually do not excel on conceptual tests. Furthermore, the post-FCI score correlates with the percentage of coherence questions asked ($r = 0.170$, $p < .033$). However, an analysis of pretest scores shows that students who score well on the pre-FCI also are likely to ask more coherence questions ($r = 0.225$, $p < .005$) and that there is no correlation between FCI gain and these questions. Therefore, we conclude that students with a fairly solid conceptual base coming into the course tend to ask questions that help them connect various pieces of their knowledge. In addition, students with high pre-FCI scores asked higher-level questions ($r = 0.177$, $p < .026$ between pre-FCI and average difficulty level; $r = 0.207$, $p < .009$ between pre-FCI and high-level questions). Again, this seems to indicate that students with more previous conceptual knowledge ask deeper questions.

Effects of Prior Conceptual Knowledge on Questioning

Given the fact that questioning behavior depends on the relation of the presented material to one's incoming knowledge base (Miyake & Norman, 1979), the population was split into two

Table 4
Correlations between question coding categories and conceptual tests (normalized)

	FCI Pre	FCI Post	FCI Gain	MBT
Minimal level	−0.116	−0.088	−0.032	−0.220**
Low level	−0.046	0.005	0.075	−0.071
Medium level	−0.005	−0.018	−0.077	0.084
High level	0.207**	0.132	0.058	0.261**
Average level	0.177*	0.113	0.020	0.288**
No. of questions	0.050	0.028	0.041	0.000
Formula	−0.139	−0.173*	−0.072	−0.169*
Concept	−0.008	0.120	0.138	0.087
Coherence	0.225**	0.170*	0.078	0.092
Knowing	−0.052	−0.154	−0.201*	−0.052
Experiment	−0.098	−0.024	−0.015	0.032
Limitations	0.144	0.085	−0.032	0.136

Note. Whole class ($N = 158$).
*$p < .05$, **$p < .01$.

groups according to FCI pretest scores. The low group consisted of students who scored $\leq$60%, whereas the high group had scores $>$60%. Sixty percent was chosen as the break point because it is considered to be the entry threshold for thinking Newtonianly (Hestenes et al., 1992).

The low pretest group contained 108 students. Table 5 lists correlations for this subgroup. Within this group, the only significant relations between the difficulty of the questions and test scores were those between the MBT and high-level questions ($r = 0.229$, $p < .017$) and between the MBT and average difficulty ($r = 0.208$, $p < .030$). The percentage of questions asked about concepts correlated weakly with both the pre- and posttest results ($r = 0.202$, $p < .036$ for pre, $r = 0.218$, $p < .024$ for post). This seems to indicate that students who are lacking in conceptual knowledge and ask questions to fill in the gaps score better on subsequent conceptual tests than those who do not. The pretest score also correlates negatively with questions about experiments ($r = -0.193$, $p < .046$), indicating that students who come in with the weakest conceptual background ask more questions about experiments done in class than those approaching the Newtonian way of thinking. One reason for this may be that students must understand the experiment before they can begin to understand the concept it illustrates. In addition, there is a negative correlation between the pretest score and the percentage of equation questions asked ($r = -0.200$, $p < .038$), showing that more of the equation-based questions came from students with very low pretest scores than students approaching the Newtonian threshold. There is also a correlation for this group between the number of conceptual questions and the number of limitations questions ($r = 0.267$, $p < .005$). Perhaps this means students with initially weak understanding who ask enough conceptual questions progress to the point where they can begin thinking about the limitations of the concepts. Also, students who ask more questions about equations tend to ask more about the experiments ($r = 0.205$, $p < .034$).

The high pretest group, 50 students, exhibited different behaviors, as shown in Table 6. The FCI pretest correlates negatively with low-level questions ($r = -0.350$, $p < .013$) and positively with the medium-level ones ($r = 0.409$, $p < .003$), indicating that students entering the class just past the Newtonian threshold asked more lower-level questions than those with a stronger conceptual background, who tended to ask questions of medium difficulty. The average difficulty level correlated with the MBT ($r = 0.378$, $p < .007$). Coupled with the fact that students in this group who asked a high percentage of minimal level questions did not perform well on the MBT ($r = -0.351, p < .012$), this indicates that students with prior knowledge who ask deeper questions continue learning. In addition, there is a significant correlation between the percentage of

Table 5
Correlations for low pretest (score <60%) group (normalized)

	FCI Pre	FCI Post	FCI Gain	MBT
Minimal level	−0.043	0.002	0.010	−0.150
Low level	0.080	0.059	0.023	−0.024
Medium level	−0.153	−0.077	−0.030	0.004
High level	0.159	0.027	−0.002	0.229*
Average level	0.059	−0.014	−0.017	0.208*
Formula	−0.200*	−0.174	−0.136	−0.147
Concept	0.202*	0.218*	0.166	0.149
Coherence	0.089	0.066	0.042	−0.042
Knowing	−0.032	−0.145	−0.130	−0.061
Experiment	−0.193*	−0.011	0.051	0.058
Limitations	0.110	0.000	−0.006	0.143

Note. $N = 108$.
*$p < .05$.

Table 6
Correlations for high pretest (score $>60\%$) group (normalized)

	FCI Pre	FCI Post	FCI Gain	MBT
Minimal level	−0.051	−0.247	−0.134	−0.351*
Low level	−0.350*	−0.139	0.165	−0.197
Medium level	0.409**	0.234	−0.151	0.366**
High level	−0.126	0.133	0.197	0.148
Average level	0.104	0.275	0.117	0.378**
Formula	−0.135	−0.201	−0.003	−0.222
Concept	−0.102	0.062	0.090	0.067
Coherence	0.276	0.250	0.162	0.231
Knowing	0.123	−0.122	−0.433**	0.102
Experiment	0.176	−0.006	−0.190	−0.008
Limitations	−0.160	0.058	−0.034	−0.055

Note. $N = 50$.
*$p < .05$, **$p < .01$.

medium-level questions and MBT score. Because the medium questions address procedural issues, it is unsurprising that a relation is found between these questions and the MBT, which is more of a problem-solving test than the FCI. The only relationship between any test score and a question topic is the fairly strong negative correlation between knowing questions and the FCI gain. One potential explanation for this is that these knowing questions, while indicating higher-level thinking, did not help students gain the conceptual knowledge needed to score well on the FCI. In addition, students in this group who ask more coherence questions also tend to ask more questions about concepts ($r = 0.552$, $p < .0001$) and justification for knowing something ($r = 0.416$, $p < .003$).

Regression Analysis

Another aspect investigated was the predictive power of student questioning. In other words, is it possible to predict anything about a student's conceptual performance at the end of the course purely based on what kinds of questions she or he asks? To determine the answer to this question, we performed several regression analyses for each of the two conceptual posttests. All regressions were carried out using a stepwise process.

The first regressions performed focused on how the difficulty level of questions asked by students influenced their MBT and post-FCI scores. Here the raw numbers of questions asked in each difficulty level were the independent variables and the test scores were the dependent variables. There was no significant regression relation between difficulty level and the post-FCI score. However, a relation was obtained between the percentage score on the MBT and the difficulty levels indicating that between 7% and 8% of the variance in student MBT scores is explained just by the distribution of difficulty in the questions asked in the weekly reports (Table 7). We see that asking high-level questions is related positively to the MBT score, whereas asking minimal questions is related negatively.

Another portion of the regression analysis explored how both the difficulty level and the topics of questions were related to the performance on the conceptual posttests. To do this, the six topic areas for the questions were divided into two groups each, based on the difficulty level of the questions. Minimal and low difficulty questions were combined to form a low group, and medium to high difficulty questions were combined to make a high group. This regrouping was done to separate questions related to clarification of concepts from the questions related to deeper

Table 7
Stepwise regression analysis for prediction of scores on end-of-quarter conceptual tests based on question difficulty

Exam	Significant Predictor Variable	β
MBT	High-level questions	0.201
	Minimal-level questions	−0.167
Post-FCI	No variables entered equation	

Note. Regression equation: MBT = 0.737 + . 0145 (High_level)− .011 (Minimal_level). $R = 0.276$, $R^2 = 0.076$.

understanding or extension of concepts. For example, ''What is the equation for *t* when a ball is fired straight up?'' would be counted as a low-level equation question, whereas ''Why can't we use the gravitational potential energy formula when the earth isn't in the system?'' would be a high-level equation question. Raw numbers of questions in each grouping were the independent variables for these analyses, and the posttest scores on the FCI and MBT were dependent variables (Table 8). We see that 10% of the variation in a student's post-FCI score can be explained merely by what sorts of questions he or she asks. In particular, asking medium- and high-level coherence questions relates to better conceptual understanding, and asking medium- or high-level equation-based questions has a negative contribution in developing conceptual understanding. On the MBT, 3% of the score variance can be explained by the questions asked in the weekly reports, still perhaps a stronger contribution than one might anticipate from something as simple as student questions. It might have been expected that this contribution would be smaller than that on the FCI, simply because the MBT involves some mathematical problem solving, in addition to the conceptual base tested by the FCI. Once again, the questions that matter are high-level coherence questions, indicating that efforts to piece knowledge together in a coherent fashion are related to doing well on the MBT.

Instructional Implications and Conclusions

Several broad instructional implications result from these findings. First, incorporating weekly reports into a course encourages students to ask more questions than they would otherwise. In a traditional classroom setting, Graesser and Person found that the average number of student questions is <3/hour for the entire class (Graesser & Person, 1994). The average questioning rate from weekly reports alone is much higher; in this study the weekly report questions are at a rate >12/contact hour. This probably is due to the more private nature of the communication between student and instructor.

However, simply encouraging students to ask questions on a regular basis does not result in learning. Students whose questions reveal a formulaic focus do not achieve increased conceptual

Table 8
Stepwise regression analysis for prediction of scores on end-of-quarter conceptual tests based on question difficulty and topic

Exam	Significant Predictor Variable	β
MBT	High-level coherence questions	0.169
Post-FCI	High-level coherence questions	0.258
	High-level formula questions	−0.209

Note. Regression equation: MBT = 0.718 + .0199 (Coherence_high). $R = 0.215$, $R^2 = 0.029$. Regression equation: Post_FCI = 0.781 + .0333 (Coherence_high)−.02,52 (Formula_high). $R = 0.323$, $R^2 = 0.104$.

understanding, indicating that this approach probably should be discouraged. Perhaps encouraging other types of questions will lead to a smaller number of these equation-based ones. Questions involving coherence and limitations are related to greater conceptual understanding for some populations of students. In addition, encouraging questions of these types may aid students in learning to phrase higher-level questions. These higher-level questions are related to deeper understanding of the subject matter.

Furthermore, the results depicted in Tables 4–6 show that although there are several significant correlations between the MBT and question difficulty, the question difficulty has no relation to the post-FCI score or FCI gain. This implies that the questioning behavior of students is not related to their acquisition of conceptual understanding only, but rather the total package of successfully applying concepts to solve simple problems, such as those on the MBT. Whether discouraging the minimal-level questions (which are correlated negatively with MBT scores) and/or encouraging the higher-level questions would relate to better scores on more difficult problem-solving tasks or not is uncertain. Certainly these results suggest this as an area for further investigation.

Because the study indicates the importance of students' questions for learning, the issue of convincing faculty of the importance of attention to student questions arises. The following arguments may be helpful in this effort. Weekly reports are an example of a formative assessment task. Black and Wiliam showed that the learning gains from systematic attention to formative assessment, if followed by feedback for the students, are larger than most of those gains found for other educational interventions (Black & Wiliam, 1998). An advantage of the questions that students ask in the weekly reports is that they allow double feedback; students provide feedback to the professor and the professor provides feedback to the students (Etkina, 2000; Etkina & Harper, 2002). This benefit can be achieved even without using the whole system of weekly reports in a big lecture course; students can be asked to submit their questions as a part of regular homework assignments so that a portion of the homework grade depends on whether a question is asked. We strongly believe that the questioning must be tied at least loosely to some sort of grade. Colleagues who have tried to make the questioning more optional (through a website or anonymous scraps of paper turned in at the end of class) report they receive few if any student questions. Because students' questions reveal problems otherwise invisible before exams, making them a part of the routine coursework might improve students' performances on these and other summative assessment activities.

How could students be aided in asking better questions? The class studied for this project received feedback only through the comments of graders who read their reports. The instructors did not devote class time to discussing desirable questioning behavior. Among strategies suggested by previous research is providing students with desirable stems to use in formulating their questions (e.g., "Why is it that _____?" or "How is _____ related to _____") (King, 1991). Another possibility is showing students a question taxonomy such as Marbach-Ad and Sokolove's (2000) and discussing it briefly in class. Integrating one or some combination of these techniques might result in more higher-level questions in the more desirable topic areas. From what has been described above, this should be related to stronger student conceptual understanding. Results of previous studies further indicate that these students ought to solve problems better (Dori & Herscovitz, 1999; King, 1991), retain more of the content material (Davey & McBride, 1986; King, 1989; Marbach-Ad & Sokolove, 2000), and become more independent learners (Marbach-Ad & Sokolove, 2000).

In addition, an alert instructor can harness student questions to motivate upcoming topics. For instance, right after the class had learned about work and energy, but had not begun impulse-momentum, a student asked, "Why is it that work is force time[s] displacement not time?" [sic].

This gave the instructor a great opportunity to involve the class in constructing the idea of impulse, fitting well with Macmillan and Garrison's "epistemological ought" (Macmillan & Garrison, 1983). The question "Is there such [a] thing as circular momentum?" could be used similarly to introduce angular momentum.

Finally, we find that it is possible to make some prediction about student achievement on conceptual tests purely based on the kinds of questions they ask in the reports. In particular, encouragement of high-level questions about how the content knowledge of the course is structured is related to better conceptual understanding. As high-level equation-related questions are related negatively to conceptual performance, this is another reason why an equation-centered approach to learning physics should be discouraged. In summary, using weekly reports to elicit and respond to student questions can assist instructors in identifying questioning behaviors that are related to both high and low conceptual achievement.

This work was supported partially by NSF Grant DUE-0088906. The authors are grateful to David Mills for help in providing feedback to the students and grading. They thank Cindy Hmelo-Silver and Xueli Zou for feedback and help in the preparation of the manuscript. They thank Alan Van Heuvelen, Ted Pavlic, Lei Bao, and the members of OSU Physics Education Research Group for input throughout the process. Finally, they acknowledge the support of Rick Freuler, John Demel, and the OSU FEH staff.

Notes

[1]This cycle is similar to the cycle of conjecture, evaluation, and modification or rejection in hypothesis development and model construction proposed by J. Clement ["Learning via model construction and criticism," in G. Glover, R. Ronning, & C. Reynolds (Eds.), Handbook of creativity: Assessment theory and research, 1989, p. 347].

[2]The comparison number was calculated from information in Hake, R.R. ("Interactive-engagement methods in introductory mechanics courses," found online at http://www.physics.indiana.edu/~sdi).

References

Arons, A.B. (1997). Teaching introductory physics. New York: Wiley.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. Assessment in Education, 5, 7–74.

Bloom, B. (Ed.). (1956). Taxonomy of educational objectives: The classification of educational goals: Handbook I, cognitive domain. New York: David McKay.

Brown, D.E. (1989). Students' concept of force: The importance of understanding Newton's third law. Physics Education, 24, 353–358.

Clement, J. (1982). Students' preconceptions in introductory mechanics. American Journal of Physics, 50, 66–71.

Davey, B. & McBride, S. (1986). Effects of question-generation training on reading comprehension. Journal of Educational Psychology, 78, 256–262.

Dillon, J.T. (1990). The practice of questioning. New York: Routledge, Chapman, Hall.

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.

Etkina, E. (2000). Weekly reports: A two-way feedback tool. Science Education, 84, 594–605.

Etkina, E. & Harper, K.A. (2002). Closing the feedback loop in large enrollment physics courses. Journal of College Science Teaching, 31, 476–480.

Gamble, R. (1989). Force. Physics Education, 24, 79–82.

Good, T.L., Slavings, R.L., Harel, K.H., & Emerson, H. (1987). Student passivity: A study of question asking in K–12 classrooms. Sociology of Education, 60, 181–199.

Graesser, A.C., Person, N.K., & Huber, J. (1992). Mechanisms that generate questions. In Lauer, T.W., Peacock, E., & Graesser, A.C. (Eds.), Questions and information systems (pp. 167–187). Hillsdale, NJ: Erlbaum.

Graesser, A.C., McMahen, C.L., & Johnson, B.K. (1994). Question asking and answering. In Gernsbacher, M.A. (Ed.), Handbook of psycholinguistics (pp. 517–538). San Diego: Academic Press.

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

Hake, R.R. (1998). Interactive engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. American Journal of Physics, 66, 64–74.

Halloun, I.A. & Hestenes, D.O. (1985). The initial knowledge state of college physics. American Journal of Physics, 53, 1043–1055.

Halloun, I.A. & Hestenes, D.O. (1985). Common sense concepts about motion. American Journal of Physics, 53, 1056–1065.

Hammer, D. (1989). Two approaches to learning physics. Physics Teacher, 27, 664–670.

Hammer, D. (1994). Epistemological beliefs in introductory physics. Cognition and Instruction, 12, 151–183.

Heady, J.E. (1993). Teaching embryology without lectures and without traditional laboratories—an adventure in innovation. Journal of College Science Teaching, 23, 87–91.

Heller, P. & Hollabaugh, M. (1992). Teaching problem solving through cooperative grouping. Part 2: Designing problems ad structuring groups. American Journal of Physics, 60, 637–644.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. Physics Teacher, 30, 159–166.

Hestenes, D. & Wells, M. (1992). A mechanics baseline test. Physics Teacher, 30, 141–158.

Johnson, D.W., Johnson, R.T., & Smith, K.A. (1991). Active learning: Cooperation in the college classroom. Edina, MN: Interaction Book Company.

Kass, H. & Lambert, P. (1983). Student preconceptions in introductory high school physics as related to course achievement. In Helm, H. (Ed.), Proceedings of the International Seminar on Misconceptions and Educational Strategies in Science and Mathematics (pp. 328–333). Ithaca, NY: Cornell University Press.

King, A. (1989). Effects of self-questioning training on college students' comprehension of lectures. Contemporary Educational Psychology, 14, 366–381.

King, A. (1991). Effects of training in strategic questioning on children's problem-solving performance. Journal of Educational Psychology 83, 307–317.

LaFrance, M. (1992). Questioning knowledge acquisition. In Lauer, T.W., Peacock, E., & Graesser, A.C. (Eds.), Questions and information systems (pp. 11–28). Hillsdale, NJ: Erlbaum.

Macmillan, C.J.B. & Garrison, J.W. (1983). An erotetic concept of teaching. Educational Theory, 33, 157–166.

Marbach-Ad, G. & Sokolove, P.G. (2000). Can undergraduate biology students learn to ask higher level questions? Journal of Research in Science Teaching, 37, 854–870.

McCloskey, M. (1983). Naive theories of motion. In Gentner, D. & Stevens, A.L. (Eds.), Mental models (pp. 299–324). Hillsdale, NJ: Erlbaum.

Minstrell, J. & Stimpson, V. (1986). Instruction for understanding: A cognitive process framework (Report No. NIE-G-83-0059). Washington, DC: National Institute of Education. (ERIC Document Reproduction Service No. ED282749).

Miyake, N. & Norman, D.A. (1979). To ask a question, one must know enough to know what is not known. Journal of Verbal Learning and Verbal Behavior, 18, 357–364.

Orr, H.A. (1999). An evolutionary dead end? Science, 285, 343–344.

Redish, E.F., Saul, J.M., & Steinberg, R.N. (1998). Student expectations in introductory physics. American Journal of Physics, 66, 212–224.

Trowbridge, D.E. & McDermott, L.C. (1981). Investigation of students' understanding of the concept of acceleration in one dimension. American Journal of Physics, 49, 242–253.

Van der Meij, H. (1994). Student questions: A componential analysis. Learning and Individual Differences, 6, 137–161.

Van Heuvelen, A. (1991). Learning to think like a physicist: A review of research-based instructional strategies. American Journal of Physics, 59, 891–897.

Van Heuvelen, A. & Maloney, D. (1999). Playing physics jeopardy. American Journal of Physics, 67, 252–256.

Van Heuvelen, A. & Zou, X. (2001). Multiple representations of work-energy processes. American Journal of Physics, 69, 184–194.

Zoller, U., Tsaparlis, G., Fatsow, M., & Lubezky, A. (1997). Student self-assessment of higher-order cognitive skills in college science teaching. Journal of College Science Teaching, 27, 99–101.