ASSESSMENT AND ACTIVE LEARNING STRATEGIES FOR INTRODUCTORY GEOLOGY COURSES

David A. McConnell Department of Geology, University of Akron, Akron, OH 44325-4101, dam6@uakron.edu
David N. Steer Department of Geology, University of Akron, Akron, OH 44325-4101, steer@uakron.edu
Kathie D. Owens Department of Curricular and Instructional Studies, College of Education, University of Akron, Akron, OH 44325-4205, kowens@uakron.edu

ABSTRACT

Educational research findings suggest that instructors can foster the growth of thinking skills and promote science literacy by incorporating active learning strategies into the classroom. Active learning occurs when instructors build learner participation into classes. Learning in large, general education Earth Science classes was evaluated using formative assessment exercises conducted by students in groups. Bloom's taxonomy of cognitive development was used as a guide to identify critical thinking skills (comprehension, application, analysis, synthesis, evaluation) that could be linked to specific assessment methods such as conceptests, Venn diagrams, image analysis, concept maps, open-ended questions, and evaluation rubrics. Two instructors conducted a series of analyses on sample classes taught with traditional lecture and inquiry-based learning methods. Qualitative and quantitative analyses show that such methods are preferred by students, improve student retention, produce no decrease in content knowledge, promote deeper understanding of course material, and increase logical thinking skills.

Keywords: active learning, inquiry-based learning, assessment, Bloom's taxonomy

INTRODUCTION

Several studies have emphasized the need to improve science literacy among non-science majors (American Geophysical Union, 1994; National Science Foundation, 1996; National Research Council, 1997) and college instructors have consistently ranked student intellectual development as a primary teaching goal (Angelo and Cross, 1993; Trice and Dey, 1997; Figure 1). Teachers can meet these complementary goals by, focusing on remedies that make content relevant to the intended audience, increasing student-student interaction in class, and encouraging conceptual understanding rather than rote memorization of facts (Chickering and Gamson, 1987; Tobias, 1990, 1992; Angelo, 1993; Astin, 1993). Such objectives can be realized by the combination of two teaching strategies, active learning and inquiry-based learning (Siebert and McIntosh, 2001). Active learning occurs when instructors build learner participation directly into classes using exercises that ask students to apply newly acquired knowledge to solve problems that may range from a single multiple-choice question to a class-length project (Silberman, 1996). Inquiry-based learning introduces elements of scientific inquiry into active learning exercises. Teaching strategies that promote inquiry-based learning (Allard and Barman, 1994; Mazur, 1997) emphasize higher-level thinking processes such as making observations, posing questions, analyzing data, making predictions, and communicating ideas (Brunkhorst, 1996; National Research Council, 2000).

This paper describes a variety of learning strategies that may be adopted in introductory geology courses to encourage the development of higher-order thinking skills. We assume the reader has no prior experience in active learning methods and provide directions for implementing these techniques in the classroom. We discuss six hierarchical levels of student learning and link them to examples of appropriate assessment tools that were used successfully in several sections of a general education Earth Science course taught by two instructors at the University of Akron. These teaching strategies have been evaluated qualitatively using peer reviews, student written evaluations and semi-structured student interviews; and quantitatively by measuring improvements in student retention, exam scores, and scores on a logical thinking assessment instrument.

TEACHING, LEARNING AND ASSESSMENT

Teaching faculty consistently rank the development of higher-order thinking skills ahead of other teaching goals (Angelo and Cross, 1993). Unfortunately, large numbers of students in introductory courses frequently find themselves in an educational setting where learning is reduced to low level intellectual skills of listening and recording information that will be memorized for a multiple choice exam (Pinet, 1995; Prothero, 2000; McManus, 2002). Students familiar with high school experiential learning strategies allied with the national science standards will be unaccustomed to lecture delivery, especially in large-class settings (Collins, 1997). Content-driven coursework that can be efficiently graded by multiple-choice tests has proven ineffective in promoting deep student understanding of basic science concepts (Tobias, 1990). Furthermore, it can have a negative impact on student attitudes about science, even among majors (Allard and Barman, 1994; Gibbons, 1994; Sundberg et al., 1994; De Caprariis, 1997). As a result, such courses are usually poor recruiting and retention tools. In many institutions, pre-service teachers make up a significant proportion of introductory science courses. Teachers in K-12 schools not only learn what they will teach in these classes, but are also exposed to teaching models by their instructors (Collins, 1997). Finally, general education science courses represent an important opportunity for students to develop the critical thinking skills that are essential for success in college.

In recent years, college science instructors have attempted to encourage in-class learning by utilizing teaching methods that promote collaborative, active learning during lecture periods (Macdonald and Korinek, 1995; Ebert-May et al., 1997; Mazur, 1997; Reynolds and Peacock, 1998; Murck, 1999; Crouch and Mazur, 2001; Wyckoff, 2001). Student interaction through collaborative learning is a key determinant of student performance (Bykerk-Kauffman, 1995; Lord,

		Learning Tool (Assessment Method)					
Bloom's Taxonomy	Learning Skill	Concep-test	Venn Diagram	Image Analysis	Concept Map	Open-ended Question	Evaluation Rubric
Knowledge	memorization and recall	•	•	•	•	•	•
Compre-hension	understanding	•	•	•	•	•	•
Application	using knowledge	•	•	•	•	•	•
Analysis	taking apart information		•	•	•	•	•
Synthesis	reorganizing information				•	•	•
Evaluation	making judgements					•	•

Table 1. Formative assessment methods and Bloom's taxonomy.

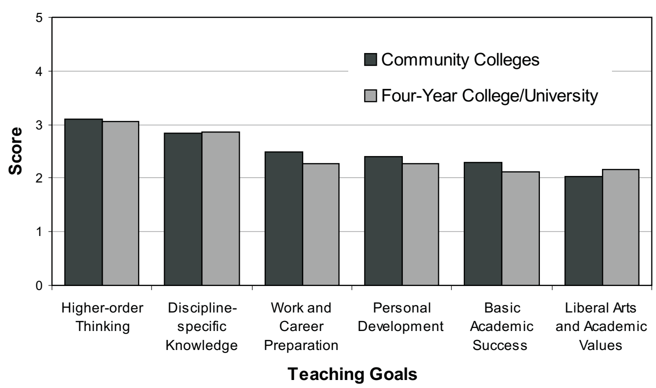

Figure 1. Relative scores on the Teaching Goals Inventory (Angelo and Cross, 1993). Instructors at both community colleges and four-year institutions ranked the development of higher-order thinking skills ahead of other teaching goals. n = 1873 for community colleges; n = 951 for four-year colleges and universities.

2001). The benefits of active learning and inquiry-based teaching methods can be seen in improvements in student attitudes about science (Gibbons, 1994; Ebert-May et al., 1997; Reynolds and Peacock, 1998) and increases in standardized test scores (Mazur, 1997; Hake, 1998).

Instructors typically assess learning by having students complete an exam following several weeks of lectures (McManus, 2002). This is a form of summative evaluation that comes at the end of a course of study, often too late to correct mistakes or identify gaps in comprehension. In contrast, formative assessment methods can be used to identify learning problems during the presentation of information while there is an opportunity to recognize and correct misconceptions. The use of formative assessment can transform a traditional passive lecture into an active learning experience, as it requires that students provide feedback on their ongoing learning, thus giving the instructor an opportunity to highlight concepts that require additional explanation. For formative assessment to be effective, we must find questions to ask that will engage students and provide answers that can be used to signal understanding or confusion. Furthermore, we can link these assessment tools to different learning skills to nurture cognitive development.

A FRAMEWORK FOR LEARNING

Over forty years ago, Benjamin Bloom and several co-workers created a taxonomy of educational objectives that continues to provide a useful structure for organizing learning exercises and assessment experiences at all levels of education (Bloom et al., 1956; Anderson and Sosniak, 1994; Anderson and Krathwohl, 2001). Bloom's taxonomy divided cognitive learning into six levels (Table 1), from lower-level thinking skills such as memorization to higher order thinking that involves the evaluation of information. The taxonomy has been used by instructors in geology courses to guide the development of questions that address a full range of cognitive skills (Fuhrman, 1996; Nuhfer, 1996). Each taxonomy level is described briefly below and examples of specific questions linked to each level are presented.

Knowledge - Answers to knowledge questions indicate if a student knows and can recall specific information. Examples of questions that assess knowledge are some types of multiple choice questions, true/false questions, definitions, matching questions, or lists. Questions that ask students to define, identify, list, or name are often

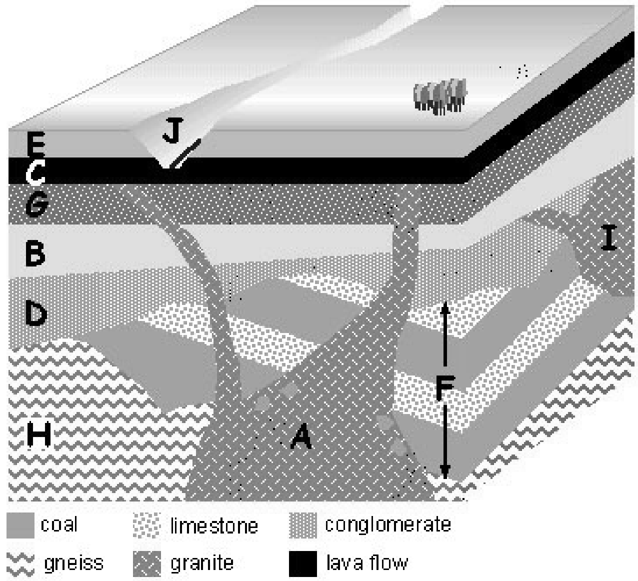

Figure 2. An application question would ask students to use the laws of superposition, cross cutting relationships, and original horizontality to determine the order of events for the labeled features in the idealized figure above.

“knowledge” questions. The following are knowledge questions.

K1. Which of the following is an igneous rock?

a) limestone b) granite c) slate d) coal

K2. List the names of three major plates.

Multiple choice, true/false, and matching questions require that students recognize information stored in memory. Listing or fill-in-the-blank questions require that a student is able to remember specific information, as the questions themselves do not provide the answer choice.

Comprehension - Responses to comprehension questions report information or observations. Students must possess some basic knowledge of the subject to correctly answer these questions. Comprehension questions can fall into several categories and require students to convert, summarize, classify, infer, compare, or explain information. Examples of questions might include the following:

C1. Draw a diagram that shows the relationships between the principal components of the Earth system.

C2. View the motion picture Dante’s Peak and summarize the principal geological concepts presented in the movie.

C3. Take four pictures or samples of igneous rocks and sort them into volcanic and plutonic rock types.

C4. Fill in the blank to complete the analogy.

The yolk is to the egg as the ____________ is to Earth.

C5. Contrast the floor of the Atlantic Ocean with the shape of a bathtub.

C6. Predict what would happen to sea level if it were to rain continuously worldwide.

Application - Application involves applying rules or principles to new situations, using known procedures to solve problems, or demonstrating how to do something. Questions that ask students to solve a problem using a known equation or to select a procedure to complete a new task would be considered application questions. An examples of an application question follows:

Ap1. Use the principles of superposition, cross cutting relationships, and original horizontality to determine the order of formation of labeled features in Figure 2.

Analysis - Answers to analysis questions may give directions, make commentaries, scrutinize data, explain how something works, or distinguish fact from opinion. Analysis requires that students break information into component parts to identify its organization. Students are expected to find links between data and interpretations and to discover which material is relevant to a task and which is extraneous. Questions that ask students to diagram, illustrate, outline or subdivide would be considered analysis questions, for example:

An1. Identify the hypothesis, observations, and conclusions in an assigned research report.

An2. Read a newspaper editorial and determine if it was written from a pro-environment or pro-development perspective.

Synthesis - Synthesis combines a series of parts into a greater whole. Good answers to synthesis questions may predict the outcome for a particular event and may involve making generalizations and developing a “big picture” view of a phenomenon or feature. Questions may ask students to create multiple hypotheses to explain a phenomenon, to develop a plan to solve a problem or to devise a procedure to accomplish a task. Examples of synthesis questions might include the following:

S1. How would you change building codes or zoning regulations in regions of volcanic activity to protect people and property?

S2. Plan an experiment to test if a landfill is polluting water from a nearby well.

Evaluation - Responses to evaluation questions use evidence and scientific reasoning to make judgments about facts, data, opinions or research results. Good answers require students to analyze and synthesize information and clarify ideas. Evaluation questions might ask a student to appraise, criticize, justify, or support an idea or concept. Examples of potential evaluation questions could include:

E1. Where is the greatest danger from an eruption of Mt. Shasta? Explain why.

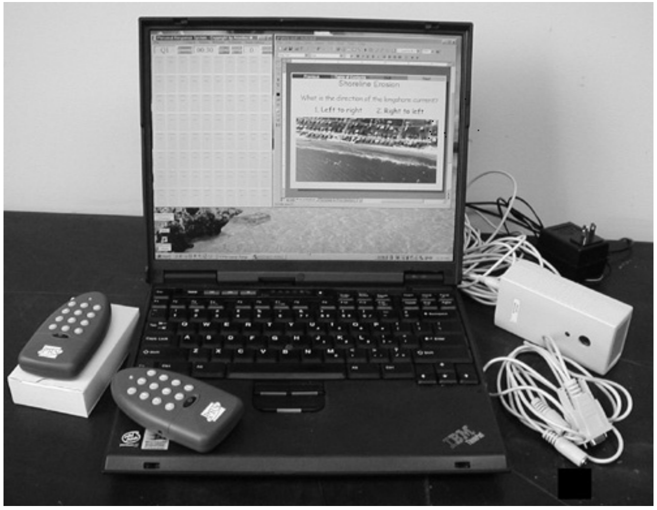

Figure 3. Students' responses to conceptest questions are collated and tabulated using a classroom communication system. The system consists of transmitters (front left) that send signals to one or more receivers (center right) linked to a computer projection system.

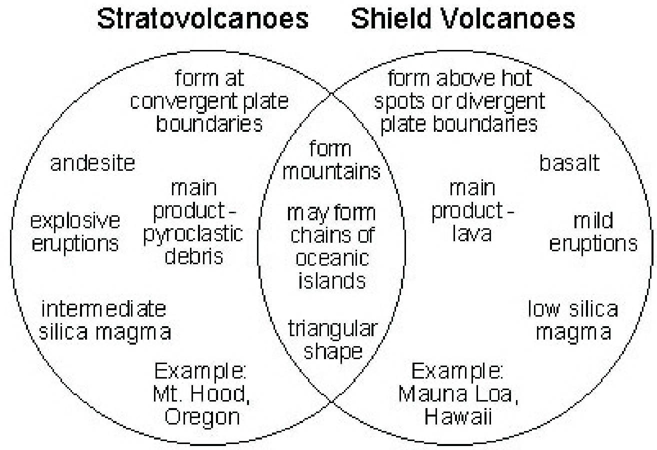

Figure 4. Venn diagrams can be used to compare and contrast the characteristics of related features. Students may be asked to complete a diagram in preparation for class, following a short reading assignment in class, or using lecture notes. Alternatively, students may be asked to locate a list of features in the correct place on the diagram.

E2. What is the most cost-efficient way to protect residents in a drainage basin from future flooding?

ASSESSMENT OF LEARNING

This section describes six methods of formative assessment aimed at recognizing and correcting misconceptions during lecture. Such learning tools can be assigned as in-class exercises or used by students outside of class in preparation for exams (Nuhfer, 1996). The assessment methods described below are keyed to Bloom's taxonomy in Table 1.

Conceot Tests - Conceptests were developed as part of the peer instruction technique used to teach physics (Mazur, 1997). This teaching method has been widely used in physics courses at a range of institutions (Hake, 1998) and has been successfully adopted by faculty in a variety of other disciplines (e.g., chemistry, biology, astronomy; Crouch and Mazur, 2001). Peer instruction divides class time between short lectures and conceptual multiple- choice questions. Conceptest questions are designed to evaluate student understanding of the basic concepts behind the lecture material. Conceptests are not simple content-based multiple-choice (conceptest) questions that rely on the student re-reading their lecture notes or memorizing a fact or definition. Instead, these questions are designed to assess student understanding of the principal concepts underlying the lecture material. Conceptests generally correspond to the comprehension level of Bloom's taxonomy but may also be suitable for application questions (Table 1).

Students were given 30-60 seconds to consider a conceptest question and to chose an answer. We had previously provided large lettered answer cards that they would use to indicate their selection. This technique has been replaced by the use of an electronic personal response system (Figure 3) that registers student answers and generates a histogram of responses. This method has the advantage of providing students with a visual display of answers for the class while keeping individual answers anonymous. Furthermore, the technology provides students with the option of declaring their level of confidence in their answer choice. Following their initial answer, students are given 1-2 minutes to discuss the reasons for their choice with their neighbors (peer instruction) in pairs or small groups before voting again. This process usually results in an increase in the number of correct answers and an increase in student confidence in their answer choices (Mazur, 1997). Finally, a group spokesperson may be given an opportunity to provide a brief explanation of the group's answer and/or the instructor may clarify or expand on the correct response.

Venn Diagrams - Venn diagrams are a graphical method for comparing and contrasting features or phenomena. Such diagrams represent an opportunity for students to identify the characteristics of classification systems or to analyze the key components of complex sets of geological features. For example, students may be provided descriptions of the geological characteristics of two volcanoes and asked to compare and contrast their features using a Venn diagram (Figure 4). The use of Venn diagrams may involve knowledge, comprehension, application, and analysis levels of Bloom's taxonomy (Table 1). Examples of other possible comparisons are, igneous vs. sedimentary vs. metamorphic rocks, hurricanes vs. tornadoes, and divergent vs. convergent plate boundaries. Instructors may choose to provide a numbered list of characteristics that could then be placed in the correct locations on a labeled diagram. This assistance reduces the analysis aspect of the exercise, as students would not be identifying key components themselves. Such an

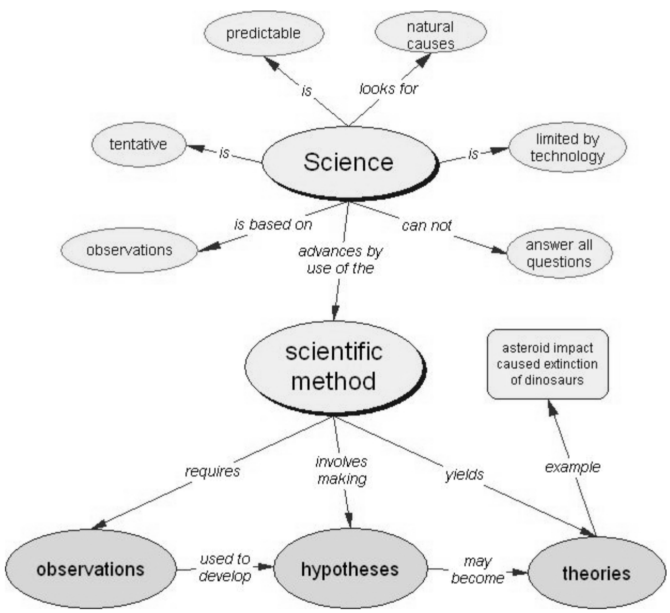

Figure 5. A simple concept map that illustrates the relationship between the elements of the scientific method. One potential scoring scheme would award 5 points per hierarchical level (5 levels present); 1 point for each reasonable linking phrase between adjacent points (12 links). Using this scheme the concept map would earn 37 points.

exercise would then be reclassified into the comprehension category.

IMAGE ANALYSIS

Image analysis is a form of slide observation (Reynolds and Peacock, 1998) where students are shown a photograph, map, or diagram and asked to make observations and interpretations. These types of exercises are an excellent way to begin a class as they immediately engage the student in the topic at hand. Image analysis involves knowledge, comprehension, application, and analysis levels of Bloom's taxonomy. Under certain circumstances exercises may also require students to synthesize and evaluate information. Readers are referred to Reynolds and Peacock (1998) for a thorough discussion of this technique.

CONCEPT MAPS

A concept map is a graphical representation of a student's knowledge about a topic (Zeilik et al., 1997). Concept maps are pictorial essays, a method of illustrating the principal concepts of a lesson, and include supporting information that indicates how a student has organized his/her ideas. Concept maps present a "big picture" view of a student's understanding of a topic. Good concept maps force their creators to challenge their own understanding and to build a strong foundation for information that follows. A poorly constructed map allows a reviewer to quickly identify gaps in logic or comprehension. Concept maps will vary from person to person, no two are alike. They allow for creative thinking in their construction.

Concept maps have two principal components: 1. Terms or concepts - often presented in boxes; 2. Directional links (arrows) and linking phrases (prepositions) - that connect the terms (Figure 5). Concept maps identify the relationships between components and therefore correspond to synthesis in Bloom's taxonomy (Table 1). The number of levels in a concept map can be readily counted. The terms are joined by logical linking phrases appropriate for the topic. The maps can be readily evaluated as good, average, or poor to speed assessment. Alternatively, one can construct a formal scoring scheme (see caption, Figure 5).

Open-Ended Questions - Open-ended or divergent questions do not necessarily have a specific correct answer. Such questions can be written by the instructor to involve almost all levels of Bloom's taxonomy (Freedman, 1994). The creation of questions can serve as a method for promoting critical thinking among students. King (1995) used a series of generic question stems (Table 2) to prompt students to generate questions related to lecture and reading assignments. The question stems can be matched to specific levels of Bloom's taxonomy (Table 2). Student-generated questions could be used for self-examination, to assess comprehension of reading assignments, or in peer questioning exercises (King, 1995).

A form of open-ended question known as a minute paper is one of the most commonly utilized assessment methods in large classes. A minute paper is a short informal writing assignment that requires little time to complete and can be assessed easily (Angelo and Cross, 1993; Macdonald and Korinek, 1995; Murck, 1999). Students may be given literally one minute or a few minutes longer to complete the writing assignments. Minute papers can be used to determine whether students have grasped the key idea(s) presented during lecture. The papers may focus specifically on an important concept that students should understand but more commonly are the students' responses to the general question "What is the most important thing we discussed today?" This question challenges students to evaluate the lecture material and identifies whether they can discriminate between critical and routine information. Another technique known as a Muddiest Point exercise may start with the question, "What was the most confusing idea (muddiest point) presented in today's lecture?" (Angelo and Cross, 1993). Rather than asking students what they know, the focus may instead be on concepts they don't understand.

Assessment of answers following an in-class minute-paper or muddiest point exercise will indicate if student perceptions of lecture material corresponded to the instructor's lecture goals. Common misconceptions or gaps in comprehension should be addressed at the start of the next class period. Prompt feedback is a hallmark of good teaching (Chickering and Gamson, 1987; Angelo, 1993).

Evaluation Rubrics - Rubrics are used widely within society. When you complete a questionnaire that asks you to judge the quality of service in a restaurant you are

Bloom's Taxonomy	Question Stems
Knowledge	What is...?
Comprehension	What would happen if...? What does...illustrate about...? What is analogous to...?
Application	How could...be used to...? What is another example of...?
Analysis	How does...affect...? What are the differences (similarities) between...? What causes...?
Synthesis	What is a possible solution for the problem of...? How does...relate to what we learned before about...?
Evalution	Why is...important? What is the best...and why? Do you agree/disagree that...?

Table 2. Critical thinking question stems, modified from King (1995).

Factors	Low Risk (1 pt)	Intermediate Risk (2pts)	High Risk (3 pts)
Proximity to fault	far (>200 km)	moderate (50-200 km)	close (<50 km)

Table 3. Template for earthquake risk evaluation rubric.

Factors	Low Risk (1 pt)	Intermediate Risk (2pts)	High Risk (3 pts)
Proximity to fault	far (>200 km)	moderate (50-200 km)	close (<50 km)
Time since last major earthquake	years	decades	centuries
Earthquake magnitude	small (<magnitude 4)	moderate (magnitude 4-5)	high (>magnitude 6)
Substrate	bedrock	rock and sediment mix	sediment
Utilization of building codes	all buildings built to code or retrofitted	building codes only partially enforced	no building codes

Table 4. Completed earthquake risk evaluation rubric. The values (distances, elevations, slopes) in the above example are arbitrary and are only intended to give an example of how quantitative data may be incorporated into a rubric.

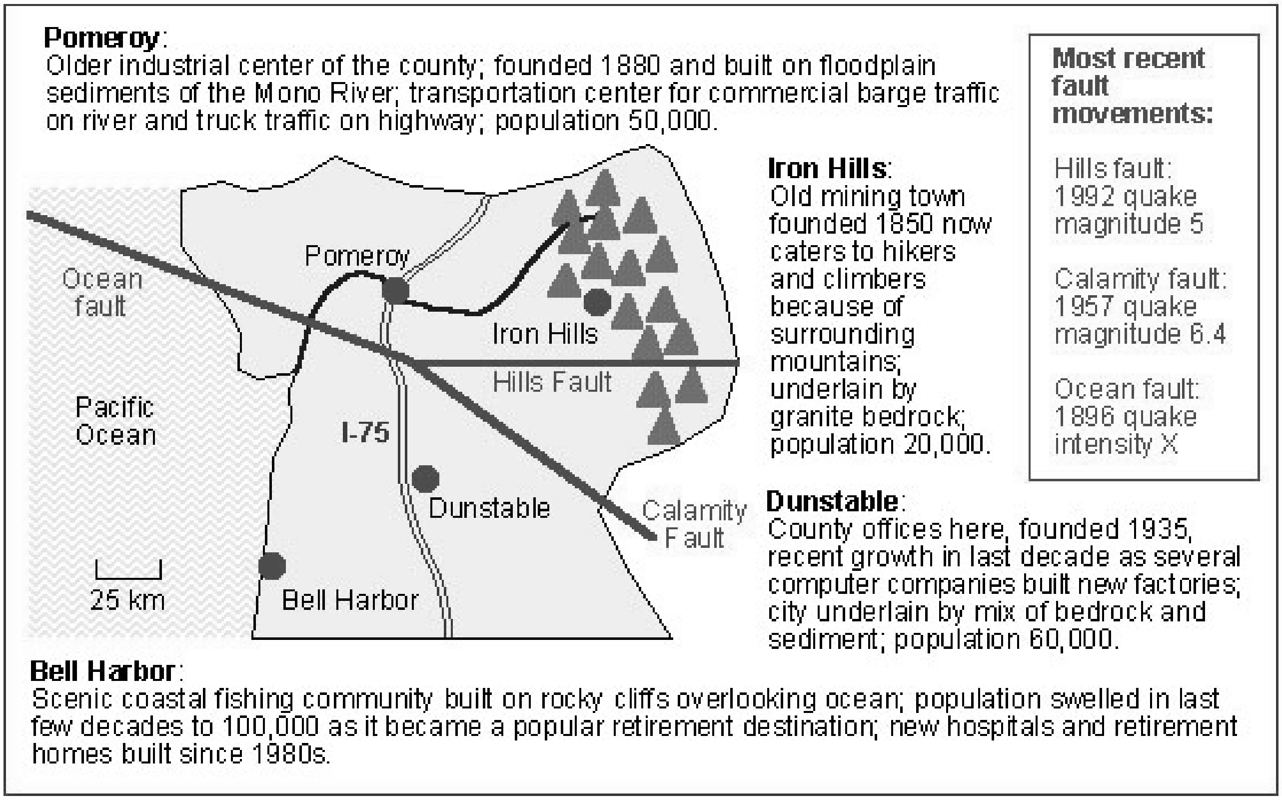

Figure 6. This idealized map of a county in California was used in the earthquake risk scoring rubric exercise. Students were asked to rank the four cities on this map in order of degree of risk of damage following a future earthquake.

using a rubric. When students judge the quality of teaching in a college class they often use a type of rubric. The relative scores on individual questions can be used to identify potential areas for improvement.

Scoring rubrics have traditionally been used by educators as assessment tools for student writing exercises. Rubrics provide a scoring scheme that can be keyed to specific performance goals and are especially useful for tasks where scoring could be subjective. The instructor compares each assignment with the standard of the rubric, ensuring a consistent scoring method. Rather than having students use an existing scoring rubric, we asked them to create their own rubrics for the purpose of evaluating specific geological situations. We termed these learning tools evaluation rubrics to differentiate them from the typical scoring rubric. Evaluation rubrics can involve all levels of Bloom's taxonomy (Table 1).

In this assessment method, students are required to generate their own rubrics. For example, students might create a rubric to assess the risk of an earthquake affecting a city (Tables 3, 4). We used two types of rubrics: 1. Rubrics with three scoring levels (Table 4); or, 2. Rubrics that required respondents to rank factors in order of significance.

Evaluation rubrics begin with a description of a specific situation. For example, the following instructions were given to students during a discussion of earthquakes. Students were asked to read the instructions and complete a partially finished rubric (Table 3).

> Following graduation you get a job working for a county planning task force in California. The task force must examine the setting of several different cities and identify which is at greatest risk for future earthquake damages from movements on known faults. You are given the assignment to create an evaluation rubric to assess factors that will influence the risk of potential damage from a future earthquake. The city that scores the highest using the rubric will receive additional county funds to protect key structures from earthquake damage.

Rubrics may be presented with one factor already identified to illustrate the scoring scheme. The quality of the student responses can be determined by the factors that are identified and the discrimination of the scoring methods (Table 4). A good rubric will identify several relevant factors and describe what constitutes a high or low score for each factor.

Students may be asked to distinguish which factor is the most important under the circumstances of the exercise. The score for this factor may be doubled. This requires making a judgment on the relative significance of the chosen factors. The final stage of a rubric exercise requires students to use their rubric in a hypothetical situation. For example, students who had completed the earthquake risk rubric (Table 4) could be given information on the geology and characteristics of four cities (Figure 6) and asked to rank the cities in order of greatest to least risk of damage from a future earthquake.

ASSESSMENT OF TEACHING

The assessment methods described above have been used in several sections of an Earth Science course taught by two instructors (McConnell, Steer) over the last two years. Course enrollment ranged from 140-180 students per section. The majority (60-70%) of students were freshmen. The classes were taught in large auditorium-style classrooms with fixed seats facing a projection screen. Both instructors projected lecture materials using presentation programs such as Powerpoint, and had access to on-line materials through classroom internet connections.

The University of Akron is a large (22,000 students), open enrollment, state institution in northeast Ohio. The majority of students commute to class from surrounding communities. Students in an equivalent introductory geology course report that they work an average of 25 hours per week outside the University. Approximately a third of incoming freshmen do not return for the subsequent fall semester (UA Factbook, 2001). The student populations in Earth Science exhibit a broad range of skill levels. Students entering the University in Fall 2000 had an average ACT score of 20 and an average high-school GPA of 2.76. Fifty-eight percent of incoming students completed the college preparatory curriculum, in comparison to an average of 71% for the thirteen principal universities in Ohio (UA Factbook, 2001).

Most of the class sections discussed herein were offered in 50-minute blocks taught three-days a week at consistent times. Students were organized into informal groups made up of nearby students or permanent formal groups assigned by the instructor in all sections that employed active learning. Ideal group size was four students but groups varied from two to five students depending on attendance. For the purposes of this paper we will divide the classes into two types:

> Traditional classes that followed a passive lecture format that did not involve groups and did not incorporate inquiry-based or active learning exercises during class;
>
> Inquiry-based learning (IBL) sections that involved students working in groups, and the incorporation of active learning methods during lectures.

The contrast between these learning environments compares with teaching-centered and learning-centered classroom models (McManus, 2002). Steer gradually increased the degree of IBL material in his courses from only traditional lecture in early sections to incorporating daily exercises in the most recent versions of the course. During this investigation McConnell taught two consecutive sections using traditional methods in the first lecture and IBL techniques in the later class. One instructor (McConnell) has taught the course for twelve years whereas the other (Steer) was in his third year of teaching. Material for both classes was divided into ten modules composed of three or four lectures each. The instructors shared many resources and identified common goals for the course, but did not necessarily use the same classroom materials or exercises.

Exams were either identical between sections or varied by a few questions due to slight differences in choice of material covered or pacing of the class sections. Exams were divided into three parts; knowledge questions, comprehension questions, and analysis questions. Equal numbers of knowledge and comprehension questions in a multiple-choice format were included on each exam and accounted for 80-90% of the total exam score. The remainder of the exam grade was from analysis questions that took several forms such as creating or completing concept maps, interpreting map data, open-ended questions, or drawing diagrams. Grading in the courses involved a combination of exam scores, homework assignments, and in-class exercises. Exams accounted for between 50-70% of the total course grade.

Peer Reviews of Teaching Methods - A colleague from the College of Education (Owens) visited the Earth Science classes on numerous occasions and collected field notes to record instructors' and students' behaviors. Table 5 compares and contrasts the classroom environments for the traditional and IBL sections of the course. In both the traditional and IBL classes students sat in groups of two to four people prior to the start of class. Students in the traditional class sat quietly or talked in hushed tones, in contrast, students in the inquiry class chatted among themselves and organized their group for the day's activities.

In the traditional class, the instructor did most of the talking about the topic of the day. Approximately 5% of the students participated when asked to give examples of phenomena or to respond to a question. Students paid attention but were passive receivers of the information. There was also some lecture in the IBL class, but it was often used to give instructions, to make transitions between one activity and the next, or to summarize the day's lesson. Students worked in groups to discuss and write responses to open-ended questions. During this time the noise level in the class increased dramatically, but a visitor listening to conversations would have discovered that students stayed on task. Conversations were peppered by technical terms but were characterized by less formal language and student idioms. After the group work ended, volunteer spokespersons were asked to report their groups' answers to the whole class. Frequently the students voted to decide a "best answer" as a way to come to an overall consensus. Sometimes the instructor presented a summary or asked the students to draw a concept map of the principal ideas of the lesson as a way to communicate their understanding.

The contrast in the methods was obvious to the College of Education observer. In the traditional class, only a handful of students participated in the discussion, whereas in the inquiry class, a substantial majority of students was actively engaged in class activities through group discussions. Essentially each class covered the same material, but the student involvement was much richer in the inquiry class.

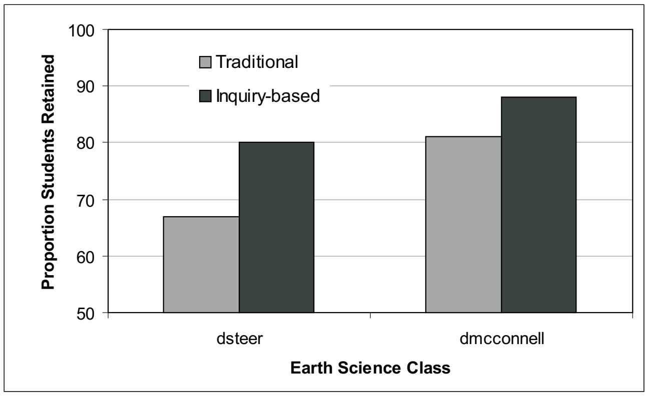

Figure 7. Student retention, the proportion of students present for the first exam who completed the last exam, was greater for the IBL classes than in traditional lecture classes. Gray bars for traditional lecture classes taught by Steer and McConnell black bars for inquiry-based learning classes.

Traditional	Inquiry-based Learning
Passive students	Active students
Quiet	Noisy
Instructor-focused	Student-focused
Information from instructor to student	Information from instructor to student and student to student
Students as individuals	Student collaboration
Competitive learning environment	Supportive learning environment
Limited assessment opportunities	Multiple assessment opportunities
Rigid setting (lack of mobility)	Mobile environment for instructor and student

Table 5. Traditional vs. IBL classroom characteristics.

STUDENT EVALUATIONS

The majority of student comments on the IBL teaching methods were positive. Written comments in student evaluations from McConnell's Fall 2000 IBL class were examined for references to in-class exercises. Forty-three students mentioned the exercises, 79% of the references were positive. Sample comments are included below.

> … truthfully, this is the only class I made myself go to because it helped me learn in a more relaxed, interesting way.
> The strongest part of the class is the group work. It helps you think about and understand the material.
> The groups were an excellent learning tool. Students teaching students is the best way for them to learn!

Even the seemingly negative comments can sometimes be interpreted with a positive spin:

> The in-class assignments were not clear…and we were expected to figure everything out for ourselves. The basic and overall outlook should be taught to understand Earth Science not to walk out as a scientist.
> The answers should have been cut and dry and not up to our imagination.

STUDENT INTERVIEWS

Several students were randomly selected for semi-structured interviews that discussed course procedures in both traditional and IBL sections. All interviews were conducted by an assistant professor or graduate student from the College of Education. Students from all grade levels (A to F) were selected, but in at least one class, no "A" students were interviewed. Students reported that although they were initially skeptical, they preferred that the instructor chose to assign working groups and that they enjoyed getting to meet new people. They stated that the group arrangement took away the impersonal feeling of a large class, provided an opportunity to participate, gave students a peer to explain the material, and let them hear the opinions of others. Most students preferred the activities to a traditional lecture class.

Student comments on the use of the electronic personal response system (Figure 3) to answer conceptest questions were universally positive. Using it gave the students an opportunity to test their understanding of course material, let them discuss their answers with others, and added vitality and interest. Students identified a strong link between class activities, the homework, and the tests. Responses to the question, "On a scale from 1 (take any other course but this) to 10 (don't miss this course), what number would you give this course?" ranged from 5 to 10 with 7 being the most common rating. Students enjoyed the participation resulting from the group work, admitted that they increased their knowledge, and would recommend the course to their peers.

Student Retention - We measured student retention by counting the number of students present for the first and last exams. Data from Steer's classes (Figure 7) showed a

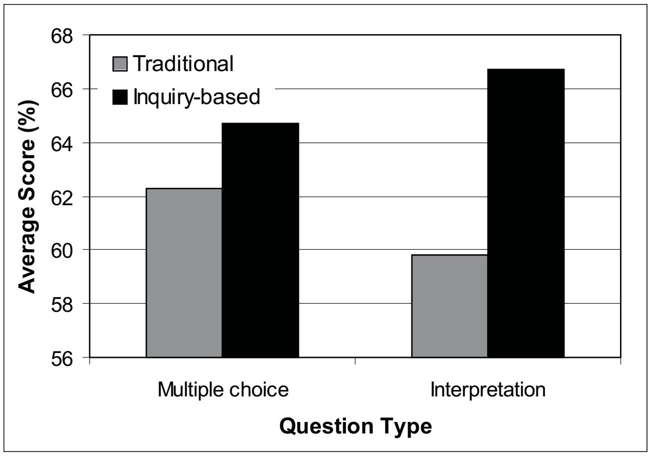

Figure 8. Average score on multiple-choice and interpretation questions on exams for traditional lecture (gray) and inquiry-based classes (black) taught by the same instructor during fall 2000. Average score was 2% higher for IBL section on multiple-choice questions even though this section received less time on content during each class period. Scores for interpretation questions were 7% higher in the inquiry-based section. Note: we attribute the relatively low scores on the multiple-choice questions to our first attempt at generating large numbers of the more challenging comprehension questions. Average scores in subsequent classes have increased by approximately 10%.

14% increase over previous years in the proportion of students who remained for the last exam as he added more IBL exercises to the course. McConnell had 8% greater retention in the IBL section in comparison to the traditional class (Figure 7) taught the same semester.

Exam Scores - Ensuring sufficient content coverage is a concern for many instructors when considering alternative teaching methods (Gold, 1988; Angelo, 1993; Ege et al., 1997). The traditional and IBL sections of McConnell's course took the same exams. Students in the IBL class slightly outperformed the traditional class on all four exams despite less direct content coverage during lecture in the IBL section (Figure 8). A more significant discrepancy was identified in the interpretation questions. Twelve short-answer interpretation questions that involved analysis, synthesis, or evaluation were distributed over the four exams. The average score on the on these questions was 7% greater in the IBL section (Figure 8).

GALT -The Group Assessment of Logical Thinking test (GALT; Roadrangka et al., 1982) is an assessment instrument that measures logical thinking skills. Higher-order thinking skills require mastery of logical operations such as proportional reasoning, controlling variables, probabilistic reasoning, combinational analysis, and correlational reasoning (Roadrangka et al., 1982). The abbreviated form of the GALT survey contains twelve illustrated questions, a pair for each of the five logical operations listed above and another two that evaluate conservation. All questions, except those dealing with combinations, are presented in a multiple-choice format where students must select an appropriate answer (four choices) and the justification for the answer (four choices). The answer is considered wrong unless both choices are correct. The combination questions require that students identify potential groupings of different objects. Student GALT scores ranged from 1-12.

The GALT is a valid and reliable instrument for measuring logical thinking in student populations from sixth grade through college and consistently yields higher scores with increasing grade level (Roadrangka et al., 1982; Bitner, 1991; Mattheis et al., 1992). Furthermore, higher GALT scores correlate with other measures of academic achievement such as grades, SAT scores, and grade point average (Bunce and Hutchinson, 1993; Nicoll and Francisco, 2001). The GALT instrument was administered as a pre- and post-test to two IBL sections in Fall 2001. Both IBL sections showed a statistically significant 6.3% improvement in average GALT scores over the length of the semester. The same instrument showed no change in score for two traditional-format sections of Earth Science taught by different instructors.

SUMMARY

A variety of learning strategies were incorporated into large, introductory Earth Science courses for non-majors. A traditional lecture course was converted into an active learning environment through the incorporation of formative assessment methods matched to different levels of cognitive development. Such a conversion can be readily accomplished through a combination of short lecture segments and group assessment exercises. Improvements in student achievement on exams, retention in courses, and logical thinking skills were documented. A majority of students viewed the active learning methods positively.

ACKNOWLEDGMENTS

The authors thank Steve Boss, Rachel Teasdale, and Germaine Taggert for their helpful reviews of a draft of this paper. This material is based upon work supported by the National Science Foundation under Grant No. 0087894.

REFERENCES

Allard, D.W., and Barman, C.R., 1994, The learning cycle as an alternative method for college science teaching, Bioscience, v.44, p. 99-101.

American Geophysical Union, 1994, Report of the AGU Chapman Conference on scrutiny of undergraduate geoscience education; 55p.

Anderson, L.W., and Krathwohl, D.R., 2001, A Taxonomy for Learning, Teaching, and Assessing: A revision of Bloom's Taxonomy of educational objectives. Longman.

Anderson, L.W., and Sosniak, L.A., 1994, Bloom's Taxonomy: A forty-year retrospective, National Society for the Study of Education.
Angelo, T.A., 1993, A "Teacher's dozen", American Association of Higher Education Bulletin, April, p. 3-13.
Angelo, T. A., and Cross, K. P., 1993, Classroom assessment techniques, A handbook for college teachers (2nd ed.). San Francisco: Jossey Bass.
Astin, A.W., 1993, What matters in college? Four critical years revisited, Jossey-Bass, 482p.
Bitner, B.L., 1991, Formal operational reasoning modes: Predictors of critical thinking abilities and grades assigned by teachers in science and mathematics for students in grades nine through twelve, Journal of Research in Science Teaching, v.28, p. 265-274.
Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., and Krathwohl, D.R., 1956, Taxonomy of educational objectives: Handbook 1: Cognitive domain, David McKay.
Brunkhorst, B.J., 1996, Assessing student learning in undergraduate geology courses by correlating assessment with what we want to teach, Journal of Geoscience Education, v. 44, p.373-378.
Bunce, D.M., and Hutchinson, K.D., 1993, The use of the GALT (Group Assessment of Logical Thinking) as a predictor of academic success in college chemistry, Journal of Chemical Education, v.70, p. 183-187
Bykerk-Kauffman, A., 1995, Using cooperative learning in college geology classes, Journal of Geological Education, v.43, p. 309-316.
Chickering, A.W., and Gamson, Z.F., 1987, Seven principles for good practice in undergraduate education, American Association of Higher Education Bulletin, March, p. 3-7.
Collins, A., 1997, National science education standards and the college instructor. In Siebert, E.D., Caprio, M.W., and Lyda, C.M., (Eds.), Methods of effective teaching and course management for university and college science teachers, Dubuque, Iowa; Kendall/Hunt, p. 51-60.
Crouch, C.H., and Mazur, E., 2001, Peer Instruction: Ten years of experience and results, American Journal of Physics, v. 69, p. 970-977.
DeCaprariis, P.P., 1997, Impediments to providing scientific literacy to students in introductory survey courses, Journal of Geoscience Education, v.45, p. 207-210.
Ebert-May, D., Brewer, C., and Allred, S., 1997, Innovation in large lectures – teaching for active learning, Bioscience, v.47, p. 601-607
Ege, S.,N., B.P. Coppola, and R. Lawton, G., 1997, The University of Michigan Undergraduate Chemistry Curriculum: 1. Philosophy, Curriculum, and the Nature of Change, Journal of Chemical Education, v. 74, p. 74-83.
Freedman, R.L.H., 1994, Open-ended questioning: A handbook for educators, Menlo Park, CA: Addison-Wesley.
Fuhrman, M., 1996, Developing good multiple-choice tests and test questions, Journal of Geoscience Education, v. 44, p. 379-384.
Gibbons, A., 1994, A time of trials and tribulations,: Science, v.266, p. 844, 845, and 851.
Gold, M., 1988, Chemical education: An obsession with content, Journal of Chemical Education, v. 65, p. 780-781.
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, v. 66, p. 64-74.
King, A., 1995, Designing the instructional process to enhance critical thinking across the curriculum, Teaching of Psychology, v.22, p. 13-17.
Lord, T.R., 2001, 101 reasons for using cooperative learning in biology teaching, American Biology Teacher, v.63, p.30-103.
Mazur, E., 1997, Peer instruction: A user's manual, Prentice Hall, 253p.
Macdonald, R.H. and Korinek, L., 1995, Cooperative-learning activities in large entry-level geology courses, Journal of Geological Education, v. 43, p. 341-345.
Mattheis, F.E., Spooner, W.E., Coble, C.R., Takemura, S., Matsumoto, S., Matsumoto, K., and Yoshida, A., 1992, A study of the logical thinking skills and integrated process skills of junior high school students in North Carolina and Japan, Science Education, v.76, p. 211-222.
McManus, D.A., 2002, The two paradigms of education and the peer review of teaching, Journal of Geoscience Education, v.49, p. 423-434.
Murck, B., 1999, Promoting and evaluating participation in high-enrollment undergraduate courses, Journal of Geoscience Education, v. 47, p. 131-134.
National Research Council, 1997, Science teaching reconsidered, National Academy Press, 88p.
National Research Council, 2000, Inquiry and the national science education standards, Washington, DC, National Academy Press.
National Science Foundation, 1996, Shaping the future: New expectations for undergraduate education in science, mathematics, engineering, and technology, 76p.
Nicoll, G., and Francisco, J.S., 2001, An investigation of the factors influencing student performance in physical chemistry, Journal of Chemical Education, v.78, p. 99-102.
Nuhfer, E.B., 1996, The place of formative evaluations in assessment and ways to reap their benefits, Journal of Geoscience Education, v. 44, p. 385-394.
Pinet, P.R., 1995, Rediscovering geologic principles by collaborative learning, Journal of Geological Education, v.43, p. 371-376.
Prothero, W.A., Jr., 2000, Keeping our focus; a perspective on distance learning and the large introductory science class, Computers in Geosciences, v. 26, p. 647-655.
Roadrangka, V., Yeany, R., and Padilla, M., 1982, GALT, Group Test of Logical Thinking, Athens, Georgia: University of Georgia.
Reynolds, S.J., and Peacock, S.M., 1998, Slide observations – Promoting active learning, landscape appreciation, and critical thinking in introductory geology courses, Journal of Geoscience Education, v.46, p. 421-426.

Siebert, E.D., and McIntosh, W.J., 2001, College pathways to the science education standards, National Science Teachers Association Press, 192p.
Silberman, M. 1996, Active Learning: 101 Strategies to Teach Any Subject.
Sundburg, M.D., Dini, M.L., and Li, E., 1994, Decreasing course content improves student comprehension of science and attitudes towards science in freshman biology, Journal of Research in Science Teaching, v.31, p. 679-693.
Tobias, S., 1990, They're not dumb, they're different: Stalking the second tier, Research Corporation, 94 p.
Tobias, S., 1992, Revitalizing undergraduate education: Why some things work and most don't, Research Corporation, 192p.
Trice, A.G. and Dey, E.L., 1997, Trends in faculty teaching goals: A longitudinal study of change, Journal of College Student Development, v. 38, p. 527-534.
University of Akron Factbook, 2001 http://www.uakron.edu/vpcio/iplan/factbook/
Wyckoff, S., 2001, Changing the culture of undergraduate science teaching, Journal of College Science Teaching: v.30, p. 306-312.
Zeilik, M, Schau, C., Mattern, N., Hall, S., Teague, K. and Bisard, W., 1997, Conceptual astronomy: A novel model for teaching post-secondary science courses, American Journal of Physics, v. 65, p. 987-996.