DOCUMENT RESUME

ED 373 998 SE 054 789

AUTHOR	Glick, Judith Gail
TITLE	Effective Methods for Teaching Nonmajors Introductory College Biology: A Critical Literature Review.
PUB DATE	21 Apr 94
NOTE	96p.
PUB TYPE	Information Analyses (070) -- Guides - Non-Classroom Use (055)
EDRS PRICE	MF01/PC04 Plus Postage.
DESCRIPTORS	*Biology; *College Science; Educational Research; Higher Education; Individualized Instruction; *Introductory Courses; Learning Modules; Literature Reviews; *Nonmajors; *Science Instruction; Science Teachers
IDENTIFIERS	Small Group Communication

ABSTRACT

Many educators are concerned about the effectiveness of traditional methods of teaching science to nonmajors. This document provides a literature review of applicable research on teaching nonmajors introductory college biology, a discussion of the information gathered, and recommendations for future research. The review examines the following topics: (1) Survey of College Science Instructors; (2) Instructional Approaches to Promote Scientific Thinking; (3) Laboratory Teaching Approaches; (4) Small-Group Discussions; (5) Individually-Paced Modular Instruction; and (5) Written Materials to Enhance Instruction. Contains 47 references. (ZWH)

Effective Methods for Teaching

Nonmajors Introductory College Biology:

A Critical Literature Review

Presented to

Doctoral students and staff
Department of Science and Mathematics Education
Oregon State University
April 21, 1994

Major Professor: Dr. Norman Lederman

Judith Gail Glick

Submitted to meet the library research paper requirement of the preliminary examination process of the Department of Science and Mathematics Education doctoral program.

Effective Methods for Teaching Nonmajors Introductory College Biology: A Critical Literature Review

Most college graduates in the United States would probably describe their postsecondary science education experience something like this: "We had lectures in an auditorium-classroom, the professor showed a lot of slides and overheads, the tests were killers, and we had a quiz at the beginning of the lab to make sure we read the manual before coming to class." There would be variations on this theme, but this pretty well describes the traditional method for teaching introductory science courses at the college level.

Concerns regarding the effectiveness of the traditional methods for teaching science are not new. Two articles were found which dealt specifically with introductory college biology teaching methods based on research done prior to 1950, and there is a substantial body of literature addressing science teaching methods at the elementary and secondary levels which begins even earlier (Boenig, 1969; Lawlor, 1970; Swift, 1969). A significant movement toward science education reform began in the late 1950s spurred largely by the success of the Soviet Union's space program and the associated perception that the United States lacked the scientific brainpower to compete in the modern world. Many new curriculum development projects for teaching science at the secondary and elementary levels were launched, and programs were implemented to better train teachers in science content and new instructional methods (Yager, 1981).

TABLE OF CONTENTS

The Commission on Undergraduate Education in the Biological Sciences (CUEBS) was funded by the National Science Foundation from 1963 to 1972. The Commission addressed the curriculum for biology majors, the role of biology in the liberal education, and the preparation and continuing education of biology teachers at the elementary, secondary, and college levels (Sundberg, 1991). One of the major recommendations of the Commission was the implementation of investigative laboratory activities. According to a recent report: "Most biology faculty still support the position set forth by CUBES more than two decades ago, but little has been done to achieve the most important objectives of laboratory instruction, that is, involving students in investigations" (Sundberg et al., 1992).

Changes in postsecondary science teaching practices have come very slowly for a number of reasons. Science courses at the college level are typically taught by scientists who have not been trained in instructional methods. Science professors have been inclined to teach the way they were taught and have had little exposure to new teaching ideas (Gottfried et al., 1993; Sundberg et al., 1992). Science faculty endorse the idea of improving students' understanding of the processes and nature of science, but their actions seem to indicate that content coverage is still considered the most important aspect of science teaching (Sundberg et al., 1992). Further, it has been assumed that college students are mature adults, and that they should be able to learn material by reading the textbook and hearing it in lecture (Fisher et al., 1986), and when students fail to do so it is blamed on their poor precollege preparation (Gottfried et al., 1993; Uno, 1988).

During the 1980s, there was a renewed concern about the scientific literacy of the nation, and an additional focus was placed on college science courses (American Association for the Advancement of Science [AAAS], 1989; 1990; Sigma Xi, 1989). Scientific professional organizations recognized the need for improved instructional methods at the postsecondary level and began offering training and publications to this effect (American Institute of Biological Sciences [AIBS], 1991; Moore, 1985). College science courses are being restructured to address the needs of the 21st century and to more appropriately serve the nonscientific population. Institutions of higher learning are offering courses that integrate the traditional sciences and address current issues such as environmental problems (Lawson, Rissing & Faeth, 1990; Malachowski, 1990; McIntosh & Caprio, 1992; Morgan, Lemons, Carter, Grumbling & Saboski, 1993).

Advances in learning and developmental theories have also shed some light on appropriate methods for teaching introductory college science to nonscience majors. Perry (1970) and Kitchener and King (1981) found that many college freshmen are dualistic (right or wrong) thinkers and are unable to judge knowledge claims based on the strength of the argument, an element critical to understanding science. It is now widely held that many young college students function below the Piagetian stage of formal operations which is essential for abstract thinking; it was earlier believed that this stage was entered in early adolescence

Lawson, 1992; Dunlop & Fazio, 1976). Cognitive psychology and the constructivist epistemology have also provided new theories regarding methods for building concepts and addressing misconceptions which have been incorporated into college science teaching (Fisher et al., 1986; Heinze-Fry, 1992).

In the 1990s we are faced with a renewed concern about the perceived scientific illiteracy of the American population and fears that we as a nation will suffer economically as a result. The concerns about scientific literacy are being championed by many major scientific and educational societies including the American Association for the Advancement of Science (AAAS), the American Institute of Biological Sciences (AIBS), the National Association of Biology Teachers (NABT), and the National Science Teachers Association (NSTA) to name a few. Through membership in these societies, many college faculty share personal concerns about the perceived inadequacy in the current teaching methods and have voiced the desire to improve, however they lack the training in specific instructional methods to affect the desired changes (Caprio, McIntosh, & Koritz, 1989; Gottfried et al., 1993). The journals of the societies listed previously and others offer forums to share teaching ideas and course descriptions, and these are usually accompanied by personal testimony as to their effectiveness. A very limited amount of empirical research has been done on college biology teaching (Gottfried et al., 1993). The question remains: What instructional methods are effective for teaching introductory college biology to nonscience majors?

Effective teaching has a variety of meanings, but for the purposes of this paper it be will broadly defined as: (a) providing the student with a working knowledge of biology content appropriate and/or necessary for members of the educated public, (b) developing in the student an understanding of the nature of scientific inquiry and its role in society, and (c) enhancing the student's thinking and reasoning skills so that she/he can evaluate alternative claims, especially in the field of biology. This definition is based on a synthesis of goal statements by the following professional groups: NABT Standards in College Biology Teaching Committee (Gottfried et al., 1993); National Science Teachers Association (1982); and AAAS (1989, 1990).

The literature on effective college science teaching is sparse and generally of poor quality. Papers were selected for this review based on the following criteria:

1. The paper described empirical research which addressed a question related to effective teaching methods and provided evidence to support conclusions. There are many opinion papers and "how-I-did-it" reports related to college science teaching which are widely cited in support of various instructional techniques; these papers were not addressed in this review.

2. The research subjects were college students in the United States. International studies were eliminated because university students in most foreign countries are a much more select group than in the U.S. Studies with high school students as subjects were not included because of differences in cognitive and intellectual development, and because students attending high

school are not representative of those who attended college.

3. Some aspect of the sample or the treatment needed to indicate that it was generalizable, in at least a small degree, to a population of students enrolled in introductory nonmajors biology. This final criterion required subjective judgement and a broad definition of generalizability. Very few studies have been done in nonmajors biology courses. Several studies were included in which the subjects were definitely, or possibly, science majors, but in these cases the treatment was not judged to affect students who had declared majors in science any differently than nonmajors. Studies were also included which dealt with chemistry or geology courses: In these cases, the nature of the content of the courses was judged to be similar to instruction in introductory biology.

4. And finally, the written report needed to be available in published journals or through the ERIC Document Service. Unpublished dissertations and theses were not included because they are not readily accessible to those hoping to improve their teaching based on research results, and the fact that they are unpublished is possibly a commentary on the quality of the information contained.

No claim is being made that this review is exhaustive based on the above criteria. An extensive search was conducted to find all applicable research, however some studies may have escaped detection. Additionally, another prudent individual, using the same criteria, may have chosen to include some of the rejected studies or eliminate some of those reviewed.

The 20 selected papers were organized by topic for the ease of presentation. These topics do not represent any *a priori* research questions, and no discussion or conclusions were developed based on topic headings. The first section includes only one paper reporting on a survey of college instructors. The second section contains reviews for four projects addressing the teaching of scientific thinking. Seven papers that dealt with teaching laboratories were combined to form the third section. The remainder of the review contains sections on small-group instruction (two papers), individualized approaches (three papers), and written aids to instruction (three papers).

A discussion of the information in the 20 articles follows the review of the literature and includes recommendations for instructional practices based on the evidence presented. A final section presents recommendations for future research.

REVIEW OF THE LITERATURE

Survey of College Science Instructors

A survey of nonmajors' science instructors serves as a good introduction to the problems faced in such courses and offers direction for future research. While there is no shortage of opinion papers expounding the difficulties faced in teaching today's college students, the paper reported in this section is the only piece of empirical research addressing instructors' concerns with regard to nonmajors science courses.

A national survey was conducted by **McIntosh and Caprio**

(1990) to determine the quality of postsecondary nonmajors' science instruction. In 1988, a questionnaire containing 17 multiple-choice items and one open-ended question was sent to 763 college and university science professors that were members of the Society for College Science Teachers (SCST). There was a 53% return rate, and the authors state that "none answered all of the questions" (p. 28). The authors conceded that the survey sample could have been biased since it was sent only to SCST members, and these may represent a more involved group of postsecondary educators. No mention was made of the low return rate, but this may have been due, in part, to SCST members who were not teaching any nonmajors' courses. The investigators felt that the large number of respondents (405) allowed for some tentative conclusions to be drawn.

Percentages for each response of the multiple-choice questions were reported in categories of course demographics, course characteristics, and science educators' opinions. Forty-one percent of the respondents taught nonmajor biology and 27% taught chemistry. There was a fairly even representation from two-year, four-year, and four-year/graduate institutions. About half taught courses with enrollments greater that 50 students, and 28% taught courses with more than 100. Fifty-nine percent of the courses were exclusively for nonmajors, and the remainder combined majors and nonmajors. It is unclear whether the term nonmajor refers to 'not majoring in any field of science' or 'enrolled in a biology course but not majoring in a biological science'. The tone of report indicated that the interest is in those not majoring in science, however, it is possible that the questions were not answered in this way.

The survey revealed that the laboratory experience is considered an important part of a nonscience major's education by the group of faculty responding. In response to: "Should Lab Work Be Required of Nonscience Majors," 85% marked *yes*. About 60% reported a laboratory component for their nonmajor science course. Only 12% reported using a published lab manual as the sole source of laboratory exercises; 45% indicated that all exercises were written by the instructor. It appears that the quality or nature of commercially available laboratory materials is not meeting the needs of nonmajor science courses.

The most critical problems in teaching nonmajor science were reported in rank order as: (1) poor preparation in reading and writing, (2) poor preparation in mathematics, (3) lack of motivation, (4) inability to reason, and (5) fear of science. These results appear to have emanated from a multiple-choice type question, and it is unclear whether additional choices were available, or how respondents selected their answer (ranking all choices or indicating only one).

The open-ended question asked the professors to report any methods they had found particularly useful in teaching science to nonmajors. These results were only reported briefly and it is impossible to determine how frequently the various suggestions occurred. General suggestions revolved around getting the students involved in the learning process using such techniques as questioning, group discussion and problem-solving sessions.

As stated above, the return rate for the nonmajors' science instructors survey was just over half, and it was originally sent to a possibly biased group of science faculty. The reported results need to be interpreted with this in mind.

Instructional Approaches to Promote Scientific Thinking

Contrary to popular belief, the idea of teaching students to think in addition to learning scientific facts and concepts is not a new one. The first two studies reviewed involve college science instruction before 1950 and both specifically addressed the need for students to learn more that just the facts. The other two studies are from the early 1980s and examine instructional approaches to enhance achievement in thinking skills. All four studies employed a technique for enhancing student evolvement and classroom discussion; they all attempted to assess content acquisition as well as some measure of scientific thinking.

The earliest published study found that dealt with postsecondary biology instruction was done by **Barnard** (1942) at New York University, School of Education. The investigation involved the biological portion of a science orientation course enrolling all undergraduate classifications, but with 67% either juniors or seniors; it is assumed that all were education majors. No additional information regarding the subjects of this study was provided.

Barnard's study compared the relative effectiveness of a lecture-demonstration method and a problem-solving method of teaching with respect to students' recall of specific information, understanding of generalizations, abilities in problem solving, and scientific attitudes. Three classes were taught by each method: Lecture-demonstration totalled 137 students; problem-solving had 145. No mention was made regarding how students or classes were assigned to treatments. The instructor or instructors were not described. No timeframe was provided for the length of treatment or number of weekly contact hours. A note in the data tables indicated that some classes completed the course first semester, some second semester, and some met once a week for the full year.

Great detail was provided on the nature of classroom instruction and assignments for each of the treatments, including similarities between the two. All classes dealt with biological problems faced by humans that were organized into six instructional units. At the beginning of each unit, bibliographies were provided including required textbook assignments and additional references. At the end of each unit, all students wrote reports on the particular problem relating appropriate biological generalizations. All classes were instructor-directed and included live demonstrations and audio-visual materials. No observations were made to document teaching approaches, however descriptions of the two treatments indicated that great care was taken to insure treatment differences.

The lecture-demonstration method used formal lectures to present the subject matter supplemented with demonstrations which illustrated the concepts. The first class meeting "stressed the meaning of science as a method of solving problems and included a lecture to the students on the elements of problem solving and the scientific attitude' (p. 122). Subsequent lectures covered major biological problems and generalizations related to the problem. Important points were outlined on the blackboard and students were instructed to record this in their notes. Students' questions were not encouraged, but when asked, were answered directly.

The problem-solving method was designed "to encourage student participation in formulating the major problems of the course, analyzing each problem into its specific parts, and proposing and carrying out the various learning activities which would develop understandings of solutions to problems" (p. 123). The instructor presented a question or problem and solicited students' ideas which were recorded on the board and then discussed. Through the instructor's direction, the students came to the same conclusions as in the lecture-demonstration classes. Once the class had identified the major unit problems, they were presented with material which they might use to analyze and understand the issues. Student-selected demonstrations and media were presented by the instructor or teaching assistant with guiding questions to assist in developing generalizations.

To assess difference between teaching methods, four types of tests were developed, one for each of the four outcomes: specific information, generalizations, problem-solving, and scientific attitude. There were three forms of each test; a battery of tests included one test for each outcome. One battery was administered before and after the first half of the course. Another battery was given at the beginning and end of the second half of the course. A third battery was administered before the beginning of the course and again at the completion of the course. No explanation was given for the seemingly excessive number of tests administered to each subject, and this design is surely susceptible to threats to validity involving test sensitization.

Objective tests for recall of specific information were constructed using the subject matter outline for the course. The tests were reviewed by a "a group of qualified jurors" (p. 126), and modified accordingly. These test forms can be said to have face validity, and depending on the rigor used in matching the items to the subject matter outline and in analysis by the jurors, they may have established content validity. (No specific validity claims are made by the author.) Items showing poor discrimination ability were discarded before calculating final scores. The Spearman-Brown coefficients of reliability for the three forms were .43, .68, and .81.

The methods of construction and validation were identical for the tests of understanding of generalizations as for subject matter tests described above. Spearman-Brown coefficients were .55, .65, and .75.

The tests on problem-solving dealt with the "abilities to recognize problems, analyze problems, evaluate information, formulate generalizations and evaluate conclusions" (p. 126) and contained objective and free-response questions. Face validity appears to have been established by having the tests reviewed and modified by a group of jurors, however no specific validity claims were made. Scores on the objective portion were simply the total number of correct answers. Free responses were judged by two different jurors who followed a system of classifying answers using defined categories. A vague system was described for assigning weights to various categories with regard to the elements of problem solving and relating relative scores of the two jurors to obtain a final score. Interrater agreement was determined, however the reported values are combined with those for the attitude test so it is impossible to report the specific coefficients. For the two tests, in 18 of 19 cases, the product-moment coefficients of correlation ranged from .70 to .94, in the other it was .57. Coefficients of stability (test-retest reliability) were determined by correlating scores made by the same student on the same version of the test administered two weeks apart. Values obtained for the three versions of the problem-solving test were .67, .53, and .51. These reliability values are unacceptably low, however higher values would also be meaningless because subjects were most likely test-wise and could have developed problem-solving skills in the intervening time.

The three scientific attitudes tests each contained eight problematic situations in which the subjects needed to devise a course of action and explain their reasoning. Face validity was established by a panel of jurors before the tests were administered. Scores were determined by "weighted opinions of jurors concerning the extent to which individual student responses showed evidence of the scientific attitude" (p. 126). Interrater agreement was determined as discussed in the preceding paragraph. Coefficients of stability calculated as for the problem-solving test yielded values of .62, .57 and .55.

The description of data analysis is quite elaborate and complex. Since this study was performed over 50 years ago, it seems unreasonable to harshly judge the statistical procedures by criteria used today, however an attempt will be made to critique and extract meaningful information from the data provided. Three areas of data analysis will be discussed below: (a) equating comparison groups, (b) significant score differences, and (c) practical significance.

Six pairs of classes were identified for comparison in each of the four dimensions of this study (24 total pair-wise comparisons). Two pairs of classes for each of the three administrations of the tests were selected based on a procedure of equating classes. For each test administration, an index was calculated for each student which combined a standardized pretest score and a standardized score on a psychological exam. Students in matched classes were paired based on their scores and this was said to create equated classes. Comparing means of combined standardized scores for paired classes using some unknown statistic seemed to produce satisfactory results, but no significance value is given, only the raw statistic. Giving the benefit of the doubt, it appears that a substantial effort was made to establish equivalent comparison groups based on pretest and psychological scores, and for the lack of strong evidence to the contrary, it will be assumed that it was achieved.

The study by Barnard used the quasi-experimental nonequivalent control-group design. Analysis of covariance should have been performed using the pre- and posttest data, however no statistical test were performed. Data charts are provided for each test area including class mean scores on the final test, difference between means for paired classes, and standard deviation. The treatment having the advantage is indicated, but no statistical tests were applied to determine significance. Interestingly enough, the following comments appeared in the text raising further question as to why statistical tests were not performed: "The differences were sufficiently great to insure practical certainty of being significant differences" (p. 130), and "there was practical certainty that the obtained differences represented true differences" (p. 131). Since the data are presented, an interested person could now perform the calculations with the aid of a computer and add strength to the conclusions made in this study.

In all six pairs, the lecture-demonstration group had higher mean scores on the final test of specific information than the problem-solving group. Mean scores ranged from 10.11 to 25.27; total possible points was not reported. Three pairs had a greater than a 2.3 point difference which may be of some practical significance if the pairs were truly equated as claimed. For the other three pairs, the difference does not appear to represent one of practical significance. Cursory examination of the data for the generalization test reveals nothing of practical significance, and seems to support the claim of no difference between teaching methods in affecting this aspect of learning.

Reported final scores on the problem-solving test ranged from 111.30 to 220.90 with no indication of total points possible. The differences ranged from 8.89 to 24.96, with all but one pair having differences greater than 17 points, and in all cases favored the problem-solving method. This would appear to be of practical significance, however the poor reliability of the instrument and great range of scores among various classes leads to questions regarding the meaning of this data.

Science attitude test mean scores ranged from 15.50 to 54.96 with differences of 5.56 to 12.19 all favoring the problem solving method. These data would appear to be of some practical significance despite the questionable reliability of the instrument. (Recall that despite low reliability values, test stability was calculated in a situation where students were most likely test-wise, so in fact, the instrument may be of better quality that indicated.) If the original data were available, internal consistency could be determined and proper statistical analysis could be performed to determine the appropriate conclusions to be drawn from these data.

It appear that no difference was revealed between instructional methods for teaching specific facts and scientific generalizations, however there may be some preliminary evidence that a problem-solving approach may contribute to higher scores on tests of problem-solving ability and scientific attitudes. These results should be considered to have the strength of casual observations and not experimental evidence. Further, the conclusions have very limited generalizability to freshmen-level biology instruction in the last decade of the Twentieth Century: Depression Era upperdivision education majors were surely dissimilar to today's entering college student in a number of educationally important attributes.

Concern about teaching the nature of science in addition to biological content led to a large scale study during the 1949 - 1950 school year at Michigan State College, East Lansing (Mason, 1952). **Mason** implemented four different treatments for a full school year and assessed differences using six different sets of instruments--all described in great detail. Despite the thoroughness of the research report, this study adds little to our understanding of appropriate instructional practices for today's postsecondary students. The results can not be expected to have population validity for American college and university classes in the 1990s because the post-World War II college population had a very different profile from today, particularly with the large number of G.I. Bill students. Because of the limited applications to present-day instructional practices, an extensive critique of this very detailed report does not seem warranted. Omissions of detail from this review are not intended to imply that the original report lacked the necessary information.

Students were apparently required to take all three terms of the General Biology sequence which met for two one-hour lectures and one two-hour laboratory each week. Each of two lecture classes had four laboratory sections and students retained the same lecture-laboratory assignment for the full year. One lecture class was taught by the scientific thinking method and the other by the descriptive method. For each lecture class, two of the associated laboratory sections were taught using the Guide for Laboratory Studies and the other two used Constructed Notebooks. The investigator taught all lectures and laboratories by the designated method for the entire year. No observations were made to verify teaching approaches.

No information was given to indicate equivalency of groups: no randomization of treatment assignments and no description of students' characteristics. It is difficult to imagine that there was no mortality of subjects considering the length of the treatment period, however this was not discussed in the report. The number of subjects in each group was given as: Thinking-Guide, 45; Thinking-Notebook, 36; Descriptive-Guide, 48; and Descriptive-Notebook, 42.

Mimeographed notes containing the same outline of "factual subject matter" (p. 274) were distributed to students in all groups at the beginning of each lecture period. For the descriptive method, the lecturer presented the subject matter following the notes given to the students and did not encourage any student involvement. For the scientific thinking method, lectures included many activities "designed to give students practice in scientific thinking by responding to factual and to attitudinal statements" (p. 274), and blanks were provided in the notes for the answers. Exhaustive descriptions left no doubt that the lecture treatments were distinct, and the scientific thinking method attempted to foster understanding in the methods and nature of science.

Both laboratory treatments were based on the activities in the departmentally prepared laboratory manual which provided general procedures for activities designed to promote scientific thinking and required very little direction from the instructor.

> All studies were written so the student must make direct observations on biological material; collect data from other reliable sources; analyze the data and draw conclusions from them. The student is asked to formulate hypotheses and to test them by further observation including experimentation. He is frequently requested to suggest two or more hypotheses to explain a given set of facts. One of these may well be the textbook explanation, but he must think for himself to devise a second hypothesis. The student is not asked for the correct explanation only, but for as many logical explanations as he can imagine (p.277).

The guide lab sections used the laboratory manual as designed. The notebook groups were presented the same material in the form of demonstration-lectures in which all interpretations and conclusions were given by the instructor. They did not have a copy of the laboratory manual and were required to construct their own notebooks following the instructor's guidelines.

To assess differences, six different instruments were administered a total of 16 times; analyses of variance and covariance were used for 88 statistical comparisons. Modified Kuder-Richardson formula reliability coefficients (specific formula not given) were reported for every administration of an instrument as were descriptions of instrument validation. Unless otherwise noted, quality of the instrumentation was not a problem. Absolutely no statistical or raw data were provided to support the claimed results.

A published college biology achievement test was administered as a pretest and at the end of each term. Final exams were given at the end of the first and second quarters, and the departmental Comprehensive Examination in Biological Science was administered at the end of the year. All of these instruments were used to measure factual biological understanding, and had good validity and reliability. The results in a nutshell: No end-of-year differences were found with respect to lecture, laboratory, or lecture-laboratory teaching method when looking at acquisition of factual information.

Scientific thinking was measured by extracting the questions on the Comprehensive Examination in Biological Sciences that specifically addressed thinking skills (recognize cause and effect, interpret data, draw conclusions, test hypotheses, identify and solve problems, critique experimental procedures, and evaluate real situations with scientific implications). Five experts validated the instruments and it had a KR value of .87. Based on this instrument, neither of two methods of instruction in lecture resulted in differences in ability to think scientifically. The author reported that the thinking method in laboratory (guide) appears to be more effective in promoting scientific thinking but does not say that a statistically significant difference was found.

<u>Scientific attitudes</u> are described as "habits of thinking and acting" (p. 271) and were specifically taught for and measured in Mason's study following the list first formulated by Victor H. Nolls in 1935:

1. Habit of accuracy in all operations, including calculations, observation, and report.
2. Habit of intellectual honesty.
3. Habit of open-mindedness.
4. Habit of suspended judgment.
5. Habit of looking for true cause and effect relationships.
6. Habit of criticism, including self-criticism.

(quoted in Mason, p. 271)

Two forms of an instrument developed by Nolls were used as a pretest and administered at the end of each term as posttests. Correlations between the forms and KR values were all admitted to be unacceptably low by the author (all values were below .60) but this was believed to be the best instrument available at the time. No mention was made of pretest sensitization. Results of this test revealed no meaningful differences between groups, but did indicate that all groups did make significant changes in scientific attitude from the beginning to the end of the year.

Another scientific attitudes instrument was developed specifically for this study and incorporated some of the same situations that were presented in the scientific thinking lectures. The instrument was reported to possess curricular validity (= content validity) since is was written using the same material presented in class. It was administered during the last scheduled lecture period of the year and had a reliability coefficient was .69. This test revealed that the scientific thinking lecture method was more effective at teaching scientific attitudes than the descriptive lecture method. (Students who were presented the material prior to the test did better than those who had no experience with it!) However, when looking at lecture-lab combinations, students in the group with no emphasis on thinking (descriptive-notebook) did better than those who had thinking emphasized in lecture or lab, but not both. These unexpected results were not explained or discussed, but may reflect the lack of a pretest and nonequivalent groups.

Since the report of Mason's study lacked the statistical data necessary to draw independent conclusions, and any inferences would have only limited generalizability, this research contributed virtually nothing to our understanding of teaching nonmajors college biology. The main significance of Mason's 1952 research article is that it <u>failed</u> to produce any strong evidence against the straight lecture for presenting factual information <u>and</u> for teaching thinking skills.

Haukoos and Penick (1983) examined the influence of teaching directness on biology content knowledge and science process achievement. The subjects were enrolled in an introductory college biology courses at a large, comprehensive, two-year community college in Illinois. No information is provided about the students regarding major, age, socioeconomic background, or academic abilities. Two intact sections were assigned to each of two treatments. An accelerated five-week class (11 students) was assigned to the Discovery Classroom Climate (DCC) along with one of the standard ten-week sections (23 students). The Nondiscovery Classroom Climate (NDCC) treatment group consisted of two ten-week classes (19 and 25 students).

Great care seems to have been taken to insure that the two treatments were applied correctly. The variable that defined the two treatments was the amount of directness or indirectness in teaching. The DCC treatment involved a large degree of student freedom in laboratory and nonjudgmental discussions in the classroom. The NDCC treatment offered exact laboratory directions and classroom lectures. Teacher behaviors were monitored using audio taping and coded to verify that the treatments were distinct. One teacher taught all classes, thus attempting to eliminate the teacher effect. The same content, textbook, laboratory equipment, and classroom visuals were used for all sections. The only obvious threat to the integrity of the treatments is use of the five-week section along with the ten-week classes.

The Science Process Inventory (SPI), Form D, was used as a pretest-posttest to assess understanding of the processes which lead to scientific knowledge. This published instrument consisted of 135 statements with forced responses of agree or disagree. The SPI was said to have predictive and constructive validity, and a variety of vague methods are described. It is unclear if the authors verified the validity of the instrument for this specific study. The SPI was reported in the literature to have a reliability of .79 established by Hoyt's analysis of variance procedures, and a Kuder-Richardson reliability of .86 was found for the described study.

Content knowledge was assessed using the Biology Achievement Test (BAT) as a posttest only. The BAT had been developed by faculty at the same institution for the purposes of offering biology credit without taking the course. The investigators stated that "neither formal validity of the test questions nor reliability of the test was measured" (p. 632). No information is provided regarding the format or content of this instrument.

The authors stated that a pretest-posttest two-treatment design was used, however no pretest was given for biology content understanding, only for process skills. In addition, the randomization process was by section, not by students, and with only four sections (one of which met for five weeks), the equivalency of groups was very questionable.

A significant difference was found between sections on SPI scores using analysis of covariance (ANCOVA), $F(1,3) = 3.86$, $p < .05$. The researchers correctly used the number of sections to determine the degrees of freedom, but make no mention of meeting the fairly rigid assumptions associated with the use of ANCOVA. For some unexplained reason, the authors performed a Duncan's Multiple Range Test on the unadjusted means by section for the SPI posttest as opposed to the net changes in scores. The ten-week DCC section was found to have a significantly higher science processes mean score, $p < .05$.

The authors graphically presented pre- and posttest science process scores for all sections, but ignored much of the information in the data. Further investigation of the graph reveals that each of the two DCC sections showed about a 2% or 3% gain in mean SPI score and the two NDCC sections had a loss of a similar magnitude. The ten-week DCC section had the highest SPI score on the pretest; the five-week DCC section had the lowest. Changes on the posttest placed the five-week DCC section within the same narrow range as the two NDCC sections (100.6 - 103.4), and the ten-week DCC with a significantly higher posttest score (111.2). As mentioned previously, statistical analyses were not performed on the net differences in SPI scores.

No significant differences in Biology Achievement Test scores were revealed by Analysis of Variance, $F(3,1) = 1.15$, $p < .33$. Since this instrument was not shown to be valid or reliable, and no pretest was given, very little can be learned from these results.

Haukoos and Penick make some broad concluding statements that were not warranted by the reported results. It can be said that no evidence was obtained to indicate that the degree of teacher directness affects biology content outcomes at the community college level. This study also indicated that teaching directness may affect students' understanding of science processes, but not at a level of practical significance.

This study has some limitations which affecr the strength of the conclusions: (a) A test-treatment interaction effect may have been present since the same form of SPI was used as a pre- and posttest, (b) the effect of one accelerated section was not addressed, (c) the subjects were not adequately described to allow for confident generalization, (d) the BAT is not a valid and reliable instrument, and (e) the Duncan's Multiple Range Test was performed on posttest scores only, not taking into account pretest differences. These researchers admirably d [illegible]gned and controlled the two treatments. A repeat of this study with randomly assigned students, a valid and reliable achievement test, and proper statistical analysis could produce meaningful results.

Moll and Allen (1982) described several studies involving an introductory biology program designed to develop knowledge of biology concepts as well as critical thinking skills. The program used short video segments to expose the student to demonstrations and experiments, followed by class discussions. The instructors guided the discussions emphasizing the use of sound reasoning based on the observations. It was said that the students were able to develop basic biology concepts based on their interpretations of the video segments, and then use previously learned concepts to derive more advanced concepts.

Students were also required to write out analyses of problems that required interpretation as well as basic recall of the concepts.

The primary study to evaluate the effectiveness of the video/discussion instructional program was done during the Fall 1980 semester at West Virginia University. Very little information was provided about the subjects except that they were the students enrolled in one section of introductory biology, and included science majors and nonscience majors. A one-group pretest-posttest design was used, and the authors offered no support for the use of this very weak design. The instrument was a 50-question test constructed and face validated by some of the introductory biology program faculty members. Approximately half of the questions focused on content recall and the other half involved application of information. The same instrument was used as the pretest and posttest.

The authors report a significant improvement in student scores for content, critical thinking and overall, $p < .001$. No mention is made of the statistical test used, however mean scores and standard errors are reported for each component on the pretest, posttest, and for improvement by major (science or nonscience), sex, and for all students combined. No significant differences were found between the mean scores of majors and nonmajors, nor those of males and females.

Students at another university were given the same pre- and posttest, however their introductory biology course did not emphasize critical thinking skills. No further information is given regarding the second set of subjects or their course. The second set of students was not claimed to serve as a control group but was used more for informal comparison. An examination of the scores on the application (critical thinking) portion of the posttest revealed that the students in the video/discussion course scored significantly higher, $p < .01$. The actual scores are not reported, nor was the statistical test described.

The paper by Moll and Allen offered some useful suggestions for improving critical thinking instruction in introductory biology courses. The video/discussion format seemed to be based on a sound theoretical framework, and it appeared to be easily implemented once the tapes were created. Unfortunately, the research described has numerous problems, including a very weak design and poor instrumentation, making it difficult to say anything about the effectiveness of the program. A well executed study using the described program as the treatment could provide some valuable information.

Laboratory Teaching Approaches

As mentioned in the review of McIntosh and Caprio (1990), there appears to be considerable importance placed on the role of laboratory as part of nonmajors science courses. The traditional lab manuals seem to be unsatisfactory, and new methods are needed. The first four studies in this section examine methods which allow the students more control over their own learning in the laboratory setting and follow discovery or inquiry approaches. The next paper examines the effect of laboratory instruction emphasizing integrated science process skills in a seemingly behaviorist style (masked somewhat by the discussion of cognitive development).

Finally, two papers will be discussed which took novel approaches to laboratory instruction which were grounded, at least in part, in constructivist epistemology. The first of these studies involved the use of an additional class meeting prior to the laboratory period to introduce the activity, a procedure which was believed to serve as an advanced organizer. The second novel approach required students to work in groups and author interactive video disc lessons to share with their peers. The use of cooperative groups and creative projects are both believed to enhance the construction of new knowledge.

In the first of two studies by **Leonard** (1983) a Biological Sciences Curriculum Study (BSCS) inquiry approach was compared to a more directive commercial laboratory program. The sample consisted of 24 laboratory sections of General Biology at a large midwestern university. The students were mostly freshmen, with both biology majors and nonmajors. Laboratory sections were randomly assigned to treatment or control groups, with each instructor teaching a treatment and a control laboratory section. A randomized (by sections) pre/posttest control group design was used with 208 students in the experimental group (BSCS-style laboratory) and 218 students in the comparison group (commercial laboratory activities). Each laboratory section consisted of approximately 20 students and met for 2 1/2 hours each week for a semester. All students attended three 50-minute lectures per week.

The BSCS-style approach used 13 laboratory investigations developed by the author. Many of the activities were adapted from Biological Science: An Ecological Approach (BSCS Green Version 4th edition, 1978). All activities were rewritten for the university level and a 2 1/2 hour laboratory period. The emphasis of this approach included: (a) the use of science processes, (b) the systematic development of concepts using questioning, and (c) increased discretionary demands on the students for planning learning strategies and for selecting procedural options. Each investigation covered basic biology concepts and skills.

The comparison group used 13 laboratory exercises from the Freeman Separates (a widely-used commercial program for university-level introductory biology) which matched conceptually the BSCS-style activities. This program was deemed to be more directive and less inquiry-oriented than the BSCS approach. No description was given of the methods used to verify the conceptual matching of the two laboratory programs or the existence of treatment differences in the two approaches.

Instructors met once a week with the investigator for training in the two approaches. Student independence was to be encouraged in the experimental group sections and not encouraged in the comparison group sections. Instructors were to give any assistance requested by the comparison group students and to politely refuse or minimally redirect the BSCS experimental group students. It is assumed that no classroom observations were made to verify treatments since it is not mentioned in the report.

The students were given a pretest during the first meeting and the same exam the 14th week as a posttest. The 60-question multiple-choice test on selected biological concepts was developed the semester prior to the study and given to 48 students in the course that term. The pilot test was analyzed for content validity by three university biologists and for reliability by item analysis. The test was then revised, given to the same students, and analyzed for internal consistency using Kuder-Richardson 20. The content validators further judged the test to measure the concepts listed and not to be biased towards either of the laboratory approaches. The following is a list of concepts, and corresponding KR-20 Coefficients: Microscope Techniques, .52; Cell Structure and Function, .61; Cell Transport, .71; Respiration and Photosynthesis, .78; Growth and Development, .70; Genetics, .65; and, Science Processes, .64. Apparently, no reliability measures were made for the administrations of the test during the study. No mention was made of validity of the test with regard to the specific laboratory concepts covered.

It was noted that laboratories were coordinated with three 50-minute lectures per week, and that some of the information learned in lecture could contribute to improved performance on the laboratory exam. However, the author stated that the lecture material was the same for all students, and should not have contributed to net differences between the two groups. No mention was made of attempts to insure that the lecture did not favor one of the laboratory approaches.

To test for group equivalency, a t-test was performed on the pretest scores, and no difference was found. The analysis of group mean scores for the posttest was done in the same way as the pretest, and indicated a significant difference ($t = 3.81$, $p < .005$) between the groups, with the experimental group scoring significantly higher. Even though it was stated that 24 laboratory sections were involved in the study, the reported *n* value for the treatment group was 6 and for the comparison group was 6. The calculation of degrees of freedom for this study was very odd and is difficult to evaluate. Furthermore, inappropriate statistical procedures were used--an ANCOVA for comparing group means using the pretest as a covariate was needed.

An analysis of the posttest by concept area revealed a significantly higher ($p < .05$) mean score by the experimental group on all concept areas except the use of the microscope. The mean overall test score differences translated to 6% or more than half of a letter grade, which would have some practical significance if the same result were revealed by correct statistical procedures.

The statistical problems and weak reliabilities of the instrument limit the usefulness of the results of Leonard's 1983 study. The theoretical framework and the research methods were described in great detail, thus facilitating future work along these lines. If the original data are still available, reliabilities could be calculated, and appropriate statistical procedures could be run, thus offering the possibility of strong meaningful results.

A second study by **Leonard** (1988) was very similar to the one described above, but compared the BSCS-style approach to an Extended Discretion (ED) approach in which the students were given laboratory assignments without detailed procedures to follow. In this study, the BSCS approach was described as guided inquiry (GI), with students given a "relatively clear and linear procedure to lead them through the activity" (p. 80). This was the same approach used in Leonard's 1983 study, yet the description of the degree of student independence seemed to differ. The purpose of this study was to examine the effect of the ED laboratory app_ ach on the learning of biology concepts.

Twenty-four laboratory sections accompanying a nonmajors general biology course at a large midwestern university were randomly assigned to one of the two treatment groups. No other information was provided about the sample. The experimental group (n = 222) used the ED laboratory approach, and the comparison group (n = 245) used the GI approach for the entire semester. Ten instructors taught the laboratories, each teaching at least one section of the experimental group, and one section of the comparison group. The instructors and the investigator met weekly to review the procedures for each laboratory approach. In this study, unlike the previous study, the instructors were permitted to answer any questions of the students using the BSCS-style approach. The students in the ED sections were limited to answers regarding the list of resources available.

The independent variable in this study was the opportunity for discretion through less direction in laboratory procedures. Treatments were examined to determine that the actually produced a difference in the independent variable. First, the number of words of required procedure were counted for each laboratory activity. The mean number of words per activity was found to be significantly different at the .01 level. Next, the two programs were subjected to the Laboratory Structure and Task Analysis Inventory which reported the percentage of laboratory activities within the program which engages the students in various science process tasks. The ED approach had students working from their own procedures in 92.9% of the activities; the GI approach never does. These analyses indicated that the ED approach did engage the students in science inquiry processes more often, however, no classroom observations were made to verify the actual treatments.

Three different assessments were used to measure the students' understanding of concepts: (a) a multiple-choice laboratory final exam, (b) laboratory reports, and (c) six laboratory quizzes. The laboratory exam contained 50 five-choice items with each laboratory topic being represented by a least four items. The exam was administered to a group of 83 students the semester prior to the study and then revised based on "intratopic correlation data, item-analysis data, and qualitative content analysis" (p. 83). No other description of instrument validation was given. The KR-20 value for the exam with the experiential population was .81.

Twelve laboratory reports were graded by the class instructors using uniform guidelines. The six quizzes contained five short-answer questions written by the investigator and were graded by the instructors using specific guidelines.

A confusing battery of multivariate and univariate analyses were performed, with degrees of freedom in the tens of thousands! Laboratory sections should have been used as the unit of analysis; it is unclear what was used to produce the reported degrees of freedom. Instructor effect, treatment effect, and interactions were examined for all three measures. Individual statistics will not be reported here, in part because of the confusing nature of the report, and in part because the hypothesis of no difference stood unrejected despite all the massaging of the data.

This study suffered from the difficult-to-control instructor effect: How do you involve large numbers in a study and avoid strong interactions between the treatments and various instructors? The investigator did take care to insure that the treatments were distinct, and attempted to control for instructor effect by weekly meetings. This does not appear to have been sufficient.

The author contended that at least the ED approach did not hinder learning, and must have some value since it fosters discretionary thinking skills. Unfortunately, no meaningful instrument is available to assess this aspect. These conclusions must be taken as theoretical, not empirical, as thinking skills were not addressed by the study.

Hall and McCurdy (1990) also conducted a study involving the

Biological Sciences Curriculum Study (BSCS) style laboratory in introductory college biology courses. The investigation was designed to replicate and extend Leonard's (1983) comparison of BSCS-style laboratory format to a directive traditional college laboratory approach. In addition to assessing students' biology content understanding, they also examined reasoning ability and attitudes toward biology.

Hall and McCurdy's experiment was performed using 119 students from introductory general biology courses at two private, midwestern liberal arts colleges. The two schools had similar student populations, and the subjects of the study were said to be "heterogeneous with respect to ability level, prior science experiences, and socioeconomic background" (p. 627), and most were of typical college freshman and sophomore ages. No additional information was given regarding the subjects. It is difficult to imagine a private, midwestern liberal arts college with a truly heterogenous population, and therefore more information would be useful for the purpose of generalizing the results of this experiment. Additionally, since Piagetian reasoning levels were reported for the students, a more detailed description of the subjects' ages would have been of value.

The authors reported using a quasi-experimental nonequivalent control group design. The students self-assigned to laboratory sections and then sections were randomly assigned to one of the two treatments, with 60 subjects in the experimental group (BSCS-style) and 59 in the comparison group (directive traditional). There appears to have been one instructor at each college, one of whom was one of the investigators. No mention was made of instructor effect, nor the fact that the subjects were actually involved in two different courses at two different colleges. The number of subjects from each college is not specified by treatment or in total.

The treatments were very similar to those used by Leonard (1983). The experimental treatment was 12 BSCS-style laboratory activities designed by Leonard and described in the discussion of his 1983 study. The comparison treatment was 12 more directive (considered traditional) laboratory investigations matched to the BSCS activities for content as was done by Leonard (1983). The traditional laboratory activities were said to be similar to the commercially available program used by Leonard, but they appear to be unpublished. Hall and McCurdy stated that the comparison activities were judged by a panel of three biology professors "to be considerably more directive and less inquiry oriented than the BSCS-style" (p. 627). As with the Leonard's 1983 study, the instructors willingly assisted the students in the comparison (traditional) group and "modeled polite refusal or minimal redirection for the experimental [BSCS] group" (p. 629), however, classroom observations do not seem to have been performed for verification.

All subjects participated in one two-hour laboratory per week for the full semester. It is interesting to note that the BSCS-style activities were designed for 2 1/2 hours and were used in a three-hour period by Leonard (1988). As with the Leonard studies, the subjects at both schools in this investigation also attended accompanying biology lectures that amounted to about one-half of their in-class time.

Three different instruments were used to assess biology laboratory concepts, reasoning ability, and attitude towards biology. The 63-item multiple-choice Test on Biology Laboratory Concepts was developed by the researchers to measure student achievement in nine concept areas common to both treatment laboratory programs. The authors reported internal consistency with a coefficient alpha of .85. Content validity was established by three science education and biology professors.

Reasoning ability was measured by the Group Assessment of Logical Thinking (GALT). This 12-item multiple-choice instrument used line drawings of Piagetian problem situations to assess cognitive development measuring six different logical operations: conservation, proportional reasoning, controlling variables, combinatorial reasoning, probabilistic reasoning, and correlational reasoning. The original authors reported a .80 correlation for the classification of subjects with the GALT and using Piagetian interviews, and a coefficient alpha of .85. The GALT can be considered to have concurrent and construct validity and to demonstrate internal consistency.

Attitude toward biology was assessed using the Biology Student Behavior Inventory, a 39-item paper-and-pencil instrument with four subscales. The original author reported the following coefficient alpha values for each subscale: curiosity, .65; openness, .71; satisfaction, .66, and responsibility, .43. Content validity was determined by an undescribed panel of judges. The unacceptably low reliability values, along with questions regarding the appropriateness of the subscales to indicate overall attitude toward biology, make results from this instrument suspect.

A two-way analysis of covariance (ANCOVA) using pre- and posttest scores on the Test On Biology Laboratory Concepts revealed a significant difference in favor of the experimental group, $F = 4.07$, $p < .05$. Unfortunately, the student was incorrectly used as the unit of analysis. No mention was made of meeting the assumptions associated with ANCOVA.

An ANCOVA failed to reveal significant differences between groups in reasoning ability. It is noteworthy that prior to the study, 60% of the subjects operated below the Piagetian formal reasoning level, and that the number of formal thinkers in both groups increased by 15% during the semester. Maturation was clearly involved with this dimension. No significant differences were found between groups with respect to attitude.

Hall and McCurdy employed reasonable instrumentation to examine content understanding and reasoning ability, administered pretests, and randomized by sections. Students may have become test-wise from the pretests, however these tests controlled for differential selection and maturation--necessary since randomization by sections was inadequate to create equivalent groups. The setting effect (college and instructor) was statistically examined, and seems to have been controlled adequately. The major limitation to the findings of this study is the use of the incorrect unit of analysis.

DeLuca and Renner (1976) conducted an experiment to compare two methods of instruction in an introductory geology laboratory course. The course enrolled about 700 students per semester, who represented a cross section of undergraduate disciplines at the University of Oklahoma. Students attended three one-hour lectures and one three-hour laboratory per week, plus a required one-day field trip. Lectures were conducted by faculty; laboratories were conducted by graduate assistants. No other information was provided about general course design or the students. The background of the students and course structure appear to be similar to that of a nonmajors introductory biology course allowing some application of the results from this study to biology instruction.

The authors stated that a "randomized, 2 X 2 factorial design" (p. 308) was used. Two instructors each taught an experimental and a control group. The description of the randomization procedure lacks the detail necessary to insure that <u>students</u> were randomly assigned to each of the four cells: "Eighty-three students were randomly assigned to two instructors, two classes to each instructor" (p. 308). It appears that intact classes were used, and if this were the case, then a quasi-experimental design was actually used, and the correct unit of analysis for statistical procedures would have been the class instead of the student.

The treatment in this study was the method of instruction in laboratory: expository approach or structured inquiry. The Expository Approach represented the traditional geology laboratory and was designated the control group in the experimental design. The Expository Approach included an introductory lecture of about an hour and extensive instructor involvement answering questions throughout the laboratory period. The Structured Inquiry method began each period with a 10-15 minute introduction followed by an activity in which the students were guided by written procedures to make observations, perform manipulative tasks, and draw conclusions. Instructors responded to students' questions with guiding questions, not direct answer.

No reference was made to specific laboratory manuals used in either treatment; it appears that they relied on unpublished materials. No mention was made of training the instructors in either of the two laboratory instructional methods, or utilizing classroom observations to insure that treatments were applied correctly. Further, no attempt appears to have been made to insure that content coverage was equivalent in the two treatments. A fairly detailed description of one of the Structured Inquiry activities is provided, but no equivalent traditional laboratory activity is available for comparison. In summary, not enough information was provided to insure that the treatments were actually implemented as described.

The dependent variables measured were achievement in geology content, students' attitude towards their course, and self-esteem as a geology student. Because the last two variables do not reflect a student's scientific literacy, they will not be considered further in this discussion.

Detailed information was provided by the authors regarding the development of the achievement test in geology content. An original version of the test was administered and modified twice based on scores of 799 students and comments from faculty reviewers. It is not indicated how the reviewer determined content validity, but such a claim was made. It is possible that only face validity was determined. For this experiment, 60 five-choice objective questions were selected from the original 99-question test. All questions on the final version of the instrument were selected because they contributed to high content validity and showed a high ability to discriminate. The final test had a Spearman-Brown estimate of reliability of .89 and item difficulty range of .92 to .27 with an average of .55.

An analysis of variance was performed and tested at the .05 level using the student as the unit of analysis. No significant differences were found between groups for instructor or methods of instruction with regard to achievement in geology content. The authors' argument in favor of the Structured Inquiry approach is that at least they did not do worse than the traditional method. (It did produce significantly higher geology self-esteem scores and better attitudes towards the course, but there were some problems with the instruments used.)

The failure of the study by DeLuca and Renner to demonstrate significant differences in achievement may be because <u>content</u> is not the area of scientific understanding that was promoted by the Structured Inquiry approach. It would be valuable to know if measurable differences existed with regard to the science process skills or the understanding of the nature of science.

Repeated reports of unsatisfactory understanding of science process skills among American students prompted **Walkosz and Yeany** (1984) to investigate the effects of specific instructional emphasis on integrated science process skills in a college biology laboratory course. Because many of the integrated process skills seem to require formal operational thinking, cognitive development of the subjects was also examined.

Subjects for the study were enrolled in Biology 102 during the Spring and Summer of 1984--most likely at the University of Georgia since both authors were housed there, but not specified in the report. An experimental group (*n* = 127) and a comparison group (*n* = 107) were selected. The methods of selection and assignment to treatments were not described, but it is suspected that the groups encompassed all students enrolled during a given term. If this is the case, comparisons between groups are all but meaningless--not only were subjects not randomly assigned, but students enrolled in summer classes are a distinct subpopulation of all college students, and summer school itself is a treatment different from the regular school year. Analysis of the cognitive development data revealed no significant differences between the groups, with 31.5% of the subjects not classified as formal thinkers. No additional information about the subjects was given that could be used to infer equivalency of treatment groups or contribute to the generalizability of the results.

It was not specified whether the course served majors, nonmajors or both. Students attended four 50-60 minute lectures and one three-hour laboratory each week. One faculty member presented all lectures; an unspecified number of teaching assistants taught the laboratory exercises. No other information was provided regarding the course used for the study.

The comparison group followed traditional laboratory exercises that had descriptions of procedures and tables for recording data. They were not expected to identify variables, state hypotheses, interpret and predict from the results, or design experiments. The laboratory activities involved data collection and display only.

The experimental group followed the same laboratory procedures, but after the data were collected and displayed, they performed additional tasks to emphasize science process skills. They were required to describe, interpret, and predict from the results, and they were to choose other variables on the same topic and design an experiment to be discussed in their lab reports. No mention was made of the additional time required by the experimental group for completing the in-class activities or in preparing reports. The nature of the involvement of the teaching assistants was not discussed.

Cognitive development was measured using the 10-item *Test of Logical Thinking* (TOLT), a published instrument. Students were classified into five different levels of thinking ranging from *concrete* to *fully formal* based on their score on the TOLT. No additional information about the TOLT was presented.

Two equivalent forms of the *Test of Integrated Process Skills* (TIPS) were adapted from published instruments--one was used as a pretest, the other as the posttest. No mention of validity or reliability was made.

Content-specific lab quizzes were given during the lab period following each exercise. Reliabilities based on the Spearman-Brown Prophecy formula were reported to be greater than .73. Validity, nature of the questions, and length of each quiz were not described.

A full suite of statistical analyses were preformed to identify differences between groups and correlations between the many variables. Inferences that significant differences between the experimental and control groups are due solely to the treatment effect are invalid for reasons of nonequivalency previously discussed. In addition, the pretest scores were not used as a covariate, the statistical procedure was not specified, and no units of analysis were given. In short, the design of this study is so weak that any statistically significant differences seem meaningless so none of those procedures will be discussed. Some correlations and subjective comparisons of mean scores may be of value.

On the 40-point *Test of Integrated Process Skills*, the experimental group gained over six points while the comparison group gained less than 1.5. This would appear to be of practical significance and can be taken as an indication that further investigations may be warranted. An alternate explanation could be regression toward the mean since the experimental group started with a significantly lower score, $p < .001$.

Laboratory quiz score means were higher for the experimental group, but even if statistical significance could be demonstrated (as the authors claim), practical significance does not exist. Additional questions are raised here of possible bias in grading quizzes in favor of the experimental group since the nature of the quiz questions was not described.

Laboratory and TIPS scores for students from both treatment groups were displayed according to the five cognitive levels determined by the TOLT. In every case, higher cognitive levels had higher mean scores. Additionally, TOLT and pre-TIPS values were correlated with a Pearson Correlation Coefficient of .44. These results from Walkosz and Yeany offer no insight into the effectiveness of a specific instructional method, but does indicate that methods designed to advance cognitive development are good candidates for having a positive effect on science achievement.

Isom and Rowsey (1986) examined the effect of a Prelaboratory Preparatory Period (PLPP) on students' academic achievement in a freshman-level introductory chemistry course at Auburn University, Alabama. The subjects were 233 students enrolled in the course over four school quarters--it is unclear whether this was the total enrollment or a selected sample. No other information is provided about the subjects. It is not known if this class served majors, nonmajors, or both.

The authors stated that a posttest-only control group design was used with laboratory sections randomly assigned to the treatment (*n* = 5) or control (*n* = 3) groups. This study falls into the gray area of experimental design: There is some question as to whether the laboratory section or the student is actually the experimental unit, and if it is in fact the student, then this is an unacceptably weak quasi-experimental design with no mechanism to control for group differences. The authors vaguely stated that a one-way analysis of variance and Scheffe multiple comparisons were performed on each group's mean scores to "insure homogeneity and to verify randomization" (p. 232). They claimed that the groups were homogeneous, but it is unclear how they were able to reach this conclusion. The philosophical issue of whether the educational unit of interest is the individual student or the mean achievement of a class of students will not be debated here, because regardless of the outcome, the design of this study did not adequately establish equivalency of experimental groups and therefore lacked internal validity.

The outcome data consisted of students' grades on laboratory reports and quizzes for each of seven laboratory exercises. The quizzes, which contained a variety of test question styles, were validated by a panel of experts, and had Kuder-Richardson 21 values ranging from .78 to .84. It was not stated whether the quizzes were actually content validated, and the validation process was not described.

The PLPP treatment involved meeting with groups of 10-12 students one or two days prior to the laboratory activity. These meetings were about 45 minutes long, with the first 25 minutes spent briefing the students regarding the upcoming laboratory. The remainder of the time was devoted to student/instructor interaction relating the laboratory material to previous lectures. This procedure was felt to have several theoretical advantages: (a) Students should be more likely to ask questions and engage in discussion with their peers in smaller groups, and (b) students would have more time to ponder questions about the upcoming laboratory, and thus the PLPP could serve as an advanced organizer.

The control group received the traditional laboratory introduction consisting of a 20-minute lecture to the full lab section of 48 student immediately prior to the laboratory activity. No mention was made of total instructor contact time, but it appears that the treatment group met an additional 45 minutes per week. It was suggested that under the traditional laboratory format, students required a prohibitive amount of individual help outside of class time, and thus the PLPP design actually reduced the amount of time required for instruction. No data were provided comparing the experimental and control treatments along this dimension.

A significant difference at the .05 level was found in favor of the PLPP treatment based on a Wilks Lambda Omnibus *F*-test and subsequently comparing group means. It appeared that the sections were correctly used as the unit of analysis, but the exact statistical procedures were unclear. A Univariate *F*-test was used to compare control and treatment groups scores on each laboratory activity. This test revealed a significant difference for only one of the seven activities. The authors stated that the data illustrate that "less abstract concepts are presented more effectively via the traditional ... laboratory lecture while more unfamiliar abstract concept or exercise requiring good laboratory technique were more effectively presented by the Prelaboratory Preparation Period" (p. 235). It is difficult to critically evaluate these conclusions since little information was given regarding the abstractness of the various laboratories.

The Isom and Rowsey study seemed to be based on sound theoretical justification, but is riddled with design limitations. Questions of group equivalency are raised by the lack of a pretest and randomization of treatments by section instead of by student. No mention was made of the number of instructors or efforts to control for any instructor effect. The sample was not described, making generalization impractical. The Hawthorne effect may also be at play here, however, the extra attention is the treatment so there is no way to avoid it. And finally, as mentioned, the amount of time required for the treatment was not balanced with the control.

The use of student-authored interactive video disc presentations in a nonmajors biology laboratory course was examined by **Ebert-Zawasky and Abegg** (1990). Sixty-six students self-assigned to laboratory sections, then two sections were randomly assigned to the experimental group and one section to the control group. To verify group equivalency, SAT scores and Group Assessment of Logical Thinking (GALT) scores were compared and no significance differences were found (no statistical information was provided). This study followed the quasi-experimental nonequivalent control group design since a pretest was given. No description was given of the subjects which would contribute to the generalizability of the results.

Students receiving the experimental treatment had the opportunity to author and present one lesson using a computer interfaced videodisc system. During the first week of laboratory the instructor presented a video disc lesson to the class and explained how the program was constructed. Working in groups of three, students selected topics from the syllabus and constructed similar interactive lessons. Each group presented their own lesson and participated in seven other lessons created by their classmates.

The control group participated in nine video disc lessons authored by a researcher and presented by the instructor. To compensate for the group project in the experimental treatment, students in the control group, working in groups of three, wrote and presented a research report, and as with the video disc project, all three group members received the same grade.

A pre- and posttest were apparently given to assess biology content acquisition and no significant differences were found between treatment groups. Unbelievably, no other information is given about the instrument or the statistical analysis.

Large amounts of data were apparently gathered, however it doesn't seem to have been used to add to our understanding of instructing nonmajor biology students. The authors stated that "age, locus of control orientation, math SAT scores, number of biology courses and computer experience appeared to have no detectable effect on student performance."

Responses to a questionnaire about the video disc authoring experience were largely positive and the majority of the students said that they would recommend that the assignment be retained for future classes. No other information was provided about the questionnaire or students' responses.

The study by Ebert-Zawasky and Abegg was so poorly designed and sketchily reported that no evidence of teaching method effectiveness can be inferred. They did provide anecdotal evidence that introductory level nonmajor biology students can successfully create and present interactive video disc lessons to their peers.

Small-Group Discussions

Since the 1980s, strategies for using cooperative groups in educational settings have been defined, popularized, and researched. Two authoritative reviews of research on college teaching concluded that discussions and lectures were equally effective in teaching content, and that discussions were somewhat better at promoting problem-solving abilities and changes in the affective domain (Kulik and Kulik, 1979; Dunkin and Barnes, 1986). It is worth noting that both of these reviews rely on meta-analyses performed prior to 1980, and little attention was given to the structure of the discussion groups at that time. The two papers reviewed in this section are recent works and reflect the current foci of research on cooperative groups used to promote conceptual change.

Scharmann (1989) investigated the influence of small discussion groups to overcome misconceptions held by college freshman biology students regarding the nature of scientific theory using evolution as an example. The subjects were enrolled in one of two concurrent general biology classes taught during a three-week summer session. The groups were selected based on the willingness of the two instructors to participate in the study. The author stated that a nonequivalent control group design was used because the two classes represented intact groups. To minimize the instructor effect, both instructors agreed to use the same course outline. No other strategy was described for controlling instructor effect or to insure fidelity of treatments.

The experimental group (*n* = 13) and the control group (*n* = 17) were said to differ along only one dimension: After an introductory lecture on evolution, during the second week of instruction for both classes, the experimental group was provided an opportunity to discuss their positions regarding evolution.

The investigator provided the students in the experimental group with a set of four questions regarding evolutionary theory versus creation origins. The students were asked to individually write responses to the questions. They were then randomly assigned to groups of three or four for discussion. Group members were asked to share their written responses and resolve conflicting opinions. The investigator then provided an interactive lecture/discussion to resolve misconceptions arising from the small group discussions.

Students in both the control and the experimental groups were given a 35 question pretest/posttest covering attitude towards evolution (5 questions), an understanding of the nature of scientific theory (20 questions), and knowledge of evolutionary content (10 questions). All items were of a five-point Likert-type format. An untitled published instrument was used for the first 25 questions; the final 10 items on evolutionary content were written by the investigator to assess instructor differences. The authors of the published instrument reported internal consistency reliabilities of .78 and .77 for the two parts of their instrument. These reliability measures are from a sample of 1,812 undergraduate students from 34 higher education institutions. Validity was established using the known group difference technique, and it was reported that "the instrument discriminated an acceptance of evolution as a function of a progressive understanding of science" (p. 4 - 5). No other information regarding validity or reliability was presented. Most notably missing was any information on the 10 questions added by the investigator. No sample questions were provided.

The posttest was administered at the end of the three-week summer term of instruction; it is not stated when the pretest was given. Nonparametric statistical techniques were used because the control group pretest scores were not normally distributed. Using the Mann-Whitney *U*-test and Wilcoxon test on pretest scores, no significant differences were found to exist between groups or within groups with respect to evolutionary content understanding, attitude toward evolution, or an understanding of the nature of science. A between group repeated measures test was not performed. A between group analysis of posttest scores using Mann-Whitney U-test found the experimental group to possess a significantly greater combined understanding of the nature of science and attitudes towards evolution, $U = 1.75$, $p < .05$. There was no significant difference found for evolutionary content items.

Within group analysis of the posttest using the Wilcoxon test for repeated measures found that the control group ($Z = 2.33$, $p < .01$) and experimental group ($Z = 2.98$; $p < .001$) both exhibited an increased understanding of the nature of science and acceptance of evolution. There was no significant difference in an understanding of evolutionary content for either group from the pretest to the posttest.

Scharmann concluded that both a traditional lecturing technique and a diversified instruction strategy using small discussion groups were effective in presenting evolutionary concepts. He failed to address the fact that no significant difference was found on the pre- and posttest for either group with regard to content understanding. The author further concluded that both provide a basis for student growth in understanding the nature of science and an acceptance of evolution as an organizing theme of biology, however, based on the between group analysis, a diversified instructional strategy was superior to the traditional lecture method.

The author failed to discuss any of the limitations of this study. There was the issue of the control group and experimental group being taught by different instructors: The author seems to feel that this had been addressed by asking the two individuals to teach from the same outline and by demonstrating that both classes have similar understanding of evolutionary content at the end of the course. It is difficult to believe that the one discussion session directed by the investigator is the only difference in treatment that the two groups received during the three-week course.

There is a strong possibility that a Hawthorne effect is occurring in this study. The investigator, apparently not a regular instructor of the course, visited the experimental group and led them in a discussion activity. The control group received no such special treatment. It was not discussed how the control group used the class time that was allotted to the treatment for the experimental group.

Scharmann's study also suffers from lack of magnitude both in duration and number of subjects. Two groups with a total of 30 students seems quite small to reveal meaningful differences in attitude changes considering they probably entered the study with diverse views, and the attempt was a directional one. Further, it is questionable that one discussion session was enough to test the effect of the teaching method. An additional limitation was the lack of information regarding the content questions on the evaluation instrument.

The effect of cooperative group work on conceptual change in a community college chemistry course was examined by **Basili and Sanford** (1991). The study used quantitative and qualitative methods to examine multiple aspects of group behavior related to conceptual change theory. Only the portion of the study which evaluated the effectiveness of the small group method for promoting conceptual change is of interest here, so the following discussion will be limited to these methods and results.

Four intact sections (62 students) were divided among the control and experimental conditions, with each of two instructors teaching one of each treatment. The sections were "heterogenous with respect to sex, age (from late teens to 40s), race (white, black, hispanic, and middle eastern), and previous experience with chemistry (never has a course, had high school chemistry, and had failed in college general chemistry)" (p. 295). Subjects were enrolled in a two-credit non-laboratory course at a suburban community college intended to prepare students for college-level general chemistry.

The authors stated that a pretest-posttest control group experimental design was used, but since intact classes were used the design is actually the weaker quasi-experimental nonequivalent control group design.

The experimental treatment involved placing the subjects in cooperative groups of three or four students on a regular basis so that they could discuss thought questions and concept maps, and hopefully engage in behavior conducive to altering misconceptions. The pattern for the course was five 50-minute class periods of regular lecture and discussion, one period of group work, followed by an exam day. The control group experienced the same treatment except that instead of group work they were given a demonstration and required to write it up for credit.

The target concepts for the study were the laws of conservation of matter and energy, and the particulate nature of gases, liquids, and solids. The authors stated that faulty understanding of these concepts has been implicated in difficulties in learning biology as well as chemistry.

An instrument for assessing and categorizing conceptual change was constructed along published guidelines. For the conservation laws, the test involved true-or-false/explain-your-answer type question. For the states of matter, students were required to draw dots representing particles in a flask. The tests were piloted and revised with chemistry students and in-service science teachers. Validity, probably only face, was determined by a group of science faculty. Answers were rated on a five step scale from no conception to correct concept. After four trials, intercoder agreement of 93% was achieved, and one coder scored the remainder of the tests.

A pretest was given to all students at the beginning of the class. The laws of conservation posttest was given at the end of the third cycle, the particulate nature posttest was given at the end of the fifth cycle. It is assumed that the pretest was identical to the combined posttests.

Because of cell size requirements for chi-square analysis, data were analyzed by placing students into two categories based on whether they held misconceptions. A concern here is that "I don't know" fell into the same category as having a complete concept understanding. Based on the chi-square analysis of the pretest, groups were equivalent, and continued to be even after

16 students dropped the course. Posttest results indicated that the experimental group held significantly fewer misconceptions than the control group for four of the five concept areas, $p < .05$. The exception was the particulate nature of gases. Statistical analysis could not be performed on the number students exhibiting correct concepts, but the experimental group exceeded the control group in all topic areas. The values are percentages, and the first number in each pair is for the experimental group: Matter (22, 9), Energy (44, 23), Gases (58, 40), Liquids (31, 5), and Solids (39, 0). It is surprising to note the low level of complete concept development after instruction.

Basili and Sanford concluded that the cooperative small group interactions were effective in reducing misconceptions commonly held by introductory college chemistry students. The data seemed to indicate that this is the case. One of the limits to this study involved possible instrumentation biases in favor of those who have practiced discussing concepts using thought questions and concept maps (although the authors questioned the value of concept maps for concept change). Another limit is the small sample size which precluded the use of meaningful statistical analysis.

Individually-Paced Modular Instruction

A Personalized System of Instruction (PSI) was developed by Keller in the mid-1960s and was the subject of much educational research during the 1970s (Dunkin & Barnes, 1986; Gifford &

Vicks, 1982; Kulik & Kulik, 1979). This body of research dealt with an array of academic subjects and led to conclusions that PSI was very effective at promoting learning. The three research articles reviewed here deal more specifically with instruction in introductory biology courses.

Robinson and Shrum (1977) used a one group pretest-posttest experimental design to examine the effectiveness of instructional modules combined with small group discussions in achieving the objectives of a college general biology course. Twelve activity-centered modules were designed for use in the first term of a two-quarter general biology sequence at Albany State College, Albany, Georgia, during the Fall of 1974. Each module contained a pretest, behavioral objectives, enabling activities (including readings, investigations, and various audio-visual media), and a posttest. Students were required to demonstrate mastery of a module by scoring 80% on the posttest before progressing to the next module. Small group discussions were held after completion of modules 4, 8, and 12.

Thirty students were randomly selected from the pool of general biology classes that met first period and assigned to an experimental class. Thirteen were females. The size and general characteristics of the population from which the sample was selected were not given. No information regarding age, academic major, general ability, or socio-economic status of the sample was provided. The use of the one-group design does not seem justified for this experiment since there was apparently a control group available in the other general biology students not selected for the experimental class.

Three different pretests were administered during the first class meeting and the identical tests were used at the completion of the course as posttests. The authors made note that instrument decay was controlled since no changes were made in the measuring devices, however the possibility of pretest sensitivity was not discussed. Another threat to internal validity not mentioned is the possibility of test burn-out from administering all instruments during the same class session.

To determine if differences existed in students' performance on the course behavioral objectives as a result of the having completed the course modules, the *Course Criterion Test* was constructed using 58 multiple-choice items from a pool of questions designed to address the objectives. Test reliabilities (method unspecified) were .38 for the pretest administration and .94 for the posttest--no explanation was given for the unacceptably low pretest reliability value. Respective difficulty levels were .36 and .57. Item discrimination indices were reported to range from .30 to .80. No additional information was provide regarding the validity, reliability, or construction of this instrument.

Two published instruments were also used as pre/posttests: the *Welch Science Process Inventory* (SPI) Form D and the *Subject Preference Scale*. The SPI was used to assess understanding of methods and processes by which scientific knowledge evolves; the preference scale was used to examine attitudes toward biology. Despite the fact that these are both published instruments, no mention was made of validity or reliability reported in the literature. Reliability values (method unspecified) obtained from pretests and posttest in this study were: SPI (135 items), .45 and .74; preference scale (45 items), .32 and .42. No additional information was provided regarding the validity, reliability, or construction of these instruments.

Data were analyzed using the correlated *t* test for the difference between two means, the recommended statistical procedure for this experimental design if the assumption of normal distribution of scores were met, however this distribution was not described. For the *Course Criterion Test* a *t*-value of 2.83 was reported; for the SPI, a *t*-value of 7.17, both with 26 degrees of freedom and both significant at the .01 level.

The practical significance of these results are less clear. The criterion test was designed to be a direct measure of the course objectives, however the mean score on the posttest was only 32.82 out of 58 items, with just a 12 point gain from the pretest. These data indicated that the course objectives were not met! If students were required to achieve mastery on the modules with an 80% correct response rate, a mean score on the final test of 56.6% seems to indicate a failure to meet the objectives. The mean SPI scores are similar: Out of 135 items, 70.55 on the pretest, 88.77 on the posttest. The posttest scores are so low on both the criterion and SPI instruments that the success of the treatment, despite the statistical differences between the pretest scores, is very questionable.

The data reported for the *Subject Preference Scale* in the text do not match the values in the corresponding table, but neither indicate a significant difference in attitudes as a result of the treatment. It was noted by the authors that the low reliability of the preference scale may be responsible for it not revealing any attitude difference.

The experiment by Robinson and Shrum offers little information of value regarding the teaching of college biology. A control group was not used, so comparisons can not be made between the effectiveness of the modular/discussion method and other methods of instruction. As discussed earlier, there does not appear to be any justifiable reason for using the weak One-Group design since control subjects were available and the nature of the treatment did not preclude their use. Descriptions of the instrumentation omitted information that would have added credibility to the data, and when reported, several reliability values were unacceptably low. And finally, the results, although statistically significant, are not of a magnitude to be of practical significance.

Langley and Bowman (1981) compared a self-paced field-oriented audio-tutorial with an illustrated lecture format for instruction in ecological concepts. Subjects were 417 introductory biology students at Wichita State University: seven classes of biology majors and four of nonmajors. No other information was provided about the students involved.

Nine of the 11 classes used the field audio-tutorial (A-T) method. The other two classes, one each of nonmajors and first-semester majors, received classroom lectures. No mention was made as to how treatments were determined. Subjects from different classes were combined into representative categories because there was no statistical difference in their test scores. Categories used were: Nonmajors A-T (174), First-Semester Majors A-T (111), Second-Semester Majors A-T (74), Nonmajors Lecture (40), and First-Semester Majors Lecture (18). It was not stated if all classes met during the same school term.

The Audio-Tutorial treatment used two instructional modules each requiring one to two hours. Study guides for each module listed key words, objectives, illustrations, questions, and activities to be performed at stations marked on a campus map. Accompanying audio cassette tapes provided general directions, suggestions for observations at each station, and possible answers to questions in the study guide. Students were given two weeks to complete the activities by checking out the materials from an independent study lab. No mention was made of any other contact with faculty or tutors during the treatment period.

Lectures were given by the two authors in their respective classes. No other information was provided regarding the lecture treatment which leaves several questions unanswered: How did instructional time compare for the two groups? How did content coverage compare? What type of student involvement was fostered by the lectures? Additionally, was the two-week A-T treatment a novel approach used in classes typically taught by the lecture method. If so, a Hawthorne effect is surely involved.

To assess the treatment effects, pre- and posttests containing ten objective questions were used. It is stated that "when the order of administrating the pre- and posttest versions was changed, no differences in student performance resulted" (p. 237), however no coefficient of equivalency was provided. The criteria for establishing no difference is not given, but assuming it was met, the instrument can be said to have alternate form reliability. No mention was made of validating the instrument and no information was provided about the administration of the tests.

Individual pre- and posttest scores were compared using a Wilcoxon matched-pairs sign test, and differences in distributions of scores between classes were compared using a median test. These nonparametric statistical tests (designed for use when the assumptions of more rigorous tests are not met) were apparently used to accommodate for the lack of randomization or any effort to establish equivalency of groups.

The authors report that both methods of instruction were effective as students in all 11 classes significantly increased their posttest score over their pretest as shown by the Wilcoxon Test, $p < .001$, for each class. This is a curious claim since that data table indicated that no pretest data were collected for the four classes of First-Semester Majors in the A-T treatment! Using a median test ($p < .05$), no statistical differences were found between scores for the Second-Semester A-T classes and the First-Semester Lecture class. (There was no Second-Semester Lecture group, and no pretest for the First-Semester A-T group, so this was the only comparison that could be made for majors.) Likewise, no difference was found when comparing classes of nonmajors exposed to the two different instructional formats.

The recommendation of the authors is the familiar: The novel instructional method didn't hurt learning, and the students seemed to have some positive attitudinal shifts, so we recommend using it. These recommendations don't seem to be warranted by the data gathered.

Gifford and Vicks (1982) examined the effectiveness of a Personalized System of Instruction (PSI) for introductory biology at a junior college serving a predominantly rural, black population. This study was designed to address the specific student population at a private school in northeast Mississippi because these students were believed to differ in learning styles, abilities, and motivation from those participating in prior studies which showed general success of PSI. The conclusions of this study are not to be applied to the general population of college students, but serve to extend research results to more specific minority populations that may not be generalizable from broader studies.

From the Spring semester of 1978 freshman population of 302 students, two intact classes of 40 students were randomly selected and assigned to either the experimental or control groups. Nonsignificant pretest differences between groups were controlled by using covariance statistical analyses. An analysis failed to reveal significant differences in motivational factors between the two groups. Data were also gathered for all subjects relating to age, sex, family income, family size, college grade point average (GPA), and California Achievement Test composite score (CAT). A regression analysis did show that GPA and CAT were significantly related to biology achievement, but this information was not used in any way to insure equivalent groups. None of the other data collected regarding the characteristics of the subjects was provided. This study follows the non-equivalent control group design, and is weakened by the failure to adequately insure that the two groups did not differ significantly in all meaningful attributes.

The experimental group was taught biological science for 12 weeks using the PSI method following the Keller plan. Students worked individually on small unit modules containing activities, readings, study-questions, and filmstrips. Proctors were available to tutor and evaluate unit tests. Once students demonstrated mastery of the material in a unit, they progressed to the next unit.

The control group was taught by the lecture method and was said to have met the same number of days per week as the experimental group. There was no indication that instructional contact time was controlled for the experimental group, so the two treatments may have differed in that variable. The two treatment groups were reported to have used the same textbook and to have covered the same topics. Depth and breath of content coverage was not mentioned.

The Nelson Biology Test, revised edition, Form E was used as a pre- and posttest to measure biology achievement. This published instrument was recommended as the best available for research applications and has reported reliabilities, *r*, in the range of .89 - .92 (type of reliability was unclear). No reliability values were established for this specific administration of the test. The content validity of this instrument for the particular course in this study was not reported, however it is said to be appropriate for high school and elementary college levels.

A significant difference, favoring the PSI method, was revealed by comparing the pre- and posttest means of the two groups. The method of statistical analysis is not specified, but it appears to be Analysis of Covariance using the pretest as the covariate. None of the assumptions implied by the selection of this method were discussed. The PSI group had a pretest mean of 11.65 and a posttest mean of 34.93; the lecture group had pre- and posttest means of 13.55 and 27.37, $F(1, 75) = 15.77$, $p < .01$. More meaningful than the statistically significant difference is the practical significance of the magnitude of the difference. Examination of the mean test scores reveals what must surely be results of practical significance in favor of the PSI method.

The study by Gifford and Vicks used the quasi-experimental non-equivalent control-group design with limited information reported about the subjects, thus weakening the conclusions that can be drawn. However, the treatments were of sufficient duration and qualitatively different to contribute to variability in outcomes, and good instrumentation appears to have been used despite the sketchy information provided. The results of this study indicated that PSI is an effective method of instruction for freshmen college biology students from a rural black population. Generalizing these results beyond that population was not the intention of the study, and would not be appropriate based on this research alone.

Written Materials to Enhance Instruction

The three papers in this final section all examine the use of written materials to enhance instruction in lower-division science courses. The first paper attempted to answer the common question posed by students: "Why don't you just give us typed notes for the lecture," and those students probably won't be happy with the answer! The other two papers examine the effects of increased student involvement through the use of adjunct questions keyed to the textbook and journal writing.

The effect of instructor-prepared handout materials on learning from lecture instruction was examined by **Petrich and Montague** (1981) in college chemistry classes. Three intact classes were randomly assigned to the three treatments: One group received verbatim scripts of the lectures, a second class was provided with an outline containing "all pertinent information" (p. 180), and a third received no aids. Material was presented using video taped lectures followed by a 10-minute question and answer period. Lecture aids were given to the students one class meeting prior to the lecture, and they were instructed to read them before coming to class.

The three classes used for this study had a total of 54 students who were enrolled in the lecture sections of a freshman-level chemistry course at San Antonio College. No additional information is provided about the students or the course. Two points of particular note are the class size of only 18 students and the lack of information regarding the students' major area of study.

The authors correctly stated that a quasi-experimental design was used with the intact classes. A pretest was administered to all three groups prior to the presentation of a series of three video taped lectures. An achievement test was given the next class meeting following each lecture presentation for a total of three posttests taking about an hour each.

The four instruments were prepared by the investigators with five-response multiple-choice questions. The pretest consisted of 55 questions tied to each of the three lectures: 23 for the first topic (PRE1), 10 for the second (PRE2), and 22 for the third (PRE3). The posttests each contained 20 questions. Content validity was determined for each instrument "by comparison with predetermined prerequisite abilities and behavioral objectives" (p. 180). The authors stated that criterion-related validity and construct validity were not determined.

Internal consistency reliabilities were measured by the split-half and Kuder-Richardson methods, producing comparable, but not identical values. Kuder-Richardson values reported were: Pretest, .84; PRE1, .77; PRE2, .62; PRE3, .74; Topic 1, .73; Topic 2, .78; and Topic 3, .89.

Since the students were not randomly assigned to the treatment groups, analysis of covariance was used to test the hypothesis of no difference in learning by students in the three groups: Script, Outline, and No Aids. The analysis was performed for each of the three posttests comparing all possible pairs of classes. The covariate was determined by examining total pretest scores, topic pretest subscores, and ACT scores for high correlation with the dependent variable (posttest score) and absence of interaction with treatment factors. For the Topic 1, the best covariate was PRE1 with correlations of .69 for No Aids, .56 for Script, and .08 for Outline. (This last value seems outrageously low, and there is suspicion of a printing error.) For Topics 2 and 3, the best choice of a covariate was ACT scores with correlations for the No Aids, Script, and Outline treatments of .78, .40 and .61; and .80, .72, and .62, respectively. No significant interactions were found for any of the covariates with the treatment groups. In this procedure, the student is correctly used as the unit of analysis since the covariate is designed to compensate for the lack of randomization.

The analysis of covariance revealed that for Topic 1, a significant difference was found between the No Aids and Outline groups, with the No Aids performing better, $p < .01$. No significant differences were found between No Aids and Script or between Script and Outline. For Topic 2, all pairs of classes showed significant differences, $p < .05$. Analyses for Topic 3 revealed a significant difference at the .05 level between the No Aids and Script groups only. Based on this analysis, the hypothesis of no difference between groups was rejected.

Mean unadjusted scores for all three topic tests followed the same pattern: No Aids had the highest scores, Outline had the lowest, with Script scoring intermediately. The treatment/class mean scores for each topic are as follows: Topic 1 -- 14.12, 12.67, 11.17; Topic 2 -- 10.94, 8.17, 5.74; Topic 3 -- 13.88, 10.94, 9.17. Examination of these values seems to indicate that the differences revealed between treatments are of practical significance. However, it must also be considered that these are unadjusted means and the groups were not assigned randomly, so the apparent differences may be the result of some underlying difference between the groups not related to the treatment.

Additional analyses were performed to determine if there was an interaction between treatment and students' prior level of knowledge or ability as indicated by pretest scores and ACT scores. No interaction was found.

Petrich and Montague's study was conducted the last two weeks of the semester; no description was given regarding the instructional procedures used in the classes up to this point. It is possible that the video taped lectures were a novelty as could be the tests every other class period. Although these novelties were experienced by all three treatment groups, it is very likely that the subjects of this study were altering their behavior as a result of participation in the experiment. Despite the limitations of this study (quasi-experimental design, small sample size, novelty effect, and weak statistical analysis), the results indicate that instructor-prepared notes do not promote learning and may in fact impede it.

Spring, Sassenrath, and Ketellapper (1986) examined the efficacy of using adjunct questions with the textbook readings for a nonmajors biology class. A static-group comparison design was used with a treatment crossover, so that during the first half of the study, Group A received the adjunct questions, and during the second half, Group B received the questions. This design was selected primarily to be fair to the students who were all enrolled in the same class and receiving treatments which may affect their grades.

Subjects for the study were enrolled in a one-quarter introductory biology course for nonmajors at the University of California at Davis. Intact discussion sections of about 20 students were randomly assigned to one of the two groups resulting in 97 students in Group A and 86 students in Group B. The groups were reported to be matched on sex, class standing, and ability; mean SAT verbal and mathematical scores were not statistically different for the two groups.

Based on prior research, the adjunct questions were designed to cover one or two paragraphs of reading and required very short written responses. About two questions per page of text were written by two graduate students and edited by the course instructor. The resulting 800 questions were printed in a workbook. About 30% of the questions were verbatim recall, 45% were paraphrased from the text, and 25% dealt with applications not specifically in the text.

Students in the experimental group were instructed to read and answer the questions immediately after reading the appropriate section of the text since this had previously been shown to be the most effective use of adjunct questions. Compliance with these instructions was determined by a questionnaire, which also determined that students in the control group did not obtain questions from students in the experimental group. Students in the experimental group were required to submit their answers for the adjunct questions on a weekly basis, and these answers were not returned to them. Experimental group subjects who did not answer the questions were eliminated from the study, and then the group was subdivided into those who followed the instruction and those who did not.

To determine group differences, scores were extracted from the multiple-choice midterm and final examinations for all questions related to the assigned textbook readings. These scores were further subdivided to create five criteria for comparison. First, the questions were classified as verbatim-recall or comprehension depending on whether the wording of the text was retained. Comprehension questions included all items that paraphrased textbook material or required applications not covered in the text. Secondly, all the text-related questions were classified as either new or old depending on whether the test item had been covered previously by an adjunct question. No mention was made of validity, reliability, or difficulty indices for any of the measures used. The numbers of questions included in each measure were:

	Midterm	Final
Text total	46	49
Old	21	25
New	25	24
Verbatim-recall	17	17
Comprehension	29	32

Results from the midterm scores were analyzed *assuming no differences* between groups: Experimental 1 (n = 60, directions followed), Experimental 2 (n = 25, directions not followed) and Control (n = 77). However, comparisons of SAT scores and scores on the lecture portion of the midterm test revealed no significant difference at $p > .20$--a probability which makes the claim of no difference in ability among groups extremely weak.

It appears that an *F*-test followed by a Newman-Keuls test were performed on the five different midterm measures, however specific statistics and probability values are vague. It was reported that no significant differences could be found between the control group and the experimental group that followed the instructions, however both of these groups scored significantly higher than Experimental 2 on several of the measures. The authors provide lengthy discussions about the meaning of these results in terms of the dangers of reading the adjunct question before reading the text; they fail to address the possibility that those who did not follow the directions may differ in some way that affected their test scores.

Treatment groups were switched for the second half of the course and effects evaluated using the final exam. Unlike the midterm examination, test scores for those not following directions were not significantly different from the rest of the experimental group, so the experimental group was kept intact for comparison to the control group. An analysis of variance of the experimental and control groups' scores revealed significant difference on all five measures, $F(1,181)$, $p < .05$. This is judged to be the correct unit of analysis for this study for two reasons: (a) Despite the fact that intact classes were assigned to the treatment groups, the treatment acted on the individual and does not seem to be related in any way to the discussion class units, and (b) satisfactory evidence exists to indicate that the distribution of students among the different classes represents the overall distribution. These results were said to be of practical significance since the experimental group scored about five percentage points higher on all types of questions.

No explanation was given to rectify the positive results for the second half of the experiment and with the failure to show differences during the first part. An analysis was performed to show that the control group for the second half performed no worse on the final exam than students in previous years, and the experimental group performed significantly better, $p = .015$, one-tailed test. These results rule out the possibility that participating in the experimental group for the first half of the class had some detrimental effect on their performance during the second half which caused them to be out-scored by the group using the adjunct questions.

The quasi-experimental static-group comparison design limits the strength of the research conclusions in that subjects were not randomly assigned to groups and no pretest was administered, so differences in posttest results can only be weakly inferred to have resulted from the treatment. The threat to internal validity from non-equivalent groups is, however, not exceedingly great in this study because the sample size was large, several intact groups were randomly assigned to each treatment, and the similarity of group characteristics was established.

The careful design of the treatment based on prior research and the use of questionnaires to determine the extent to which the experimental and control subjects followed the design provide assurance that experimental diffusion did not occur and contributed to the validity of this study. The crossover of treatment groups, while implemented for reasons of educational fairness, may have served to ameliorate threats to validity such as the Hawthorne Effect. A major threat to internal validity is the failure to account for the differences in study time that may have existed as a result of assignment to one of the treatments. It may have been the increased academic engagement time required to answer the adjunct questions that led to the significant differences on the final exam, and not the specific activity of writing the answers to the questions. And finally, since no reliability or validity information was provided for the instruments (class exams), results revealed by the study must be taken as tentative until such information is provided.

Trombulak and Sheldon (1989) looked at the effect of journal writing on college biology students' grades and attitudes about science courses. The experimental subjects were the students in freshman-level general ecology (n = 77) and sophomore-level vertebrate biology (n = 25) at Middlebury College in Vermont. General ecology was required of biology majors and satisfied a general distribution requirement of the college; vertebrate biology satisfied no specific course requirements in biology or for the college. It was not stated to what degree nonscience majors might be involved in either of these courses. No demographic information was provided for the subjects or the college as a whole.

It appears that a pretest-posttest control group design was used for attitudes, and posttest-only for content learning, however there is some question regarding randomization. The authors stated that each class was divided into two groups of about equal number and matched according to sex, grade on the midterm, and whether they were enrolled in the accompanying lab. The remainder of the study does not reveal any reason for the students to be matched after assignment to groups, so it is most likely that matching was an attempt to create equivalent groups. It was never stated that matched pair were randomly split into the treatment or control groups. It is possible that this study lacks the necessary randomization and is therefore following the weaker non-equivalent control group and static group comparison designs.

The attitudes assessment instrument was a Likert Scale survey with seven questions about various aspects of the course. The only information regarding this instrument was a copy of the survey: no reliability, validity, or source information were provided.

The effect of the treatment on learning was measured by using the students' letter grades in the course. No information was provided to support the validity of this measure, nor was it explained how the grades were derived.

The students in the treatment groups were asked to write for five minutes about their biology lecture at some time during the day. They were provided with spiral notebooks and given suggested journal topics at the end of each lecture. These journals were not handed in and were done strictly on a voluntary basis. Students in the control group were told that they were free to do the writing assignments since the ethics of withholding a learning opportunity could have been an issue. Only one person in a nonwriting group did any writing. About one-third of the ecology writing group did no writing at all; every subject in the vertebrate writing group made some journal entries.

Letter grades were compared using the Mann-Whitney *U*-test. It was reported that the vertebrate biology writing group performed better than the comparison group by two-thirds of a grade, $p = 0.038$. No significant differences were found in the ecology class between those in the writing group and the comparison group, nor between those who actually wrote in their journals and those who did not.

The attitude survey did not reveal any attitude differences between groups on the pretest nor on the posttest, and no measurable changes occurred within groups during the study. Good attitudes were reported throughout the study with average responses to all questions between agree and strongly agree. It seems likely that this survey with no reported validity or reliability was unable to assess true attitudes.

The only significant differences found by Trombulak and Sheldon with regard to journal writing effectiveness was in letter grades of students in a sophomore-level elective biology course, and this had several limitations. Questions needed to be answered regarding assignment of subjects to treatments. The components of the grade needed to be described and some indication of the reliability and validity of the components needed to be established. A stronger analysis involving a meaningful pretest and/or clear establishment of equivalent groups would provide stronger evidence on the effectiveness of journal writing to enhance learning.

DISCUSSION

No study was reviewed which produced strong evidence regarding effective methods for teaching college biology to nonscience majors. Several studies found no significant differences between tested methods, most suffered from design problems, and the instrumentation used typically lacked content validity and/or had low reliability values. Any conclusions offered must be taken as tentative and suspect.

As indicated by McIntosh and Caprio's (1990) survey and the nature of the studies reviewed, increased student involvement in the learning process is believed to be important. Several different methods to this effect have been tried and tested including classroom discussions, inquiry-oriented laboratories, cooperative group work, individualized instruction, and written assignments.

The use of a classroom discussion technique instead of an instructor-controlled lecture was the focus of several studies. Barnard (1942) and Mason (1952) both conducted studies with large classes using interactive techniques to involve students with the intention of promoting scientific thinking. Barnard provided some early evidence that discussions could enhance thinking while remaining equally as effective as lectures at conveying content information. Mason's study revealed no difference between lecture and discussion-based methods for instruction in content or in promoting thinking skills. These two early studies produced results which remained consistent with more recent reviews of the educational literature (Dunkin & Barnes, 1986).

More recent papers examining the use of instructor-directed discussions also indicated the possibility of an advantage in using such methods, but if this advantage was found, it was not large and the research evidence is weak. Moll and Allen (1982) produced evidence that short video segments followed by a guided discussion could improve biology content learning and critical thinking skills, but their study design was very weak.

Haukoos and Penick (1983) failed to find any significant difference in biology content achievement when comparing directed lectures and laboratories with a less directed classroom climate including informal discussions. Performance on a test of science process skills was significantly improved, yet not at a level of practical significance. No content pretest was used in this study and statistical errors were made limiting the value of the results.

Isom and Rowsey (1986) found a significant difference in favor of using prelaboratory discussions with groups of 10-12 as opposed to a lecture introduction at the beginning of the lab period. Their study, however, lacked a pretest and had poor randomization. They also concluded that the lecture was equally effective if concrete information was presented.

Based on the research reviewed, weak evidence exists in favor of classroom discussions instead of formal lecturing, and the advantage is more likely to be found in domains of scientific thinking and attitudes as opposed to content acquisition. None of the studies was able to provide evidence regarding specific discussion techniques the instructor could employ to enhance students' learning. The success of an instructor-led discussion surely depends on the instructor's actions and involvement of the students. Any broad statements about discussion methods of instruction are more than likely over generalized.

Traditional laboratories utilized cookbook-style manuals with specific procedures for the students to follow. A national survey of 405 nonmajors science instructors (McIntosh & Caprio, 1990) indicated that laboratories are an important part of such courses, and that there is a general dissatisfaction with materials available for these labs. Interest has grown in inquiry-oriented laboratory activities which allow students to determine appropriate procedures.

Two studies compared a BSCS-style inquiry laboratory program to a more traditional approach. Leonard (1983) found a statistical, as well as a practical significance in favor of the BSCS-style program. Unfortunately, inappropriate statistical tests were performed, thus these results, as they stand, can add little to our empirical understanding of inquiry laboratories at the college level. Hall and McCurdy (1990) performed a similar experiment and also found a significant difference in favor of the BSCS group. However, the student was incorrectly used as the unit of analysis. Neither of these studies offer strong experimental evidence because of statistical problems, however an indication exists that the BSCS-style inquiry laboratory program may be more effective than the traditional approach and this deserves some consideration when selecting a biology program for a group of students.

DeLuca and Renner (1976) found no differences in student performance when using a traditional laboratory with a high degree of instructor input and a structured inquiry approach in which only guiding questions were given as responses to students' questions. Similar results were found by Haukoos and Penick (1983) regarding the directness of laboratory instruction provided by the instructor. Leonard (1988) compared the BSCS-style laboratory to a program that offered greater student freedom (less instructor involvement) and found no significant differences in student biology content achievement. The lack of significant differences with regard to the degree of students' independence in laboratory raises the question of whether this variable even affects students' learning. A point for educators to consider is that this variable may be learning-style dependent, and that individual students perform differently under different treatments, but the means of students' scores reveal no difference.

Walkosz and Yeany (1984) examined the effect of laboratory instruction designed to promote science process skills in relationship to cognitive development levels. Their experimental results comparing two teaching methods offer no insight because of serious statistical and design problem. However, positive correlations were found between students' levels of cognitive development and measures of understanding of science process skills, as well as laboratory quiz scores. This evidence appears to indicate that teaching methods designed to promote cognitive development may also enhance students' science achievement.

A sketchy report by Ebert-Zawasky and Abegg (1990) indicated no measurable differences between laboratory groups who authored interactive video disc lessons and laboratory groups who worked on traditional laboratory reports. Their experience did offer evidence that nonmajor biology students are capable of producing video disc lessons for use in class, and informal questionnaires indicated that it was a positive experience for all involved. This offers promise as an instructional technique that involves the students and incorporates computer technology.

Only two studies were reviewed which focussed on the effectiveness of cooperative groups. Scharmann's (1989) study was so weak that no information can be gained from it. Basili and Sanford (1991), on the other hand, did very thorough work examining the effect of small cooperative groups and students' misconceptions. Because of the nature of the data, statistical analysis was limited. However, their conclusion that small cooperative groups were effective in reducing misconceptions among college students seems to be sound.

As with classroom discussion methods, small group discussions may take many forms. The Basili and Sanford (1991) research involved specific guidelines for the students within the cooperative groups, and monitored group dynamics to ensure that behaviors conducive to reducing misconceptions occurred. It seems critical to take care in assembling groups and to provide some underlying structure for their operation.

If additional research shows small groups to be an effective teaching method for introductory college science, this may offer a viable means of instruction because a large number of students can be served at the same time and still be involved in learning.

Individualized programmed instruction is one of the few instructional methods which has strong research support for its effectiveness in college level instruction (Dunkin & Barnes, 1986; Kulik & Kulik, 1979). The three papers reviewed in this report provided no evidence that this method is specifically appropriate for nonmajors introductory biology courses serving a heterogeneous population. The papers by Robinson and Shrum (1977) and Langley and Bowman (1981) failed to reveal any difference between individualized instructional methods and any other method of instruction. The paper by Gifford and Vicks (1982) did offer support for the use of a personalized system of instruction in nonmajors biology, but their results were intended to be generalizable only to small southern rural black colleges.

The reading and writing skills of introductory college science students are an impediment to their learning of science (McIntosh & Caprio, 1990). Though this is not the domain of science faculty, these deficiencies need to be dealt with if optimum learning is to occur.

Students often ask for printed lecture notes to compensate for their poor note-taking skills. The research by Petrich and Montague (1981) indicated that this practice may be detrimental to the students' learning.

A proposed method to improve the effectiveness of textbook reading assignments is the use of adjunct questions keyed to small sections of text material. Spring, Sassenrath, and Keller (1986) indicated that this method improved students' test sccres on textbook related questions, but did not address the issue that the extra time spent studying the text may have been the important factor, and not the adjunct questions *per se*.

Trombulak and Sheldon (1989) examined journal writing as a means of improving biology content achievement. They were unable to show any positive results that could be generalized to nonmajors introductory biology.

A final consideration is the special nature of the nonscience major. Any teaching method used at the postsecondary level needs to serve a broad range of students with respect to reasoning and cognitive maturity. Hall and McCurdy's (1990) study found that 60% of the college students in an introductory biology course were functioning below the Piagetian formal operations level. They found equal growth in this dimension with both treatments, so were unable to offer any insight into effective methods for improving reasoning. In a study not within the scope of this review, Kitchener and King (1982) found that critical judgement skills continued to improve with increased schooling through the college years, but ceased to develop further once a person was removed from the school setting. In light of this information, more needs to be known regarding the teaching methods that are appropriate and effective for the diverse population that constitutes nonmajors college science courses.

RECOMMENDATIONS FOR FUTURE RESEARCH

The body of research available regarding effective methods for teaching nonmajors college science is quite limited. This review and discussion has indicated five areas for future research:

1. Specific techniques for instructor-directed classroom discussions for use with large groups need to be designed using a theoretical framework. These techniques need to address specific aspects of scientific literacy, and all domains affected need to be assessed using sound research techniques.

2. The effectiveness of small groups to facilitate desired

learning outcomes in a college setting needs to be examined further. This research should also address the preferred structure of these groups.

3. The desired nature of the laboratory instructor's involvement in inter-active instruction needs to be addressed and clear guidelines need to be established. This is of particular importance because labs are often taught by teaching assistants who are untrained and inexperienced in educational methods. Research is needed to provide insight into the teaching techniques which could be recommended to these novice instructors, as well as to programs undergoing revision.

4. Methods of instruction need to be examined in light of learning-style and cognitive development theories. It needs to be determined if certain methods favor students of a given profile at the expense of learning by others. Also, methods of instruction that promote development, in addition to meeting the course objectives, need to be identified.

If any advances are to be made, the problems that plagued the research in this review must be avoided. Sound experimental designs must be used with randomization by student if possible, and the administration of pretests. Careful attention should be given to instrumentation, as the use of a measure that is not valid and reliable can provide little information. Finally, a well executed experiment is useless if incorrect statistical analyses are performed. It is highly recommended that future research be designed with the aid of a statistician.

Additionally, it needs to be pointed out that all research addressing college nonmajors science instruction has followed the process-product model and has attempted to add to our knowledge by quantifying some perceived effect. This model has not provided the answers that are needed, and it is strongly recommended that nonquantifiable evidence be gathered and evaluated to supplement any future research efforts.

REFERENCES

American Association for the Advancement of Science. (1989). Project 2061: Science for all Americans. Washington, D.C.: Author.

American Association for the Advancement of Science. (1990). The Liberal art of science: Agenda for action. Washington, D.C.: Author.

American Institute of Biological Sciences. (1991). Annual meeting information: Workshops. BioScience, 41, 420-422.

Barnard, J.D. (1942). The lecture-demonstration versus the problem-solving method of teaching a college science course. Science Education, 26, 121-132.

Basili, P.A., & Sanford, J.P. (1991). Conceptual change strategies and cooperative group work in chemistry. Journal of Research in Science Teaching, 28, 293-304.

Boenig, R.W. (1969). Research in science education: 1938 through 1947. New York: Teachers College Press.

Caprio, W.M., McIntosh, W., & Koritz, H. (1989). Science education of the nonmajor: The problems. Journal of College Science Teaching, 18, 424-425.

DeLuca, F.P., & Renner, J.W. (1976). Structured inquiry vs. expository approach in introductory geology laboratory. Journal of College Science Teaching, 5, 307-309.

Dunkin, M.J., & Barnes, J. (1986). Research on teaching in higher education. In M.C. Wittrock (Ed.), Handbook of research on teaching, 3rd ed. (pp. 754-777). New York: Macmillan.

Dunlop, D.L., & Fazio, F. (1976). Piagetian theory and abstract preferences of college science students. Journal of College Science Teaching, 5, 297-300.

Ebert-Zawasky, K., & Abegg, G.L. (1990, April). Introductory computer interfaced videodisc systems in introductory college biology. Paper presented at the annual meeting of the National Association of Research in Science Teaching, Atlanta, GA. (ERIC Document Reproduction Service No. ED 324 240)

Fisher, K.M., Lipson, J.I., Hildebrand, A.C., Miguel, L., Schoenberg, N., & Porter, N. (1986). Student misconceptions and teacher assumptions in college biology. Journal of College Science Teaching, 15, 276-280.

Gifford, V.D., & Vicks, J. (1982). A comparison of the personalized system of instruction and a conventional biology course on the achievement of junior college freshmen. Journal of Research in Science Teaching, 19, 659-664.

Gottfried, S., Hoots, R., Creek, R., Tamppari, R., Lord, T., & Sines, R.A. (1993). College biology teaching: A literature review, recommendations & a research agenda. The American Biology Teacher, 55, 340-348.

Hall, D.A., & McCurdy, D.W. (1990). A comparison of a Biological Sciences Curriculum Study (BSCS) laboratory and a traditional laboratory on student achievement at two private liberal arts colleges. Journal of Research in Science Teaching, 27, 652-636.

Haukoos, G.D., & Penick, J.E. (1983). The influence of classroom climate on science process and content achievement of community college students. Journal of Research in Science Teaching, 20, 629-637.

Heinze-Fry, J.A. (1992). Using student self-assessment of biological concepts in an introductory biology course. Journal of College Science Teaching, 22, 39-43.

Isom, F.S., & Rowsey, R.E. (1986). The effect of a new prelaboratory procedure on students' achievement in chemistry. Journal of Research in Science Teaching, 23, 231-235.

Kitchener, K.S., & King, P.M. (1981). Reflective judgement: Concepts of justification and their relationship to age and education. Journal of Applied Developmental Psychology, 2, 89-116.

Kulik, J.A., & Kulik, C-L.C. (1979). College teaching. In P.L. Peterson & H.L. Walberg (Eds.), Research on Teaching: Concepts, findings, and implications (pp. 70-93). Berkeley, CA: McCutcheon.

Langley, W.M., & Bowman, B.L. (1981). Portable audio-tutorial vs. lectures in presentation of ecological concepts. Journal of College Science Teaching, 10, 236-238.

Lawlor, E.P. (1970). Research in science education: 1953 through 1957. New York: Teachers College Press.

Lawson, A.E. (1992). The development of reasoning among college biology students--a review of research. Journal of College Science Teaching, 21, 338-344.

Leonard, W.H. (1983). An experimental study of a BSCS-style laboratory approach for university general biology. Journal of Research in Science Teaching, 20, 807-813.

Leonard, W.H. (1989). An experimental test of an extended discretion laboratory approach for university general biology. Journal of Research in Science Teaching, 26, 79-91.

Lawson, A.E., Rissing, S.W., & Faeth, S.H. (1990). An inquiry approach to nonmajors biology. Journal of College Science Teaching, 19, 340-346.

Malachowski, M.R. (1990). A cluster approach to teaching nonmajors science. Journal of College Science Teaching, 20, 22-26.

Mason, J.M. (1952). An experimental study in the teaching of scientific thinking in biological science at the college level. Science Education, 36, 270-284.

McIntosh, W.J., & Caprio, M.W. (1990). Survey results on the quality of nonmajors science instruction. Journal of College Science Teaching, 20, 27-30.

McIntosh, W.J., & Caprio, M.W. (Eds.). (1992). Successful approaches to teaching introductory science courses. Cedar City, UT: Society of College Science Teaching.

Moll, M.B., & Allen, R.D. (1982). Developing critical thinking skills in biology. Journal of College Science Teaching, 12, 95-98.

Moore, J.A. (1985). Science as a way of knowing: Report to the cosponsoring organizations. Journal of College Science Teaching, 14, 530-532.

Morgan, P., Lemons, J., Carter, J., Grumbling, O., & Saboski, E. (1993). A scientific learning community at the University of New England. Journal of College Science Teaching, 22, 171-177.

National Science Teachers Association. (1982). Science-technology-society: Science education for the 1980s. Washington, D.C.: Author.

Perry, W., Jr. (1970). Intellectual and ethical development in the college years: A scheme. New York: Holt, Rinehart & Winston.

Petrich, J.A., & Montague, E.J. (1981). The effect of instructor-prepared handout materials on learning from lecture instruction. Journal of Research in Science Teaching, 18, 177-187.

Robinson, D.C., & Shrum, J.W. (1977). Assessment of a competency based modular biology course. Journal of College Science Teaching, 6, 228-231.

Scharmann, L.C. (March, 1989). The influence of a diversified instructional strategy on an understanding of the nature of scientific/evolutionary theory. Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco, CA.

Sigma Xi. (1990). Entry level undergraduate courses in science, mathematics and engineering: An investment in human resources. Research Triangle Park, N.C.: Author.

Spring, C., Sassenrath, J., & Ketellapper, H. (1986). Ecological validity of adjunct questions in a college biology course. Contemporary Educational Psychology, 11, 79-89.

Sundberg, M.D. (1991). Are we reinventing the wheel? BioScience, 41, 779-783.

Sundberg, M.D., Kormondy, E.J., Carter, J.L., Moore, J.A., Postlethwait, S.N., & Thorton, J.W. (1992). Reassessing the Commission on Undergraduate Education in the Biological Sciences. BioScience, 42, 442-447.

Swift, J.N. (1969). Research in science education: 1948 through 1952. New York: Teachers College Press.

Trombulak, S., & Sheldon, S. (1989). The real value of writing to learn biology. Journal of College Science Teaching, 18, 384-386.

Uno, G.E. (1988). Teaching college and college-bound biology students. The American Biology Teacher, 50, 213-216.

Walkosz, M., & Yeany, R.H. (1984, April). Effects of lab instruction emphasizing process skills on achievement of college students having different cognitive development levels. Paper presented at the annual meeting of the National Association for Research in Science Teaching, New Orleans, LA. (ERIC Document Reproduction Service No. ED 244 805)

Yager, R.E. (1981). Prologue. In N.C. Harms & R.E. Yager (Eds.), What research says to the science teacher, Volume 3 (pp. 1-4). Washington, D.C.: National Science Teachers Association.