ENHANCING COLLEGE STUDENTS' CRITICAL THINKING: A Review of Studies

James H. McMillan

Twenty-seven studies are reviewed that investigate the effect of instructional methods, courses, programs, and general college experiences on changes in college students' critical thinking. Only two studies used true experimental designs; most were nonequivalent pretest-posttest control group designs. The results failed to support the use of specific instructional or course conditions to enhance critical thinking, but did support the conclusion that college attendance improves critical thinking. What is lacking in the research is a common definition of critical thinking, good instrumentation to provide specific measurement, and a clear theoretical description of the nature of an experience that should enhance critical thinking.

It has been recognized for many years that one of the primary aims of education, especially at the college level, is to foster students' ability to think critically, to reason, and to use judgment effectively in decision making. Two recent, significant national reports highlight a renewed interest in college students' critical thinking skills and thus emphasize that an essential component of undergraduate education is the development of these skills. *Involvement in Learning: Realizing the Potential of American Higher Education* (National Institute of Education, 1984), an NIE report directed toward the improvement of undergraduate education, recommends that the curriculum ensure "the development of capacities of analysis, problem solving, communication, and synthesis" (p. 43). The report concludes that a college education should enable students to adapt to a changing world and that successful adaptation requires "the ability to think critically, to synthesize large quantities of new information" (p. 43). In addition, this document strongly recommends that

James H. McMillan, Educational Studies, Box 2020, Virginia Commonwealth University, Richmond, VA 23284.

> "faculty and academic deans . . . design and implement a systematic program to assess the knowledge, capacities, and skills developed by students . . . [and] . . . it is especially important that these assessments allow for judgments of the impact . . . on analytic reasoning, and the ability to synthesize." (p. 55)

In February 1985, the Association of American Colleges issued *Integrity in the College Curriculum: A Report to the Academic Community.* The report proposes as part of a minimum program that all students learn "to reason well, to recognize when reason and evidence are not enough, to discover the legitimacy of intuition, to subject data to the probing analysis of the mind" (Association of American Colleges, 1985, p. 15). Further emphasis is placed on inquiry skills, abstract logical thinking, and critical analysis.

In addition to these two reports there are other indications suggesting that increased emphasis will be put on critical thinking. In California, a graduation requirement in critical thinking has been established at the nineteen campuses of the state university system. Within the last year, three issues of *Education Leadership* and one issue of the *National Forum* have been devoted to thinking skills, and several major associations have developed workshops and networks of individuals to share ideas about this subject.

One of the primary means used to enhance critical thinking is classroom instruction. It has been assumed that if teachers use appropriate instructional methods and curriculum materials, students will improve their critical thinking skills (Young, 1980). Indeed, this view was formalized more than thirty years ago by Dressel and Mayhew (1954), who identified five critical thinking skills and conducted research to show how a college curriculum and teaching strategies could be developed to enhance critical thinking. Yet in the seminal volume *The Impact of College on Students,* Feldman and Newcomb (1969) did not summarize any studies that investigated the effect of curriculum or teaching strategies on critical thinking. The term "critical thinking" does not even appear in the index of this book.

Despite both the long-standing and more recent interest in developing college students' critical thinking ability, few published studies document the development of critical thinking or demonstrate how to improve it with specific curriculum materials or instructional methods (Terenzini, Theophilides, and Lorang, 1984). Moreover, what has been published discusses critical thinking along with other forms of higher level thinking (Pascarella, 1985). Based on a comprehensive review of research on college teaching from 1924 to 1970, McKeachie (1970) concludes that "in general, large classes are simply not as effective as small classes for critical thinking" (p. 2). That this conclusion appears to be based on three studies in marketing and economics, only one of which demonstrated statistically significant differences, is indicative of the paucity of studies on critical thinking.

McKeachie also cites seven studies to show that discussion classes are more effective than lecture classes in promoting "higher level retention and thinking," along with three other studies indicating a similar, though nonsignificant difference in outcomes for the two modes of instruction, and he finds some indication that student-centered classes, rather than instructor-centered classes, promote higher level cognitive outcomes. For the most part, other variables, such as the use of printed materials, programmed learning, independent study, simulation, and student characteristics, were found to be unrelated to critical or higher level thinking outcomes. In his conclusion, McKeachie suggests that "we need to go well beyond the simplest measures of knowledge . . . I hope this review demonstrates the value of measures of critical thinking . . ." (p. 12). In a later review by Kulick and McKeachie (1975), essentially the same conclusions are reached.

The purpose of this review is to analyze research that has investigated programs designed to change college students' ability to think critically. This article will provide a summary of the findings from these studies and analyze their methodology, examine curriculum materials and teaching strategies that may be related to improvements in critical thinking, and explore implications of these studies for needed research and documentation of changes in the critical thinking ability of college students. This author conducted an exhaustive search of literature from 1950 through 1985, using *Dissertation Abstracts, Current Index of Journals in Education,* and *Research in Education,* and books. The studies included used the term "critical thinking" to describe the dependent variable. Studies investigating related concepts, such as problem solving, reasoning, and formal operational thought, are not included in this review, although results from research in these areas are discussed in interpreting the findings here. The studies discussed here are divided into three categories: The first category includes studies of specific instructional variables; the second examines how entire courses may affect critical thinking; and the third contains studies that investigate the impact of general programs, studying the combined effect of many courses. Table 1 (see p. 20) summarizes the research problem, design, subject selection, instrumentation, and results of each study.

SPECIFIC INSTRUCTIONAL VARIABLES

The largest number of studies (13) compared classes that were different in one or two ways with respect to how the classes were taught. Six of the studies reported no significant differences. These included instructional differences in university chemistry lab classes (Coscarelli and Schwen, 1979), values clarification compared with traditional approaches (Jones, 1974), science classes that did and did not stress critical thinking objectives (Dressel

and Mayhew, 1954f),[1] BSCS compared with traditional mini-courses (Hayden, 1978), guided design compared with traditional classes (Hancock, 1981), and self-paced compared with lecture instruction (Hardin, 1977). Five of these studies used a nonequivalent pretest-posttest design, and one study used a pre-experimental posttest only design. Five of the studies used the Watson-Glaser Critical Thinking Appraisal (CTA) as the criterion measure of critical thinking.

Three studies reported mixed findings. Dressel and Mayhew (1954c) found that while the use of different course materials did not affect critical thinking, significant differences were found among sections of the same course taught by different instructors. This suggests that instructors may be an important factor, although the design used in the study is quasi-experimental with nonrandomized groups. Smith (1977), using a pretest-posttest design, found no significant difference in scores of critical thinking, but did find a significant relationship between changes in critical thinking and some instructor behaviors. Bailey (1979), in one of only two true experimental designs used in the twenty-seven studies, found significant gains from pretest to posttest of biology students taking a course emphasizing problem solving and critical thinking, but also reported a similar gain in a comparison class not emphasizing these skills. Unfortunately, although students were randomized to two classes in this study, one class covered zoology and the other covered botany, and it is not clear whether both classes had the same instructor.

Of the four studies that reported significant differences, one found that, contrary to the research hypothesis, the control students showed better critical thinking than the experimental group (Shuch, 1975). The weak design used in this study, a two group posttest-only design, may account for the findings. Fishbein (1975) found that students grouped in classes on the basis of personality variables outgained students assigned randomly. The results, which appear contradictory, showed that critical thinking, as measured by the Watson-Glaser CTA increased significantly if students were put in homogeneous or heterogeneous groups, as compared to grouping students without regard to personality. A more plausible reason for the finding may be selection differences in the groups.

Suksringarm (1976) used a true experimental Solomon design and found that students taught with BSCS materials scored significantly better on the Watson-Glaser CTA than students taught in a traditional manner. A limitation of this finding is that the Watson-Glaser CTA needed to be translated into Indian, and the reliability and validity of the translated version was not documented. Logan (1976), using a cohort-group design, found that students taking a course specifically focused on fostering critical thinking obtained significantly better scores than a comparison group. Nonequivalency

of the groups and the nature of the dependent variable, which was a locally developed twenty-item test, suggest that these results should also be viewed with caution.

In summary, then, the majority of studies of specific instructional variables fail to support the notion that implementing such instructional changes enhances students' critical thinking. Three factors may account for this overall finding. One, a time period of one semester may be too short to show changes in critical thinking, especially as measured by an instrument like the Watson-Glaser CTA, which provides a generalized measure. Second, the measures used in such studies should probably be curriculum specific. A broad measure like the Watson-Glaser CTA will be influenced by many factors, and thus the amount of variance left that will be accounted for by only the specific instructional difference is minimal. This results in the need for a very strong treatment, in much the same way as attempting to change broad, general constructs like self-concept or values. Third, because of the nature of most of these studies, in which two to six classes are compared, the effective number of treatments, or units of analysis, is the number of classes, not the number of students. With such a small number of replications of the treatments, it is not surprising to find few significant differences.

COURSES

Seven studies examined the impact of taking specific courses or programs on critical thinking. Three of the studies, by Gressler (1976) on the effects of a research course, by Williams (1951) on the effect of a debate program, and by Lyle (1958) on the effect of a specially designed psychology course, found no significant differences. Three other studies, Beckman (1956), Jackson (1961), and Dressel and Mayhew (1954e), report mixed results. In the Beckman study, students taking courses in argumentation and critical thinking were compared with students not enrolled in such courses on the Watson-Glaser CTA; no significant differences were found between the groups. However, Beckman did find differences in critical thinking gains when colleges were compared. Jackson compared gains on critical thinking of students in a debate program to matched control students. In five colleges the debaters outgained the controls on the Watson-Glaser CTA, while in four colleges the controls outgained the debaters. Given the nonrandomized nature of the study, these results suggest a strong overall college effect. Dressel and Mayhew investigated science reasoning and found that the results depended on the college studied. In two colleges, students taking a science course demonstrated the same science reasoning skills as students not taking science. One college reported significant gains for students taking science,

gains which were those of students taking a logic course, and in three colleges, students taking a specific science course showed significantly better science reasoning skills than students taking no science or general science. All three of these studies used a nonequivalent pretest-posttest control group design. The Jackson and Dressel and Mayhew results suggest that selection may be a critical plausible rival hypothesis, and that there are factors related to the college as a whole which affect critical thinking.

Only one study reported significant differences. Dressel and Mayhew (1954a) found a significant gain in critical thinking for freshmen enrolled in social science courses. However, the pretest-posttest nature of the design suggests that the conclusion that the social sciences courses caused the change should be viewed with caution.

Overall, these results are consistent with those concerned with specific instructional variables, and they further attest to the difficulty of finding instructional variables related to gains in critical thinking. The quasi-experimental nature of the studies suggests many plausible rival hypotheses which make the results difficult to interpret. What appears to be supported, however, especially by the Jackson and Dressel and Mayhew studies, is that the college experience as a whole may have a significant effect on critical thinking.

GENERAL PROGRAMS

These seven studies examine the impact of participation in one or more years of a specific program on critical thinking. A series of two studies used nonequivalent pretest-posttest control group designs to assess the effect of a special one-year program designed to enhance the critical thinking of freshmen. The first study used a locally developed measure of critical thinking and found no significant difference between the treatment and comparison groups (Tomlinson-Keasey, Williams, and Eisert, 1977). The second used the Watson-Glaser CTA and reported a significant gain for the treatment group and no gain for the control group (Tomlinson-Keasey and Eisert, 1977). Because of the nonequivalency of the groups and the nature of the statistical analysis that compared gains scores, however, the results of the second study should be viewed with caution.

The other five studies report a statistically significant gain in critical thinking over time but used no control groups. Dressel and Mayhew (1954d,g) report significant gains over the freshman year in critical thinking and science reasoning. Both of these studies had a large sample size (1002 and 990 students) and used locally developed instruments to assess critical thinking. In both studies, large gains were reported for students whose entering scores were low, whereas students scoring high initially showed little

or no gain. Whitla (1977) used nine cohort groups of freshmen, sophomores, and senior students and reported significantly better critical thinking for upperclassmen. Dressel and Mayhew (1954b) showed that, among upperclass students critical thinking in social science increased as students progressed through their course of study.

In a study of a very comprehensive and intensive effort to affect critical thinking throughout the curriculum at Alverno College, Mentkowski and Strait (1983) used longitudinal and cross-sectional data with a variety of established measures of critical thinking. Although the cohort groups in the cross-sectional comparisons showed no difference in critical thinking, significant gains in critical thinking were demonstrated by increasingly better scores on the Watson-Glaser CTA. Two other instruments, however, failed to show significant gains in critical thinking. The program at Alverno College is perhaps the best example of an entire college devoted to teaching and measuring critical thinking. At Alverno, each course specifies those aspects of critical thinking that are intended outcomes and systematically assesses them for each student. The curriculum is built around a theoretical framework of critical thinking, which allows for multiple perspectives within a broad definition. Also, multiplicity in formal, out of class measurement techniques, conducted at three points in time over four years, provides triangulation in validity and diagnostic information that is individualized to each student.

Overall, these studies suggest that seniors, in the main, are probably better at critical thinking than freshmen. However, since the most compelling data were gathered through weak pretest-posttest or longitudinal designs, it is difficult to separate the effect of college from the maturational effects that occur despite college. It is even more difficult to identify a particular curriculum or academic program that may be the causal factor in developing critical thinking, given the well documented, strong impact of out-of-class experiences.

MEASURING CRITICAL THINKING

It is worth noting that 16 of the 27 studies used scores obtained from the Watson-Glaser CTA as the dependent measure. Since the seven Dressel and Mayhew studies (1954a–g; see note 1) used instruments developed by the investigators, only four other studies of critical thinking have used instruments other than the Watson-Glaser CTA. Considering the heavy reliance on the Watson-Glaser CTA, it is worthwhile to examine this measure in evaluating the research.

The Watson-Glaser CTA consists of a series of objective items that include problems, statements, arguments, and interpretations of data similar

to those encountered in daily life. The test has five subtests designed to measure different, though related, aspects of critical thinking. These subtests include inference (discriminating among degrees of truth or falsity of inferences), recognition of assumptions (unstated or presuppositions), deduction, interpretation (weighing evidence in light of generalizations or conclusions), and evaluation of arguments (strong or relevant vs. weak or irrelevant). These scales are weighted equally to derive the total score. What is measured on the basis of these subtests is very broad and general. The Watson-Glaser CTA was developed to provide a sample of the ability to think critically about statements encountered in daily work, magazines, and newspapers. Consequently, it is not surprising that it would be difficult to show how a specific curriculum, course, or teaching strategy affects such general critical abilities in a different context than what was studied in class.

Two recent reviews of the Watson-Glaser (Abo El-Nasser, 1978; and McPeck, 1981) criticize the test on a technical basis, pointing out weaknesses in construct validity, normalization data, and the assessment of inferences as true or false rather than valid or invalid. In the *Ninth Mental Measurements Yearbook* (Mitchell, 1985), several limitations are also noted. The first is that the test appraises critical thinking only through reading, and there is no evidence that a similar score would be obtained through listening. The scope and content of the test is judged to be narrow, with a combination of neutral and controversial content that is difficult to identify. Another caution relates to the indirect evidence to support equivalency of forms A and B. Reliability is judged as adequate but not as high as other cognitive tests, and construct validity is not as thorough and systematic as it could be. In addition, a recent analysis of existing measures of critical thinking (Morante and Vlesky, 1984) revealed limited information on the reliability, norms, and other statistical data to adequately evaluate the results.

In sum, although some reviewers judge the Watson-Glaser CTA to be well constructed for use with groups of students (Modjeski and Michael, 1983), there may be sufficient technical limitations to significantly weaken research that uses this measure. In any event, it seems unlikely that a study designed to change one or even a few environmental conditions among many influences will show a difference on the basis of scores on the Watson-Glaser CTA. In this way the research in this area suffers from similar designs of investigations studying self-concept, intelligence, or creativity. It is simply very difficult to demonstrate a change in a broad, generalized construct which is influenced by many factors over a long period of time by altering one, relatively small factor. It is noteworthy to point out that in the studies that reported significant differences, the instruments were either locally developed or standardized with college students.

DISCUSSION

While common sense and the experiences of hundreds of college professors suggest that attending college results in improved critical thinking of students, there is little research reported here to suggest how such improvement takes place. The fact that only 27 studies of this important outcome could be found suggests that there may be little concern about demonstrating what college experience affects critical thinking. This paucity of research in the area, as well as the results, suggests strongly that greater attention needs to be devoted to the measurement of critical thinking skills of college students and to demonstrating the conditions that maximize positive change in these skills.

On the other hand, in examining current conceptualizations of critical thinking, additional information is needed about the development of some critical thinking skills. Two recent volumes (by Chipman, Segal and Glaser, 1985, Segal, Chipman, and Glaser, 1985), for example, provide an excellent summary of recent research in reasoning and problem-solving, which has implications for teaching critical thinking. Although not directly related to programs that seek to improve college students' critical thinking, issues addressed include the generality or domain specificity of skills, the influence of developmental differences, social background, and prior knowledge and skill levels of students.

According to a recent paper by Presseisen (1986), there appear to be three prominent positions today regarding critical thinking. Glaser (1985) believes that critical thinking involves recognition of assumptions and values, evaluating arguments and evidence, drawing inferences, and altering judgments when justified. These skills involve knowledge of logical inquiry and reasoning and an appropriate attitude of being disposed to consider problems in a thoughtful, perceptive manner. R. W. Paul (1984) suggests two levels of critical thinking. The first involves "micro-logical, analytic critical skills," such as the skills assessed in the Watson-Glaser CTA. At a second level Paul maintains that there is a focus on *dialectical* or *diaglogical reasoning* (1985), in which one becomes skillful at using the perspective of others to develop a holistic sense of rationality. Paul is critical of approaches which treat critical thinking as a series of discrete, technical skills without incorporating broader issues of values and consequences. Robert Ennis, a long-standing authority in critical thinking, has recently expanded his definition of critical thinking to include a set of thirteen dispositions and twelve abilities (1985). He defines critical thinking generally as "reflective and reasonable thinking that is focused on deciding what to believe or do" (Ennis, 1985, p. 45).

Presseisen (1986) points out that while each of these conceptualizations is

derived from the field of philosophy, there are aspects of agreement and similarity with other views of critical thinking that have a psychological or educational orientation. It is in this overlap that research in related programs has implications for developing critical thinking. It also helps to explain why a philosophically derived measure like the Watson-Glaser CTA may be insensitive to psychologically or educationally based educational programs (see Robert Sternberg, 1985, for a discussion of different theoretical perspectives on the nature of critical thinking). For example, Resnick's (1985) recent research into the nature of higher order thinking includes components that seem to be critical thinking skills, such as the use of judgment, self-regulation, and synthesis.

The study of meta-cognition (Glaser, 1984) and self-management skills (Segal, Chipman, and Glaser, 1985) has clear implications for problem solving and engaging in self-monitoring that Ennis might refer to as dispositions. Beyer (1985) suggests that meta-cognition can be important in evaluating the process of critical thinking. Glaser (1984) reviews recent research on cognition and problem solving in knowledge-rich domains and concludes that knowledge of content is a significant factor in problem solving. He specifically cites research on expert problem solving to show that process is to some extent dependent on degree of knowledge. Perkins (1985) argues that many intellectual skills are context-specific, and in the medical field there is evidence that physicians' knowledge of content is more critical than mastery of generic problem-solving processes in analysis, diagnosis, and decision making (Elstein, Shulman, and Sprafka, 1978). This finding has implications for critical thinking, since many of the steps involved in problem solving and general intelligence are also used in critical thinking (Yinger, 1980). Thus, previous research that did not take into account knowledge or context may have failed to control a confounding variable that weakened the results.

From the research summarized by Chipman, Segal, and Glaser (1985), which reviews the issue of generality or specificity of cognitive skills, it seems that instructional programs need to teach thinking skills within specific subject domains. General skills, such as those emphasized in most previous research in critical thinking, may develop as an outgrowth of skills developed to an advanced state in several areas. What may be needed are skills—operationalized and measured—that are specific to the content areas, rather than generally defined and measured skills.

Sternberg (1985) also emphasizes the role of knowledge in problem solving as a way to conceptualize critical thinking. His theory of intelligence stresses the difference between thinking in the classroom and in more practical, everyday situations. His point is that programs designed to teach critical

thinking skills need to reflect the realities of everyday problem solving and decision making if students are to effectively apply the skills to their everyday lives. Perkins (1982) adopts a similar view, and argues for programs that focus on informal reasoning rather than formal logic. The studies reviewed in this article emphasized philosophically based formal logic, and what seems to be suggested by several writers is that the programs need to focus more on the process encountered in everyday situations.

Another related area of research that overlaps with critical thinking is reflective judgment. Reflective judgment is a model of thinking that incorporates Perry's (1970) conceptualization of student's thinking (which is similar to what Paul calls the second level of critical thinking and what Ennis terms dispositions) with the work of Broughton (1977), Loevinger (1976), and others. According to Kitchener and King (1981) and Brabeck (1983), reflective judgment is an adaptation of Piaget's (1972) theory of cognitive development, especially the formal operational stage, and involves epistemological and metaphysical assumptions about the ways in which people use evidence and justify beliefs. In one study, Brabeck (1983) found that critical thinking skills such as deduction, influence and recognition of assumptions, were "necessary but not sufficient for attainment of the highest levels of reflective judgment" (p. 33).

There are several programs that have applied some of these related theoretical positions to the development of thinking skills, and to at least some extent, critical thinking skills seem to be enhanced. The ADAPT program at the University of Nebraska was structured to provide a multidisciplinary freshman curriculum that emphasized reasoning-skill development. Based on Piaget, the program enhanced formal operational thinking by teaching self-regulation of the thinking process (Tomlinson-Keasey and Eisert, 1977). Another Piagetian-based program, Project SOAR at Xavier University, has for many years offered a program to enhance the problem-solving ability of entering freshmen. The focus of SOAR is on developing cognitive reasoning, probabilistic reasoning, combinatorial reasoning, and correlations, and emphasizes active student engagement in problems. Research shows significant increases in reasoning ability as a result of participation in the program (Carmichael et al., 1980), although the evidence is weak, based on single group, pretest-posttest designs. Guided Design, developed by Charles Wales at West Virginia University, is a program that teaches the steps involved in successful problem solving. Students work in self-paced groups and receive modeling and feedback from faculty to learn each step in the process. While the program is widely adopted, and according to Stonewater (1980) has achieved impressive results, the earlier summarized study by Hancock (1981) fails to support guided design as enhancing critical thinking as measured by

the Watson-Glaser CTA. It is quite conceivable, however, that the program is successful in affecting some of the problem-solving skills that are needed in critical thinking.

What are the implications of this related research in light of the findings from the present review of studies? First, it seems clear that there is a great amount of work being done, both in program development and research, on topics that contain some of what most experts refer to as "critical thinking." This is encouraging, since the results of the studies reviewed here show weak effects on critical thinking. What is needed, perhaps, is greater specification of what thinking skills are being developed, with specific measurement of those skills. Second, it appears that older philosophical definitions of critical thinking, which are used extensively in such measures as the Watson-Glaser CTA and the Cornell Test of Critical Thinking, are changing to conform with present-day cognitive research. This seems particularly true in three respects. One is the stress on thinking in the context of everyday problems and decision-making situations, a second is the emphasis on meta-cognitive skills, and a third is the importance of developing critical thinking skills in specific content domains. It seems that current programs that hope to affect critical thinking should incorporate these concepts in program development and measurement. Third, there is a continuing need to clarify the differences between terms such as thinking, reasoning, problem solving, and critical thinking. Clear operational definitions are needed, and philosophical, educational, and psychological traditions need to be merged.

It would be helpful for researchers to develop a theoretical basis or orientation to explain why a particular experience should enhance critical thinking. For example, one study mentioned above found that grouping students by personality traits made a difference in critical thinking, another found that science students taking a specifically designed course emphasizing critical thinking skills gained more, and a third study found that debaters showed a greater gain in critical thinking scores than nondebaters. What is lacking in these studies is a theoretical description of the nature of the learning experience that led to the significant gains. Such descriptions would help to improve the construct and external validity of the research.

It may well be that the Watson-Glaser CTA is not sufficiently sensitive to pick up changes in critical thinking, but an alternative hypothesis is that the treatments are not strong enough to produce a change. Unfortunately, because of the small number of replications, lack of random assignment of subjects, manipulation of treatment, and control, it is premature to suggest that the treatments as summarized in these studies do not affect critical thinking. Given the weak designs employed and the use of the Watson-Glaser CTA, it seems best not to accept the null hypothesis. What is needed

is a stronger set of studies with more sensitive instrumentation to begin to draw conclusions about what types of instructional methods, curriculum materials, and courses enhance critical thinking. It should be pointed out, however, that like many educational programs for children of all ages, these studies are done in applied settings. This means, of course, that researchers must contend with nonrandom assignments, classes with different teachers, subject mortality in longitudinal studies, intrusion into normal academic programs, and other difficulties.

CONCLUSIONS

A few conclusions and recommendations appear to be warranted, based on this review. First, consistent with the conclusions of Pascarella (1985), it appears from these studies that college students' critical thinking improves while attending college, but it is not clear what factors affect this change. Lehmann (1963) and Keeley, Browne, and Kreutzer (1982) provide additional evidence that critical thinking improves, without attempting to suggest why or how. Second, the best work in this area appears to be provided by Dressel and Mayhew (1954). Although done more than 30 years ago, this research represents the most comprehensive program that has been undertaken to study college students' critical thinking. This research used large samples, carefully developed instruments, and adequate designs. Researchers in this area would be well advised to carefully study these investigations. Third, when the Watson-Glaser CTA is used as the dependent measure, it is likely that nonsignificant differences will be found. To improve research in this area, it may be necessary that the measurement of critical thinking coincide closely with what the intervention seeks to change. Fourth, instructors and content areas may be significant factors affecting critical thinking; designs should control for these variables. Fifth, if students begin college or a specific course with high scores on a measure of critical thinking, it is unlikely that an intervention will statistically improve that score. For selective colleges, this is a crucial factor. It may well be that such students change in ways that are difficult to assess, and currently available instruments may not be sufficiently difficult or discriminating to measure changes. Finally, the research relies on a single measure of critical thinking. What is needed is a set of multiple measures of critical thinking that can be used to triangulate the results. For instance, measures of student and teacher perceptions, judgmental analyses of essay answers, and locally devised instruments could be used in addition to appropriate standardized tests.

Unfortunately, the research on enhancing college students' critical thinking does not provide much of a foundation for implementing instructional and assessment programs that are currently being called for. What is sug-

gested from this review is that the design of programs to enhance critical thinking include features to permit strong causal inference and measurement that will be sensitive to the specific changes desired. Such designs are not easy or convenient to implement but appear to be the best approach for contributing to our understanding of the conditions that improve critical thinking skills.

Acknowledgments. The author wishes to thank Carey Chirico for her assistance in locating some studies and Katie Esham and an anonymous reviewer for their suggestions on an earlier draft. Support to complete this review was provided by the Center for Educational Development and Faculty Resources of Virginia Commonwealth University.

NOTES

1. Dressel and Mayhew (1954) contains seven separate studies (a–g), which are referred to in the text as 1954a, 1954b, etc.

REFERENCES

Association of American Colleges. (1985). *Integrity in the College Curriculum: A Report to the Academic Community.* Washington, D.C.

Abo El-Nasser, M. E. M. (1978). Conflicting concepts of critical thinking. (Doctoral dissertation, George Peabody College of Teachers). *Dissertation Abstracts International, 39,* 480A.

Bailey, J. F. (1979). The effects of an instructional paradigm on the development of critical thinking of college students in an introductory botany course. (Doctoral dissertation, Purdue University). *Dissertation Abstracts International, 40,* 3138A.

Beckman, V. E. (1956). An investigation of the contributions to critical thinking made by courses in argumentation and discussion in selected colleges. (Doctoral dissertation, University of Minnesota). *Dissertation Abstracts International, 16,* 2551A.

Beyer, B. (1985). Critical thinking: What is it? *Social Education,* 49(4): 270–276.

Brabeck, M. M. (1983). Critical thinking skills and reflective judgment development: redefining the aims of higher education. *Journal of Applied Developmental Psychology* 4(1): 23–24.

Broughton, J. (1977). Beyond formal operations: theoretical thought in adolescence. *Teachers College Record* 79(1): 87–97.

Carmichael, J. W., Hassell, J., Hunter, J., Jones, L., Ryan, M. A., and Vincent, H. (1980). Project SOAR (stress on analytical reasoning). *The American Biology Teacher* 42(3): 169–173.

Chipman, S. F., Segal, J. W., and Glaser, R., eds. (1985). *Thinking and Learning Skills (Volume 2: Research and Open Questions).* Hillsdale, NJ: Lawrence Erlbaum Associates.

Coscarelli, W. C., and Schwen, T. M. (1979). Effects of three algorithmic representations on critical thinking, laboratory efficiency, and final grade. *Educational Communication & Technology* 27(1): 58–64.

Dressel, P. L., and Mayhew, L. B. (1954). *General Education: Explorations in Evaluation.* Westport, T: Greenwood Press.

Elstein, A. S., Shulman, L. S., and Sprafka, S. A. (1978). *Medical Problem Solving.* Cambridge, MA: Harvard University Press.

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational Leadership* 43(2): 44–48.

Feldman, K. A., and Newcomb, T. M. (1969). *The Impact of College on Students* (Vol. 1). San Francisco: Jossey-Bass, Inc.

Fishbein, E. L. (1975). The effect of three patterns of small group assignment in promoting critical thinking, open-mindedness, and creativity in community college students. (Doctoral dissertation, University of Miami). *Dissertation Abstracts International, 36,* 7946A.

Glaser, R. (1984). Education and thinking: the role of knowledge. *American Psychologist* 39(2): 93–104.

Glaser, E. M. (1985). Critical thinking: educating for responsible citizenship in a democracy. *National Forum* 65(1): 24–27.

Gressler, L. A. (1976). The effect of research courses upon the attitudes and critical thinking abilities of graduate students. (Doctoral dissertation, Mississippi State University). *Dissertation Abstracts International, 37,* 3994A.

Hancock, B. W. (1981). The effect of guided design in the critical thinking ability of college level administrative science students. (Doctoral dissertation, Southern Illinois University at Carbondale). *Dissertation Abstracts International, 42,* 4275A.

Hardin, L. D. (1977). A study of the influence of a physics personalized system of instruction versus lecture on cognitive reasoning, attitudes, and critical thinking. (Doctoral dissertation, University of Northern Colorado). *Dissertation Abstracts International, 38,* 4711A.

Hayden, V. M. B. (1978). A study of the effects of traditional biology and selected biological science curriculum study (BSCS) minicourses on the attitudes, achievement levels, and critical thinking abilities of students at Alcorn State University. (Doctoral dissertation, University of Southern Mississippi). *Dissertation Abstracts International, 39,* 2167A.

Jackson, T. R. (1961). The effects of intercollegiate debating on critical thinking ability. (Doctoral dissertation, University of Wisconsin). *Dissertation Abstracts International, 21,* 3556A.

Jones, J. T. (1974). An experimental study of four interdisciplinary approaches to promoting critical thinking skills and personal development in the college classroom. (Doctoral dissertation, The University of Florida). *Dissertation Abstracts International, 35,* 5216A.

Keeley, S., Browne, and Kreutzer, J. (1982). A comparison of freshmen and seniors on general and specific essay tests of critical thinking. *Research in Higher Education* 17(2): 139–154.

Kitchener, K. and King, P. (1981). Reflective judgment: concepts of justification and their relationship to age and education. *Journal of Applied Developmental Psychology* 2(1): 89–116.

Kulick, J. A., and McKeachie, W. J. (1975). Effective college teaching. In F. Kerlinger (ed.), *Review of Research and Education,* Vol. 3, Itasca, IL: F. E. Peacock.

Lehmann, I. J. (1963). Changes in critical thinking, attitudes, and values from freshman to senior years. *Journal of Educational Psychology* 54(6): 305–315.

Loevinger, J. (1976). *Ego Development.* San Francisco: Jossey-Bass, Inc.

Logan, G. H. (1976). Do sociologists teach students to think more critically? *Teaching Sociology* 4(1): 29–48.

Lyle, E. (1958). An exploration in the teaching of critical thinking in general psychology. *Journal of Educational Research* 52(4): 129–133.

McKeachie, W. J. (1970). Research on college teaching: a review. Washington, D.C.: ERIC Clearinghouse on Higher Education. (ERIC Document Reproduction Service No. ED 043 789)

McPeck, J. E. (1981). *Critical Thinking and Education.* New York: St. Martin's Press.

Mentkowski, M., and Strait, M. J. (1983). A longitudinal study of student change in cognitive development, learning styles, and generic abilities in an outcome-centered liberal arts curriculum. (Final Report to the National Institute of Education: Research Report Number Six). Milwaukee: Alverno College, Office of Research and Evaluation. (NIE-G-77-0058)

Mitchell, J. V. Jr., ed. (1985). *The Ninth Mental Measurements Yearbook.* [Reviews of the Watson-Glaser Critical Thinking Appraisal]. Lincoln, NE: The Buros Institute of Mental Measurements of the University of Nebraska-Lincoln, pp. 1692–1694.

Modjeski, R. B., and Michael, W. B. (1983). An evaluation by a panel of psychologists of the reliability and validity of two tests of critical thinking. *Educational and Psychological Measurement* 43: 1187–1197.

Morante, E. A., and Vlesky, A. (1984). Assessment of reasoning abilities. *Educational Leadership* 42(1): 71–74.

National Institute of Education. (1984). *Involvement in Learning: Realizing the Potential of American Higher Education.* Washington, D.C.: U.S. Government Printing Office.

Pascarella, E. (1985). College environmental influences on learning and cognitive development: a critical review and synthesis. In J. Smart (ed.), *Higher Education: Handbook of Theory and Research,* Vol. I, pp. 1–62. New York: Agathon Press.

Paul, R. W. (1984). Critical thinking: fundamental to education for a free society. *Educational Leadership* 42(1): 4–14.

Paul, R. W. (1985). The critical thinking movement: a historical perspective, *National Forum* 65(1): 2–3, 32.

Perkins, D. N. (1982). Difficulties in everyday reasoning and their change with education. Report to the Spencer Foundation, Harvard University, Cambridge, MA.

Perkins, D. N. (1985). General cognitive skills: why not? In Chipman, S. F., Segal, J. W., and Glaser, R. (eds.), *Thinking and Learning Skills (Volume 2: Research and Open Questions),* pp. 339–364. Hillsdale, NJ: Lawrence Erlbaum.

Perry, W. (1970). *Forms of Intellectual and Ethical Development in the College Years: A Scheme.* New York: Holt, Rinehart & Winston.

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human Development* 15: 1–12.

Presseisen, B. Z. (1986). Critical thinking and thinking skills: State of the art definitions and practice in public schools. Paper presented at the 1986 annual meeting of the American Educational Research Association, San Francisco.

Resnick, L. B. (1985). Education and learning to think. Unpublished paper. Learning Research and Development Center, University of Pittsburgh.

Segal, J. W., Chipman, S. F., and Glaser, R., eds. (1985). *Thinking and Learning Skills (Volume 1: Relating Instruction to Research).* Hillsdale, NJ: Lawrence Erlbaum.

Shuch, M. L. (1975). The use of calculators versus hand computations in teaching business arithmetic and the effects on the critical thinking ability of community college students. (Doctoral dissertation, New York University). *Dissertation Abstracts International, 36,* 4299A.

Smith, D. G. (1977). College classroom interactions and critical thinking. *Journal of Educational Psychology* 69(2): 180–190.

Sternberg, R. J. (1985). Teaching critical thinking, part 2: possible solutions. *Phi Delta Kappan* 67(4): 277–280.

Stonewater, J. K. (1980). Strategies for problem-solving. In Young, *Fostering Critical Thinking,* pp. 33–58.

Suksringarm, P. (1976). An experimental study comparing the effects of BSCS and the traditional biology on achievement understanding of science, critical thinking ability, and attitude toward science of first year students at the Sakon Nakorn Teachers College, Thailand. (Doctoral dissertation, Pennsylvania State University). *Dissertation Abstracts International, 37,* 2764A.

Terenzini, P. T., Theophilides, C., and Lorang, W. G. (1984). Influences on students' perceptions of their academic skill development during college. *Journal of Higher Education* 55(5): 621–636.

Tomlinson-Keasey, C. A., and Eisert, D. (1977). Second year evaluation of the ADAPT program. In *Multidisciplinary Piagetian-Based Programs for College Freshmen: ADAPT.* Lincoln: University of Nebraska.

Tomlinson-Keasey, C. A., Williams, V., and Eisert, D. (1977). Evaluation report of the first year of the ADAPT program. In *Multidisciplinary Piagetian-Based Programs for College Freshmen: ADAPT.* Lincoln: University of Nebraska.

Whitla, D. K. (1977). *Value added: Measuring the Impact of Undergraduate Education.* Cambridge, MA: Office of Instructional Research and Evaluation, Harvard University.

Williams, D. E. (1951). The effects of training in debating on critical thinking ability. Unpublished master's thesis, Purdue University.

Yinger, R. J. (1980). Can we really teach them to think? In Young, *Fostering Critical Thinking,* pp. 11–32.

Young, R. E., ed. (1980). *Fostering Critical Thinking.* San Francisco: Jossey-Bass.

Received October 31, 1986

TABLE 1. Studies Investigating Changes in College Students' Critical Thinking

Problem	Design	Subjects	Instruments	Results
		Bailey (1979)		
To study the effect of a special instructional paradigm emphasizing problem solving on critical thinking	True experimental pretest-posttest control group	University students randomly assigned to either a zoology or botany course	Watson-Glaser Critical Thinking Appraisal	Significant gains in critical thinking were obtained for the treatment group; no significant differences were reported in comparing the two classes.
		Beckman (1956)		
To study the extent to which courses in argumentation and discussion improve critical thinking	Nonequivalent pretest-posttest control group	303 students in 8 colleges and universities	Watson-Glaser Critical Thinking Appraisal	No significant difference between the experimental and control classes. The differences in mean gain between colleges was significant.
		Coscarelli and Schwen (1979)		
To ascertain the effects of three representation modes on critical thinking	Nonequivalent pretest-posttest control group	190 introductory chemistry students in 10 lab classes at a large university	Watson-Glaser Critical Thinking Appraisal	No significant differences among the three groups.
		Dressel and Mayhew (1954)		
a) To assess the gain in critical thinking of freshmen students enrolled in social science courses	Pretest-posttest	1,752 freshmen students attending 11 colleges	Test of Critical Thinking in Social Science developed by a team of experts for a comprehensive evaluation of general education.	Students from all institutions showed a significant gain; students scoring lower on the pretest showed the greatest gain.

TABLE 1. *(Continued)*

Problem	Design	Subjects	Instruments	Results
b) To assess the gain in critical thinking of upper-class students	Repeated measures (three times)	236 students from several institutions	Test of Critical Thinking in Social Science	Students' scores continued to increase for three groups of students and remained constant for one group.
c) To evaluate the effect of different course materials or instructors on critical thinking	Nonequivalent pretest-posttest comparison group	Approximately 680 students in 14 different groups	Test of Critical Thinking in Social Science	There was no significant difference among classes that used different instructional methods; significant differences were found among sections of the same course taught by different instructors.
d) To assess the gain in science reasoning over a one year period for freshmen taking general education courses	Pretest-posttest	990 freshmen from 7 colleges	Test of Science Reasoning and Understanding. Developed for this study to assess science problems and conclusions.	Significant gains were reported for six colleges, with significant variations between the colleges. Large gains were found for initially low students; small or insignificant gains reported for initially high scoring students.

TABLE 1. *(Continued)*

Problem	Design	Subjects	Instruments	Results
e) To assess the relationship between gain in science reasoning over a one year period and type and amount of science taken for freshmen and sophomore students	Nonequivalent pretest-posttest control group	Approximately 470 students from 7 colleges	Test of Science Reasoning and Understanding	Two colleges reported no significant differences when comparing students taking science with students not taking science. One college reported gains of students taking science equal to gains of students taking a logic course. Three colleges showed significantly higher gain scores for students taking specific science courses such as biology or physical science than for students taking general science or no science.
f) To assess the effect of science classes emphasizing critical thinking objectives as compared to science classes not stressing these objectives on gain scores of science reasoning	Nonequivalent pretest-posttest control group	1075 students from 3 colleges	Test of Science Reasoning and Understanding	No significant differences were found comparing posttest scores of nonrandomized groups.

TABLE 1. *(Continued)*

Problem	Design	Subjects	Instruments	Results
g) To assess change in critical thinking of college freshmen after one year of general education coursework	Pretest-posttest	1002 students from 7 colleges	A Test of Critical Thinking. Prepared to assess the abilities of defining a problem, selecting pertinent information, recognizing assumptions, formulating hypotheses, and drawing conclusions.	Significant gains reported for all colleges. Students scoring low initially showed the greatest gain. Students scoring high initially showed little or no gain.
		Fishbein (1975)		
To determine the effect of three methods of grouping students on critical thinking	Nonequivalent pretest-posttest control group	359 community college students in six classes	Watson-Glaser Critical Thinking Appraisal	Students grouped complimentarily, in which they were alike on all four personality functions, or heterogeneously, with wide variations of personality variables, scored significantly higher than students grouped randomly.
		Gressler (1976)		
To investigate the effect of a foundations research course on critical thinking	Matched two group posttest only design	62 graduate students	Watson-Glaser Critical Thinking Appraisal	No significant differences between students who had taken the research course and students who had taken a similar number of graduate hours, but not the research course.

TABLE 1. *(Continued)*

Problem	Design	Subjects	Instruments	Results
		Hancock (1981)		
To study the effect of Guided Design in the development of critical thinking by comparing a class taught with Guided Design to a traditionally taught class	Nonequivalent pretest-posttest control group	234 in two sections of the same class	Watson-Glaser Critical Thinking Appraisal	No significant differences between the two classes.
		Hardin (1977)		
To test the effect of a self-paced system of instruction compared to a lecture approach to instruction	Nonequivalent pretest-posttest control group	62 university students enrolled in two physics classes	Logical Reasoning Test. Watson-Glaser Critical Thinking Appraisal	No significant differences were obtained between the pretest and posttest of both groups, and no significant differences were obtained between the self-paced and lecture groups.
		Hayden (1978)		
To compare critical thinking abilities of students taking BSCS minicourses to students taking traditional minicourses	Two group posttest only	74 university nonscience majors enrolled in a biology course	Watson-Glaser Critical Thinking Appraisal	No significant differences were obtained in comparing the two groups.

TABLE 1. *(Continued)*

Problem	Design	Subjects	Instruments	Results
		Jackson (1961)		
To compare gains on critical thinking of college debaters to comparable control groups	Nonequivalent pretest-posttest matched control group	Students from 9 colleges	Watson-Glaser Critical Thinking Appraisal	In five colleges, debate students scored significantly higher gains than controls. In four colleges the controls outgained the debaters.
		Jones (1974)		
To study the differences in critical thinking between traditional and values clarification methods of teaching two interdisciplinary general education courses	Nonequivalent pretest-posttest control group	163 university freshmen	Watson-Glaser Critical Thinking Appraisal	Traditional social science course had significantly lower posttest scores than either experimental group. No significant pretest-posttest differences were reported for the experimental groups.

TABLE 1. *(Continued)*

Problem	Design	Subjects	Instruments	Results
		Logan (1976)		
To investigate the relationship between amount of sociology and critical thinking	Eight cohort groups compared	874 students ranging from freshmen to graduate level	Locally developed 20 item test to assess the inclination and ability of students to, without prompting or directions, think critically about statements on contemporary social issues	Students at all levels scored very low on both inclination toward and competence in critical thinking. Graduate students scored slightly higher than undergraduates. Undergraduates taking a course specifically designed to develop habits of critical thinking showed significantly higher scores for both inclination toward and ability to think critically than did a comparable group just beginning the course.
		Lyle (1958)		
To study the effect of a special psychology course designed to enhance critical thinking compared to a traditionally taught psychology course	Nonequivalent pretest-posttest control group	55 students in two sections taught by the same instructor	A Test of Critical Thinking developed by Dressel and Mayhew (1954)	No significant differences between matched students.

TABLE 1. *(Continued)*

Problem	Design	Subjects	Instruments	Results
		Mentkowski and Strait (1983)		
To assess changes in critical thinking of students taking a curriculum structured to enhance critical thinking and other cognitive achievement outcomes of college	Longitudinal repeated measures over 3½ years; cross sectional, freshmen and senior cohort groups compared	350 college students	Test of Thematic Analysis, Analysis of Argument. Watson-Glaser Critical Thinking Appraisal	No significant differences were found on any of the instruments comparing freshmen and senior cohort groups; the repeated measures showed significant positive gains on the Watson-Glaser; no significant differences for the Test of Thematic Analysis or the Analysis of Argument.
		Shuch (1975)		
To compare achievement in critical thinking of students using electronic calculations to students using paper and pencil calculations	Two group posttest only	112 community college students in four classes	Watson-Glaser Critical Thinking Appraisal	Contrary to the research hypothesis, students using paper and pencil calculations achieved significantly greater scores than students using electronic calculations.

TABLE 1. *(Continued)*

Problem	Design	Subjects	Instruments	Results
		Smith (1977)		
To study the relationship between specific classroom behaviors and critical thinking	Single group pretest-posttest	138 undergraduate students in 12 classes in a small, liberal arts college	Modified Flanders Interaction Analysis. Watson-Glaser Critical Thinking Appraisal. Chickering Critical Thinking Behaviors	No change in scores on the Watson-Glaser. Significant positive relationship between change in critical thinking scores (Watson-Glaser) and reported critical thinking behavior (Chickering), and student participating, faculty encouragement, and use of student ideas, and peer-to-peer interaction.
		Susksringarm (1976)		
To study the effect of BSCS (emphasizing inquiry and higher order thinking) compared to traditional biology and instruction on critical thinking	True experimental Solomon Four Group	152 university students in four classes	Watson-Glaser Critical Thinking Appraisal (translated)	Students taught with BSCS instructional materials scored significantly better than students taught using traditional methods.
		Tomlinson-Keasey and Eisert (1977)		
To evaluate the impact of the second year of a comprehensive program to enhance critical thinking of freshmen	Nonequivalent pretest-posttest control group	104 university freshmen	Watson-Glaser Critical Thinking Appraisal	Treatment group showed statistically significant gains while no gain was reported for control groups.

TABLE 1. *(Continued)*

Problem	Design	Subjects	Instruments	Results
		Tomlinson-Keasey, Williams, and Eisert (1977)		
To evaluate the impact of the first year of a comprehensive college program to enhance critical thinking of freshmen	Nonequivalent pretest-posttest matched control group	184 university freshmen	Locally developed test of logical operations, including drawing conclusions and probability	No significant difference between the treatment and two comparison groups.
		Whitla (1977)		
To assess changes in cognitive abilities during the under-graduate college years	Nine cohort groups compared	182 freshmen in 3 cohorts; 142 seniors in 2 cohorts; 43 sophomores; 1192 alumni in 4 cohorts	Test of Logic and Rhetoric Analysis of Argument. Test of Thematic Analysis. Thematic Apperception Test (cognitive maturity section)	With the exception of natural science seniors, upperclass students composed more forceful and logical essays than freshmen. Upperclass students are significantly more able to compose more effective and logical arguments, use analysis skills, and use causal explanations than freshmen.
		Williams (1951)		
To compare gains on initial thinking of college debaters to comparable control groups	Nonequivalent pretest-posttest matched control group	Debate students and control students from one university	Watson-Glaser Critical Thinking Appraisal	No significant differences between the treatment and control groups