Research on College Teaching: The Historical Background

Wilbert J. McKeachie
Center for Research on Learning and Teaching
and National Center for Research to Improve
Postsecondary Teaching and Learning
University of Michigan

Experimental research on college teaching began with single variable studies of class size and lecture vs. discussion. During the 1930s, research on student ratings of teachers began, and following World War II, studies of college teaching and learning became more common. In the decades from then to the 1980s, research moved to concern with a broader range of variables, to analyses of interactions between student and classroom variables, and to attention to processes as well as products resulting from teaching. Research on college teaching clearly meets Conant's criteria for a scientific field: progress in theory, methods, and established knowledge. Moreover, we now have demonstrated that educational research can contribute to educational practice.

Has research on college teaching made progress? Conant (1947) argued that a field could be called scientific when knowledge has accumulated, progress is evident in the development of new conceptual schemes resulting from experiments and observations, and the conceptual schemes lead, in turn, to more research. He suggested that one of the tests of whether a field qualifies would be to imagine the reaction of the pioneers in the field if they were to be brought back to life and viewed the current status of the research and theory. Would they acknowledge that there had been progress?

What would Carl E. Seashore, Edward L. Thorndike, or J. B. Edmondson say if they were to examine the research on college teaching today? In this article, I will choose a few of the significant studies in major areas of research on college teaching to subject our field to Conant's (1947) test.

World War I gave American psychologists their first opportunity to demonstrate the usefulness of psychological methods on a large scale. By the end of the war, 1,700,000 soldiers had taken Army Alpha or Beta tests of intelligence. Psychologists came out of the war with a greater sense of competence and confidence that empirical methods could solve important practical problems such as college teaching (e.g., Jones, 1923). Although educators had written scholarly works about college teaching in prior years (Klapper, 1920; Seashore, Angell, Calkins, Sanford, & Whipple, 1910), it was in the decade of the 1920s that researchers began a sustained empirical attack on the problems of college teaching.

As in other areas, the focus of research on college teaching shifted from decade to decade; therefore, this article will not follow a strictly chronological order. Rather than lose the sense of continuity in research on a particular problem, I will follow selected topics from their beginnings to the present; the topics, which will be discussed in the order in which they first appeared on the research scene, include class size, lecture versus discussion, student-centered discussion, independent study and peer learning, evaluation of teaching, teaching and technology, and the impact of cognitive psychology.

In addition to summarizing the substantive results, I shall highlight methodological lessons to be learned. My background in social, personality, and cognitive psychology undoubtedly influences both the selection and the interpretation of the studies reported. I view effective teaching and learning as a function of teaching methods interacting with student, group, and subject matter variables. I place heavier value on the importance of motivation and mindfulness in learning than would be typical of college faculty members. My hope is that this brief history will give a sense both of the progress our field of research has made and of its continuing promise and that this article will provide a useful perspective for the articles in this special issue. It is not, however, intended to be a comprehensive "Psych. Bulletin type" research review. For such a review, see Dunkin and Barnes (1986), McKeachie, Pintrich, Lin, and Smith (1986), and the triennial reviews of instructional psychology in the Annual Review of Psychology. Also useful is Menges and Mathis's (1988) Key Resources on Teaching, Learning, Curriculum, and Faculty Development.
Class Size

Early Research

Are small classes more effective than large classes? This was probably the first major question that research on college teaching tried to answer. Size does not seem to be a conceptually exciting variable. Yet, as we shall see, psychologists have, as in other studies, developed a conceptual understanding that makes the size variable more psychologically interesting.

Among the first investigators of class size were Edmondson and Mulder (1924), who compared the performance of students enrolled in a 109-student class with that of students enrolled in a 43-student class of the same course in education. Achievement of the two groups was approximately equal; the small class had a slight edge on an essay and the midsemester tests, and the large class had a slight edge on quizzes and the final examination. Students reported a preference for small classes.

Edmondson and Mulder’s (1924) results stimulated the Committee of Research at the University of Minnesota to begin a classic series of studies of class size. In 59 experiments that involved such widely varying subjects as psychology, physics, accounting, law, and education, the results of 46 favored the large classes for achievement measured largely by classroom examinations (Hudelson, 1928).

Support for small classes, however, came from studies in the teaching of French conducted by Cheydleur (1945) at the University of Wisconsin between 1919 and 1943. With hundreds of classes ranging in size from 9 to 33, Cheydleur found consistent superior performance on objective departmental examinations for the smaller classes.

Post-World War II experiments were also favorable to small classes. Macomber and Siegel’s (1957a, 1957b, 1960) experiments at Miami University are particularly important because, in addition to conventional achievement tests, they included measures of critical thinking and problem solving, scales measuring stereotypic attitudes, and tests of student attitudes toward instruction. Statistically significant differences favored the smaller classes (particularly for high-ability students). When retention of knowledge was measured 1–2 years after completion of the courses, small differences favored the smaller classes in eight of the nine courses compared (Siegel, Adams, & Macomber, 1960).

Few instructors are satisfied with the achievement of knowledge if it is not remembered, if the students are unable to use it in solving problems in which the knowledge is relevant, or if the students fail to relate the knowledge to relevant attitudes. If one takes these more basic outcomes of retention, problem solving, and attitude differentiation as criteria of learning, the weight of the evidence favors small classes, a conclusion consistent with Glass and Smith’s (1979) classic but controversial meta-analysis of class size research at all educational levels.

What Did We Learn?

The methodological lesson for both researchers and teachers is to measure higher level outcomes as well as knowledge.

What about theory? The early researchers were simply interested in answering a practical question, but there are both practical and theoretical reasons why class size should make a difference. The larger the class, the less the sense of personal responsibility and activity, and the less the likelihood that the teacher can know each student personally and adapt instruction to the individual student. Nonetheless, it seems plausible that the effect of class size on learning depends on what the teacher does. We lack descriptive studies of teacher behavior in college classes of differing sizes, but it seems likely that in larger classes, faculty members typically require less written work and spend more time lecturing and less in discussion. Thus the effect of class size also depends on the relative effectiveness of lecture and discussion, and it is not surprising that shortly after the first research on class size, experiments comparing lecture and discussion began to appear.

Lecture Versus Discussion

Early Research

Only a year after Edmondson and Mulder’s (1924) and Mueller’s (1924) studies on class size, Bane (1925) published a comparison of lecture and discussion teaching methods. Not only was Bane’s research pioneering in studying teaching methods in college; he also introduced an important methodological advance by obtaining a measure of delayed recall after the course examination as well as the conventional final examination score. There was little difference between his groups on the immediate test, but there was a significant superiority for discussion on the measure of delayed recall.

In a well-designed study of teachers of adult education courses, Solomon, Rosenberg, and Bezdek (1964) found that those who stressed lectures produced higher achievement on a factual test but were not as effective as other teachers on a test of comprehension. Other studies also supported the superiority of discussion for higher level outcomes (reviewed by McKeachie, 1984).

What Did We Learn?

In general, later research supported Bane’s (1925) earliest findings: Lecture tends to be at least equal to, and often more effective than, discussion for immediate recall of factual knowledge on a course examination, but discussion tends to be superior for long-term retention. Current cognitive theory would explain this finding in terms of greater likelihood of elaboration or deep processing when students are actively engaged in discussion.

1 Recitation, discussion, and seminar are sometimes used interchangeably. I think of recitation as involving student participation in which students answer fairly specific questions assessing their knowledge of the assigned lesson; discussion, on the other hand, is more likely to involve broader questions with openness to alternative points of view; a seminar typically involves a student presentation followed by discussion and usually implies a small group with advanced students.
The method of testing delayed recall, such as Bane used, is all too seldom used. The typical use of final course examinations, for which students have crammed, as the primary outcome measure is a major reason for the small effects often found in research on college teaching.

**Student-Centered Discussion**

**Early Research**

Interest in discussion methods seems to be a cyclical function, peaking in even-numbered decades: the 1940s, 1960s, and 1980s. Although there were a number of experiments on lecture versus discussion in the 1930s, the next peak of interest occurred after World War II when two independent movements in psychology converged to produce a spate of experiments on nondirective, group-centered, or student-centered discussion methods. Carl Roger’s nondirective approach to counseling and Kurt Lewin’s “group dynamics” movement both spilled over into research on “student-centered” or “group-centered” teaching.

Both the Rogerian and the Lewinian position emphasized movement away from the teacher’s role as expert and authority to a role of facilitating student responsibility for learning. Thus student-centered discussion went beyond conventional discussion methods in emphasizing (a) much more interaction between students and (b) student responsibility for decisions about goals and activities of the class. Accompanying this was a shift away from thinking of subject-matter knowledge as the only goal of education. Although cognitive goals were still recognized as being important, student-centered classes accepted the expression of feelings and development of group cohesion as important mechanisms for achieving “gut” learning: learning linked to motivation, attitudes and deeper understanding (McKeachie, DeValois, Dulany, Beardslee & Winterbottom, 1954).

A pioneering study of this genre preceded World War II. Zeleny (1940) found that a group-centered method was superior to a recitation method not only in cognitive outcomes but also in changes in self-confidence, leadership, and other personality variables.

Another early comparison of student-centered and instructor-centered instruction was that made by Faw (1949). A class was divided into three discussion groups; one group was taught by a student-centered method, one was taught by an instructor-centered method, and one alternated between the two methods. In comparison with the instructor-centered class, the student-centered class was characterized by more student participation, no correction by the instructor of inaccurate statements, lack of instructor direction, and more discussion of ideas related to personal experiences.

Scores on the objective course examination based on the textbook showed small but significant differences favoring the student-centered method. In the area of major interest, emotional growth, Faw’s (1949) method of evaluation was to ask students to write anonymous comments about the class. In general, comments indicated that the students felt that they received greater social and emotional value from the student-centered discussion groups.

Following the model of Lewin, Lippitt, and White’s study (1939) of authoritarian, democratic, and laissez faire group climates, the staff of the University of Michigan’s general psychology course set up an experiment in 1947, using three styles of teaching: recitation, discussion, and group tutorial (Guettel, Kelly, & McKeachie, 1954; McKeachie, 1951). In comparison with discussion and tutorial methods, the more autocratic recitation method proved to produce not only better examination scores but also greater interest in psychology, as measured by the election of advanced courses in psychology. Furthermore, students liked the recitation method better than the other methods. The greater gains in knowledge produced by the recitation method fit in with the general principle that feedback aids learning, for students in the recitation sections had weekly or semiweekly quizzes.

Despite the immediate superiority of the recitation method, two motivational outcomes favored the discussion and tutorial methods: (a) The students in discussion sections were significantly more favorable than the other groups in attitude toward psychology, and (b) a follow-up of the students 3 years later revealed that 7 men each from the tutorial and discussion groups majored in psychology, whereas none of those in the recitation group did so. Women majoring in psychology came about equally from all three groups.

Wispe (1951) carried out an interesting variation of the student-centered versus instructor-centered experiment. Instead of attempting to control the instructor personality variable by forcing instructors to teach both instructor-centered and student-centered classes, Wispe selected instructors who were rated as naturally permissive or directive. He then compared their sections of the Harvard course in “Social Relations.” He found no difference in final examination scores between students taught by different methods. Students preferred the directive method, and the poorer students gained more in directive classes.

Whereas scores on objective final examinations seem to be little affected by teaching method, there are, in addition to the positive changes in adjustment reported by Asch (1951), Faw (1949), Moore and Popham (1959), Slomowitz (1955), and Zeleny (1940), other indications that student behavior outside the usual testing situation may be influenced. Bovard (1951a, 1951b) and McKeachie (1954) found that student-centered classes showed greater insight (as rated by clinical psychologists) into problems of the young women depicted in the film “The Feeling of Rejection.” As in the studies of class size and lecture versus discussion, we find that the favorable effects of student-centered teaching methods emerge in the more subtle, “higher level” outcomes rather than in factual knowledge. These results parallel those found by Giaconia and Hodges (1982) in their meta-analysis of research on open education.

Although most of these studies were carried out in the late 1940s, they were not published until the 1950s and were thus still in the current literature when a new impetus to student-centered learning came to the fore—indeed, independent study. This movement in the late 1950s, followed by the sensitivity training movement in the 1960s, maintained interest in student-centered teaching despite the post-Sputnik panic about out-of-date and ill-organized scientific content.
What Did We Learn?

The studies of student-centered classes carried methodology forward by pointing to a broader range of outcomes involving attitudes, motivation, and personality variables. In addition, Wispe's (1951) method of studying natural variation rather than manipulated differences in teaching led to greater use of this method in later studies.

One methodological feature was important for the field of psychology in general. These were among the first studies to look at attribute-treatment interactions (ATIs). Even before Cronbach's (1957) classic article, "The Two Disciplines of Scientific Psychology," Remmers (1933) had reported that more intelligent students achieved more with a higher proportion of class time in recitation, whereas less intelligent students profited from a larger proportion of time in lecture.

Independent Study and Peer Learning

The World War II veterans bulge had barely passed when leaders in higher education began warning colleges and universities that they would face even greater strains on their facilities when the postwar "baby boomers" reached college age. The projected rapid increase in enrollments would not be matched by an increase in PhDs to teach; so mechanisms to handle more students with few faculty needed to be found. Advocates of television, teaching machines, and computers predicted enormous gains in efficiency if faculty conservatism could be overcome. The research on the effectiveness of technology, disappointingly, provided little support for these proposed panaceas.

One potential solution, however, went beyond the problem of teaching greater numbers to emphasize an important goal of education—training autonomous learners (students who would presumably be better equipped for lifelong learning because they experienced less teaching in college and practiced independent study). Thus the period spawned a number of studies of learning with reduced formal classroom instruction. Some of these sent individual learners off to study by themselves; others involved small-group, peer-learning experiences.

With the support of the Ford Foundation's Fund for Advancement of Education, a number of colleges experimented with large programs of independent study. As with other comparisons of teaching methods, few large differences were found between the achievement of students working independently and that of students taught in conventional classes. Moreover, the expected gains in independence also often failed to materialize. Students taught by independent study did not always seem to develop greater ability or motivation for learning independently. Nevertheless, a number of encouraging results emerged.

Favorable results on independent study were obtained in the experiments carried out at the University of Colorado by Gruber and Weitman (1960). In a course in freshman English in which the group met in class about 90% of the regularly scheduled hours and had little formal training in grammar, the self-directed study group was significantly superior to control groups on a test of grammar. In a course in physical optics, groups of students who attended class without the instructor learned fewer facts and simple applications but were superior to students in conventional classes in performing difficult applications and learning new material. Moreover, the areas of superiority were maintained in a retest 3 months later when the difference in factual knowledge had disappeared. In educational psychology, an experimental class meeting once a week with the instructor and twice a week in groups of 5 or 6 students without the instructor was equal to a conventional three-lecture-a-week class in mastery of content, but tended to be superior on measures of curiosity. In another experiment, students in self-directed study paid more constant attention to a lecture than did students in conventional classes.

Different kinds of learning may take place out of class than in class. The experiment reported by McKeachie, Lin, Forrin, and Teevan (1960) involved a fairly high degree of student-instructor contact. In this experiment, "tutorial" students normally met with the instructor in small groups only weekly or biweekly, but students were free to consult the instructor whenever they wished. The results of the experiment suggest that the "tutorial" students did not learn as much from the textbook as students taught in conventional lecture periods and discussion sections, but they did develop stronger motivation both for course work and for continued learning after the course. This was indicated not only by responses to a questionnaire administered at the end of the course but also by the number of advanced psychology courses later elected.

Webb and Grib (1967) reported six studies in which student-led discussions were compared with instructor-led discussions or lectures. In two of the six studies, significant differences in achievement tests favored the student-led discussions. In the other four, differences were not significant. Both students and instructors reported that the student-led discussions increased student motivation. Students reported that the sense of freedom to ask questions and express their own opinions is a major advantage of the student-led discussions. It makes theoretical sense that this opportunity to expose individual conceptions and misconceptions and compare one's ideas with those of others should contribute to learning if the group contains sufficient resources of knowledge and higher level thinking.

The Pyramid Plan

The most impressive findings on the results of student-led discussion came from the research on the Pyramid Plan at Pennsylvania State University (Carpenter, 1959; Davage, 1958, 1959). The basic plan may be represented by a description of their experiments in psychology. Each "Pyramid Group" of psychology majors consisted of 6 freshmen, 6 sophomores, 2 juniors (who were assistant leaders), and a senior (who was group leader). The group leaders were trained by a faculty supervisor. A control group received comparable special attention by being given a special program of lectures, films, and demonstrations equal to the time spent in discussion by the Pyramid Groups. The results on such measures as attitude toward psychology, knowledge of the field of psychology, scientific thinking, use of the library for scholarly reading, intellectual orientation, and resourcefulness in prob-
lem solving were significantly favorable to the Pyramid Plan. Moreover, a follow-up study showed that more of the Pyramid students continued as majors in psychology. Unfortunately, this program was never widely publicized and apparently was not adopted on other campuses.

The independent study movement thus extended student-centered teaching to peer teaching and learning. The sensitivity training movement of the 1960s was also based on the underpinnings of student-centered teaching—Lewinian and Rogerian theories.

**T-Groups**

During the 1960s, sensitivity training (T-groups, encounter groups) became the fad for high-level business executives as well as for government workers, teachers, and students. Originating in the group dynamics theories and practice of Kurt Lewin and his followers, sensitivity training groups met the 1960s generation's desire for self-analysis, confrontation of stereotypes, and overthrowing norms restricting the expression of personal needs and feelings. Many universities developed courses involving sensitivity training, and many faculty members incorporated elements of sensitivity training in conventional courses.

A number of studies of T-group effectiveness were carried out, some in academic settings. In general, the results were favorable, particularly in participants' self-reports of gains in self-understanding and sensitivity to others' feelings and behavior (see review by P. B. Smith, 1975).

**The College Classroom**

The decade of the 1960s was a period when federal funding of educational research resulted in substantial progress in our understanding of teaching and learning. The study to which I most often refer dates from this period. Combining quantitative and qualitative approaches, using data from observations and questionnaires, focusing on feelings underlying verbal expressions, and yet concerned with the goal of productive educational "work," *The College Classroom* (Mann et al., 1970) provides an unequalled source of insights with respect to student characteristics, teacher roles, and the development of the class as a working group over the course of a semester. The book describes the development of four introductory psychology classes over a semester; that development began with a phase in which the teachers' behaviors emphasized the roles of "formal authority" and "facilitator" while the student "heroes" and "snipers" tested the teachers' tolerance for student autonomy. During the first 2 weeks, much of the teacher's effort was to socialize the students into the methods and viewpoints of the field.

In a second phase, the teachers became dissatisfied with the students' lack of work, the dependence of the "anxious-dependent" students, and the excessive irrelevance of the participation of the "attention getters." Teachers, who had previously been trying to facilitate independent, autonomous work, were likely to become more punitive and authoritarian. Nonetheless, during this second period, the students and teachers began to gain a better understanding of the kind of work and involvement that was needed, and in the third phase, coming after 4 to 5 weeks, anxious dependence diminished, the classes became more collegial and cooperative, and effective work occurred.

As the term continued, the teacher assumed more control but now moved more toward a situation in which the teacher was leading a joint exploration of the subject matter with contributions from both students and teacher. The last phase, "Separation," was characterized by warmth, yet also by the beginnings of withdrawal.

Although the phases described in *The College Classroom* are not likely to be precisely replicated in other classes, most teachers will recognize some resonance with their own experience. *The College Classroom* thus helps both teachers and students understand and think about the unique development of their own classes and the sort of interactions that can facilitate productive development.

**Peer-Learning, Cooperative-Learning**

The forces of the cognitive revolution in psychology and education swept over the "touchy-feely" movement of the 1960s, but student-centered, peer-group learning survived and emerged in new cognitively oriented forms in the 1970s and 1980s as the most effective method for helping students to achieve cognitive goals.

One of the best-developed systems for helping pairs of students learn more effectively is the "Learning Cell" developed by Marcel Goldschmid of the Swiss Federal Institute of Technology in Lausanne (Goldschmid, 1971). The Learning Cell, or student dyad, refers to cooperative learning in pairs in which students alternate asking and answering questions on commonly read materials:

1. To prepare for the learning cell, students read an assignment and write questions dealing with the major points raised in the reading proper or other related materials.
2. At the beginning of each class meeting, students are randomly assigned to pairs, and one partner, A, begins by asking the first question.
3. After having answered and perhaps having been corrected or given additional information, the second student, B, puts a question to A, and so on.
4. During this time, the instructor goes from dyad to dyad, giving feedback and asking and answering questions.

A variation of this procedure has each student read (or prepare) different materials. In this case, A "teaches" B the essentials of his or her readings and then asks B prepared questions, whereupon they switch roles.

The effectiveness of the learning cell method was first explored in a large (250-student) psychology course (Goldschmid, 1970) in which four learning options were compared: seminar, discussion, independent study (essay), and learning cell. Students in the learning cell option performed significantly better on an unannounced examination and rated their ongoing learning cell's effectiveness higher regardless of the size of the class, its level, or the age of the students (Schirmerhorn, Goldschmid, & Shore, 1975).
“Pay to be a tutor, not to be tutored” is the message from these studies of peer tutoring. For example, Annis (1983) compared learning under five conditions:

1. Students read a textbook passage.
2. Students read the passage and were taught by a peer.
3. Students did not read the passage but were taught by a peer.
4. Students read the passage and prepared to teach it to other students.
5. Students read the passage and taught it to another student.

The results demonstrated that teaching resulted in better learning than did being taught. In a similar study, Bargh and Schul (1980) also found positive results; the largest part of the gain in retention was attributable to students’ deeper studying of material when preparing to teach. The research in higher education is congruent with research at other levels of education demonstrating the effectiveness of peer teaching and learning (Johnson, Maruyama, Johnson, Nelson, & Skon, 1981).

What Did We Learn?

These results fit well with contemporary theories of learning and memory. Preparing to teach, teaching, questioning, and explaining involve active thought, analysis, selection of main ideas, and processing the concepts into one’s own thoughts and words. The greater freedom, found by Webb and Grib (1967), of students to admit confusion and ask questions in peer learning groups also fits with current cognitive-instructional theory stressing the importance of restructuring incorrect or inadequate cognitive structures or schemata. Motivational effects of peer learning and independent study also fit well with motivation theorists’ emphasis on the importance of personal control of one’s environment. The results of these studies complement those of peer tutoring in elementary and high school classes. In their meta-analysis, Cohen, Kulik, and Kulik (1982) found favorable effects on learning, attitudes, and self-concept.

Methodologically, the studies in this group extended even further the list of relevant outcome measures to include, in addition to factual learning, simple versus difficult applications of concepts, retention of knowledge months later, scientific thinking, resourcefulness in problem solving, intellectual orientation, attention to lectures, ability to learn independently, student acceptance of responsibility for learning, student motivation for continued learning (assessed by both questionnaires and election of advanced courses), choosing to major in the area of the course, use of the library for scholarly reading, curiosity, attitudes, self-understanding, sensitivity to others’ feelings and behavior, and effectiveness as a leader or group member.

But the most important methodological advance evidenced in these studies was the move toward more detailed analysis of the processes responsible for outcomes. Whereas in earlier studies, such as McKeachie’s (1951, 1958), researchers had used observers and student ratings to describe classroom processes involved in the experimental comparisons, the detailed inferential coding of every verbal act represented in The College Classroom study by Mann et al. (1970) carried the analysis of interaction between teacher and students to a new level. Similarly, the processes responsible for the success of peer learning were analyzed in the studies by Webb and Gribb (1967), Gruber and Weitman (1960), and Bargh and Schul (1980). Thus these studies were early representatives of the process-product paradigm (Mitzel, 1960).

Evaluation of Teaching

Early Studies

Anyone who carries out research on teaching effectiveness quickly runs into the problem of evaluating the outcomes of teaching. Obviously, one first looks at student learning, and the studies cited above typically include measures of student achievement. But if one takes increased interest and motivation for learning as important outcomes, it is hard to come up with better measures than the students’ own perceptions of their interest. Not only can students provide data about the effects that instruction has had on them, but they also have an excellent opportunity to observe what the teacher does and what the course requires. Thus student reports of instruction have commonly been used as a source of data, not only for research, but also to improve teaching and to evaluate teaching for personnel decisions.

Once again we go back to the 1920s for the beginnings of this active field of research on college teaching. In 1927, Herman Remmers published the first of an impressive series of reports on his Purdue Rating Scale for Instructors. In that series, basic questions about validity and reliability were confronted and answered with a good deal of clarity. Unfortunately, even today many faculty raise questions that were well handled by Remmers and his students and by later research that reinforced and extended his conclusions (see Marsh, 1987):

1. Do students’ judgments agree with those of peers or administrators? Yes.
2. Do students change their minds after they have been out of college long enough to appreciate the sterling qualities of teachers whom they failed to appreciate while enrolled? No.
3. Can the poorer students’ judgments be disregarded? No. When a teacher is particularly effective with the poorer students, these students rate the teacher higher than do other students.
4. When several teachers are teaching sections of the same course, do the teachers whose students score highest on the achievement tests get the highest ratings? Yes, a result most convincingly demonstrated by Cohen’s (1981) meta-analysis of 68 validity studies.

One might suppose that the area of research on students’ rating of faculty would have been exhausted by the extensive series of research studies carried out in the past 60 years and comprehensively reviewed by Costin, Greenough, and Menges (1971), by Feldman (1978), and by Marsh (1984, 1987). However, two current lines of research seem particularly worthy of special mention.
One of these is the series of well-controlled laboratory studies involving videotaped presentations of teaching carried out by Raymond P. Perry and his associates at the University of Manitoba (Perry, Abrami, & Leventhal, 1979). Originally, the methodology was devised to investigate the "Dr. Fox effect"—named after the pseudonymous "Dr. Fox," an actor whose animation and expressiveness seduced an audience of professionals into giving high ratings to a lecture devoid of content. The Manitoba research group showed that expressiveness was positively related to learning with content held constant but that student ratings of an expressive teacher were perhaps overly generous. However, the basic findings turned out, as most research does, to reveal complex interactions. More recently the method has been turned to investigations of students' "learned helplessness"—a drop in performance following failure.

A second line of research investigates the relationship between instructor personality, instructor behavior, and teaching effectiveness. Erdle, Murray, and Rushton (1985) found two personality factors, "Achievement Orientation" and "Interpersonal Orientation," that relate to classroom behavior factors "Charisma" and "Organization." Research on student rating of instructors has often yielded somewhat similar factors (e.g., Cranton & Smith, 1990 [this issue]; Smalzreid & Remmers, 1943). Personality characteristics related to effective teaching vary, depending on the type of course (Murray, Rushton, & Paunonen, 1990 [this issue]).

What Have We Learned?

Despite faculty doubts about the ability of students to appreciate good teaching, the research evidence indicates that students are generally good judges—surprisingly so, in view of the fact that most research on student evaluation has been carried out in introductory classes, in which one would expect the students to be less able to evaluate than in more advanced classes. Moreover, ratings are robust. Potentially contaminating variables, such as time of day, class size, or required versus elective classes, make a difference, but not a large enough difference to cause researchers to misclassify a good teacher as "poor." Although one should also get evidence from other sources if a teaching evaluation is to lead to an important personnel decision, student ratings are the best validated of all the practical sources of relevant data.

But, in addition to what we have learned about evaluating teaching, the student rating research has contributed to the broader field of research on college teaching. The student rating literature led the way in the substantive identification of classroom processes affecting learning outcomes. This area of research also made it clear that different students reacted differently to the same teacher. This truth was at first taken to be a telling blow against the validity of student ratings, but, as early as 1949, attribute–treatment interactions had been demonstrated in student learning, and these interactions paralleled the ATIs in student ratings (Elliott, 1949). Furthermore, the research indicated not only that there were general attributes of effective teaching, such as clarity of explanations and enthusiasm, but also that there are a variety of ways in which teachers can be effective.

Finally, an important methodological contribution was the use (by Raymond R. Perry and the University of Manitoba group) of films to enable researchers to manipulate variables in well-controlled laboratory experiments.

Teaching and Technology

From Films to Computers

Instructional films came into widespread use during World War II, and research on uses of film in instruction continued after World War II. Carpenter and Greenhill (1955, 1958), working at Pennsylvania State University, had produced a series of studies of instructional films, and so they were well positioned to take advantage of the interest of the Ford Foundation's Fund for the Advancement of Education in encouraging the use of television for college-level instruction. During the mid-1950s, television seemed to offer great promise for coping with the great hordes of students expected to arrive as a result of the baby boom. Although many faculty members decried the loss of face-to-face contact with students, others embraced the chance to be among the first to appear on television. (In fact, I was one such [McKeachie, 1952]!)

Grants by the Fund led to a large number of well-controlled studies of the effectiveness of television, particularly as an alternative to large lectures for semester-long courses.

The results of this research may be used either to exalt or to damn television. Essentially, they indicated that although students learn nearly as much information in courses taught by television as in courses taught conventionally, live classes tend to be superior (e.g., Schramm, 1962; Sullivan, Andrews, Hollinghurst, Maddigan, & Noseworthy, 1976). Most television students learned the information needed to pass examinations, and most did not object strongly to the televised classes, although they preferred live instruction.

A course adapted for television by the addition of supplementary visual aids proved to be no more effective than televised lecture-blackboard presentations. In fact, at both Pennsylvania State University (Carpenter & Greenhill, 1955, 1958) and New York University (Adams, Carter, & Smith, 1959), the "visual" productions tended to be less effective than "bare bones" television.

The bloom of hopes for educational television had hardly begun to fade when a new technology threatened to eliminate the need for television. Teaching machines were to revolutionize education, increasing the efficiency of teaching manyfold. Skinner (1954) wrote,

We are on the threshold of an exciting and revolutionary period, in which the scientific study of man will be put to work in man's best interests. Education must play its part. It must accept the fact that a sweeping revision of educational practices is possible and inevitable. (p. 97)

Skinner further stated, "The technical problem of providing the necessary instrumental aid is not particularly difficult" (p. 95), and "If the teacher is to take advantage of recent advances in the study of learning, she [or he] must have the help of mechanical devices" (p. 95).
So persuasive were Skinner’s writings that many of the major electronic and book companies moved quickly into the teaching machine movement. Programs for teaching were written in accordance with carefully specified behavioral objectives, and, in some of the early research, human beings simulated teaching machines to test the material. It soon became apparent that programmed learning material could be presented in textlike form as effectively as by machine. Programmed books and booklets were designed to permit students to learn without formal classroom instruction or, in some cases, to be used as an adjunct to other teaching materials.

The research on teaching machines and programmed learning failed to reveal the dramatic gains expected by Skinner and the corporations that invested in the field. Nonetheless, programmed learning was not as unsuccessful educationally as it was commercially. Students do learn from the programs, but learning is generally slower than with conventional printed materials (but faster than with lectures; N. H. Smith, 1962). Reviews by Kulik, Cohen, and Ebeling (1980), Lange (1972), Nash, Muczyk, and Vettori (1971), and Schramm (1964) show programmed instruction to be slightly superior to traditional instruction in about 40% of the over 100 research studies reported, equally effective in about half the studies, and seldom less effective.

Although programmed learning was not an enormous success, two related methods proved to be more effective. One of these was the Keller Plan, or Personalized System of Instruction, a self-paced, mastery oriented, modular system of instruction that produced not only superior end-of-course achievement but also superior retention (Kulik, Kulik, & Bangert-Drowns, 1988).

The other was Postlethwait’s Audio-Tutorial method, which also involves a modularized self-instructional approach. Postlethwait developed slides, audiotapes, demonstration experiments, and other materials for modules of his introductory botany course (Postlethwait, Novak, & Murray, 1972). The results were so encouraging that the method was widely used in laboratory courses. A meta-analysis by Kulik, Kulik, and Cohen (1979) showed a significant positive effect on student achievement. It should be noted that there is little evidence of the value of conventional laboratory courses, although a few studies suggest that specific attention to scientific thinking does produce gains (see review by Hofstein & Lunetta, 1982).

The teaching machine revolution was quickly lost in the apparent brilliance of an even more glamorous technological competitor—the computer. By the early 1960s, the claims for teaching machines paled beside those made for the computer. Clearly, the flexibility offered by the computer could individualize learning much more than could the teaching machine. In fact, however, the first educational programs for computers differed little from the branching programs offered in many printed programmed-learning materials. Once again, manufacturers, research sponsors, and those interested in educational technology found that it is easier to envision the potential of educational technology than to develop the educational software that achieves this potential. By the early 1970s, it was apparent that the initial hopes had not been sustained. Nonetheless, some progress was made and, with the widespread use of microcomputers, the 1980s have seen a revival of optimism, with somewhat less grandiose visions of the role of the computer in education. In their meta-analysis of research on computer-based education, Kulik, Kulik, and Cohen (1980) found that computers made small but significant contributions to achievement. The greatest successes, however, were achieved with drill and practice programs—not the stuff of our dreams.

What Have We Learned?

Probably the most important thing we learned from the research on technology was to be skeptical about claims for revolutionary advances in education brought about by technology. We also learned that technology does not substitute for teachers; there is little likelihood that the classroom will be robotized. However, technological tools can facilitate student learning. How much learning takes place still depends on student activity and thought.

Technology has been methodologically useful in helping us examine the microprocesses of education, as exemplified in the research of Tobias (1988, in press).

Although “think-aloud” techniques and protocol analysis have less commonly been used in research on college student learning than in research with children, these methods, often used in connection with research with computers, provide an additional tool in our research kit, particularly as we move into cognitive approaches concerned not only with classroom processes but also with cognitive processes of students and teachers.

The Cognitive Era

Applications of Cognitive Psychology

The shift from behaviorism to cognitivism in psychology occurred gradually, even though it seemed to move with exciting speed during the decades of the 1960s and 1970s. Its impact on instructional research came in various forms, but one of the most striking effects was that deriving from research carried out at the University of Gothenburg by Ference Marton and his associates (Marton & Saljö, 1976a, 1976b; Svensson, 1976). Using a phenomenological-like approach that Marton calls phenomenography, the Swedish researchers described the differing ways in which students approach textbook assignments. “Surface processors” read the assignment straight through with little attempt to think about the purpose of the author or about the relationship between the assigned reading and their own previous knowledge; “deep processors,” on the other hand, are more likely to look for cues to the organization and purpose of the reading and to relate it to previous chapters or other learning. Svensson (1976), in a study of student learning over the course of a semester, similarly found contrasts between those who approached study holistically and those who were characterized as atomists. Examinations in different disciplines differed in the extent
to which a holistic approach was necessary, but the atomistic students generally did less well. These studies represented well the move from a focus on instructional materials to a focus on the student.

The Gothenburg studies stimulated researchers in Great Britain, Australia, and the United States to devise questionnaires and develop remedial programs for students whose approaches to learning were rigid and ineffective. Thus the focus in research on instruction shifted from the teacher to the student. Diagnosis of deficiencies in study skills has shifted to a greater emphasis on deep processing (or elaboration) and on meta-cognition—the ability to think about one's own learning and thinking and to choose effective strategies for different learning situations (Biggs, 1976; Entwistle, Hanley, & Hounsell, 1979; Weinstein, Underwood, Wicker, & Cubberly, 1979). Courses designed to teach students how to be more effective learners have been devised (see McKeachie, Pintrich, & Lin, 1985; Weinsteiin, Goetz, & Alexander, 1985; Weinstein & Mayer, 1986). Traditional topics of research such as test anxiety have benefited from theoretical analysis based on cognitive psychology (McKeachie, 1984; Tobias, 1985). Research on college teaching has become more closely integrated with that at other levels of education (e.g., De Corte, Lodewijks, Parmentier, & Span, 1987; Goldschmidt, 1971; Palincsar & Brown, 1984). Cognitive theory and research has spilled over into the areas of college classrooms so that "teaching thinking" has become a major theme of educational discussions in higher education, as well as in business and in other levels of education (McMillan, 1986).

**What Have We Learned?**

Cognitive theory has provided a conceptual base for understanding the results of the earlier studies and for providing guidance for instruction (e.g., Bjork, 1979). Small classes and discussion methods tend to be effective because students are actively processing material rather than passively listening and reading. Cognitive research, both in the laboratory and in the classroom, has given us a much more detailed account of how problem solving occurs in different disciplines, and this in turn is influencing textbooks and teachers. Motivation theory, too, is beginning to help us understand why some students fail to achieve up to capacity. We are now beginning to see the relationships between active deep processing and intrinsic motivation for continued learning as these relationships are affected by teaching methods that provide guidance and yet also provide opportunities for students to feel responsible and efficacious as learners.

**What of the Future?**

Recently someone asked me if we had made any progress in learning about college teaching since I began doing research in 1946. The answer is obvious, I hope, in the preceding pages. Moreover, the advances continue. But, as I told my questioner, the greater the advance, the greater the complexity ahead. The circle separating what we know from the unknown becomes even larger.

We now realize that the variables influencing learning are almost numberless. Because their interactions change from day to day, we need to move from pretest–posttest measures to studies of ongoing processes, from single-variable studies to individual students interacting in groups, and from studies of outcomes of learning to studies of what goes on in the thoughts, feelings, and desires of students. The frontier of knowledge about college teaching thus becomes even more challenging. And I suspect that the progress in the next decade will be even greater than in past decades. We have more researchers, better tools, and more comprehensive cognitive, motivation, and instructional theories. We have a clearer and more comprehensive grasp of the goals of education—the intertwining of intrinsic motivation for learning with elaboration, metacognition, and "mindful" learning.

We now know that we can teach thinking skills; in the next decade, we will gain a better understanding of how to go beyond discipline-specific skills to more broadly transferable intelligence.

We now know learning skills and strategies that generally help students to learn more effectively; in the next decade, we will better understand which strategies are most effective for which students with which material and which goals.

We now know that intrinsic motivation and a sense of self-efficacy have much to do with learning strategies and the mindfulness of student learning; in the next decade, we will gain a better understanding of the kinds of instructional methods that facilitate such motivation and that integrate learning with basic values.

We now know that peer-teaching is a very effective method of learning; in the next decade, we will gain a better understanding of how to structure peer-learning groups and when and where to use them.

We now know more about the processes leading toward educational outcomes; in the next decade, we will gain a better understanding of the way in which individual differences in motives and cognition interact with teaching methods and how the interactions change as courses develop, learners learn, and learning is assessed (see Covington & Omelich, 1988; Mann et al., 1970; Tobias, 1988).

We now know that students can evaluate teaching effectively; in the next decade, we will gain a better understanding of the validity of peer review of course materials, how various aspects of teaching contribute to learning outcomes, and how to teach students to evaluate their own gains in thinking and in knowledge.

We now know that faculty teaching can be improved with consultation; we will gain a better understanding of how consultation, training, and feedback can be combined to achieve greater effectiveness.

We now know that educational research can be of practical help to faculty members and institutions; in the next decade, we will better understand how to avoid the misuse of research and how to get more effective links to practice.

Do we meet Conant's (1947) criteria for a scientific field? I believe that if Edward L. Thorndike, Carl E. Seashore, J. B. Edmondson, and Herman Remmers could see us today, there would be a ringing chorus, "Yes"!
References


Bane, C. L. (1925). The lecture vs. the class-discussion method of college teaching. School and Society, 21, 300-302.


Tobias, S. (1988, August). *Adapting instruction to student characterization*. Presidential address of Division of Educational Psychology, American Psychological Association, Atlanta, GA.


Received July 19, 1988
Revision received July 31, 1989
Accepted August 29, 1989