



Learning Sociology and Assessing Critical Thinking

Author(s): Paul J. Baker

Source: *Teaching Sociology*, Vol. 8, No. 3, Knowledge Available, Knowledge Needed: Improving Sociology Instruction (Apr., 1981), pp. 325-363

Published by: American Sociological Association

Stable URL: <http://www.jstor.org/stable/1317344>

Accessed: 30-09-2015 16:04 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



American Sociological Association is collaborating with JSTOR to digitize, preserve and extend access to *Teaching Sociology*.

<http://www.jstor.org>

This article reviews the work of philosophers, educators and sociologists who have examined critical thinking as central to the teaching/learning process. Critical thinking is a complex term which is not associated with any single pattern of teaching and testing. The many facets of critical thinking have led to the development of various measuring instruments and numerous educational experiments. In both high schools and colleges, educators have demonstrated repeatedly that instructional arrangements can be designed to promote the learning of critical thinking.

Learning Sociology and Assessing Critical Thinking

PAUL J. BAKER

Illinois State University

Sociological inquiry rests on the noble claims of human reason (Bendix, 1970). Sociologists who make their living in the academy often assert this commitment to reason in the company of undergraduate students. Sharing the educational goals of their learned colleagues, sociologists join the chorus of academicians who believe that students “should be able to think clearly, to read critically, and to write and speak effectively” (University of Illinois, 1979: 6). The rhetoric of curriculum committee reports and college catalogs is also supported by a recent national survey of sociology departments. According to Bradshaw and McPherron (1978: 17), chairpersons in universities, four-year colleges, and community colleges agree that the most important instructional goal in the sociology curriculum is the development of critical thinking. (1978:17).

Author's Note: *I am grateful to Robert Ennis, Mark Felson, and Everett K. Wilson for critical comments to an earlier draft of this article. Research for this article was supported by a grant from Illinois State University and the Fund for the Improvement of Post-Secondary Education, HEW, Grant # G007804637.*

Emphasis on critical thinking in sociology raises some interesting questions. If sociologists have placed critical thinking at the top of their curricular goals, why has the topic been ignored in their professional literature on the assessment of learning outcomes? If critical thinking is so important, why have so few sociologists designed testing methods which can demonstrate the achievement of such educational aims? These questions are rhetorical; but they point out the need to explore uncharted territory in the field of sociological instruction. We have little systematic knowledge about the testing of learning outcomes in sociology courses; and even less about the assessment of critical thinking.

This article is divided into two sections. First, *knowledge available* on the topic of critical thinking as related to sociology will be explored. While sociologists have spent little time on the topic, they can benefit from several theoretical and empirical investigations conducted by educators and philosophers. After reviewing knowledge available, I will discuss crucial areas in which *knowledge is needed* to improve the sociology teacher's capacity to assess critical thinking.

KNOWLEDGE AVAILABLE

STYLES OF TEACHING AND TESTING TECHNIQUES

Testing critical thinking is not an isolated instructional task. Rather, it is one component of a larger set of teacher-student interactions. This dyadic relationship is often examined by educators according to types of teaching styles and types of testing procedures. During the past half century, many educators have described a common pattern in which certain types of instruction are associated frequently with particular methods of evaluating learning outcomes (Beyer, 1971; Chickering, 1976; Ausubel, 1961). In its simplest form, the educators' model is bipolar. Expository teaching is related to objective exams in which students are expected to replicate the exposition of textbooks and lectures. Critics of expository teaching charge that

it involves little more than listening, reading, copying, and memorizing discrete items conveniently classified as *T* or *F*; or as *A*, *B*, *C*, or *D*. At the other extreme, educators describe the inquiry approach; students engage in problem-solving tasks which challenge their capacity to extend the realm of school knowledge to new areas outside the classroom. William Gartner (1970) has described these teaching/learning strategies as tell-recite contrasted with excite-discover.

Recognizing the conceptual crudeness of this polar typology, educators often stress the fact that teaching techniques can be placed along a continuum. The degree of elaboration varies, but educators often mention one or two intermediate types (Bigge, 1976: 327-28; Chickering, 1976: 90-91). Most typologies hold clear hierarchical learning concepts that follow Bloom's taxonomy of educational objectives or Gagné's eight conditions of learning. Students climb a progressive ladder of sophistication from the primitive skill of recalling (memorization) to the sophisticated tasks of analysis, evaluation, and problem-solving.¹

Academic folklore suggests that objective exams are concerned only with the assessment of fact retention; conversely, essay exams are always aligned with the evaluation of critical thinking. Such impressions from common sense and polar typologies must be examined with a keen eye for qualifying exceptions. Some educators construct critical thinking tests in an objective format, and many teachers administer essay exams which require little more than direct replication of information found in textbooks and lecture notes. Teachers who devote their time to reading essay exams are not necessarily evaluating the critical thinking of their students. In short, critical thinking is not associated with any single pattern of teaching and testing.

The educational goals of critical thinking can never be divorced from the broader context of student-teacher expectations; but, it is difficult to determine the occurrence of critical thinking by merely examining various types of teaching and testing techniques. The problems of placing critical thinking into a broader learning context must await clarification of the term itself. It is time to examine the nature and meaning of critical thinking.

DIVERSE DEFINITIONS OF CRITICAL THINKING

Like many important terms (e.g., scientific method, sociology), critical thinking defies precise definition. The vagueness of the term, however, does not nullify valuable work by educators and philosophers who have identified a set of intellectual skills essential to the learning process both within and outside formal school settings. I shall begin the task of definition by discussing some mental activities which are not included under the general rubric of critical thinking. While scholars have always found imaginative and creative thinking worthwhile, these thought processes often lack sufficient logical rigor to be included as part of critical thinking. Many intellectual abilities identified by Bloom at levels 1, 2, and 3—knowledge, comprehension, translation, interpretation, extrapolation, and application—would also be excluded from any list of critical thinking skills. These skills are important, but considered to be too elementary. There are many complex thinking processes which require more effort than rote memory yet are not included in various lists of critical thinking abilities.² Finally, the intellectual tasks associated with symbolic logic are rarely included in the province of critical thinking.³ Some contemporary logician-educators link critical thinking with the informal logic movement of recent years (Ennis, 1979).

No simple schemes embrace all topics which have surfaced as aspects of critical thinking. Nevertheless, I shall clarify the subject by suggesting that educators have adopted one of two approaches. First, following the inspiration of John Dewey, educators have defined critical thinking as a problem-solving process. This approach is sometimes identified by science and social science teachers as the discovery method. Second, other educators have acknowledged Dewey's contribution, but have made an extension to include additional topics of deductive logic and argument analysis. Some of the thinking skills developed by educators in these two loosely-defined learning orientations follow.

Dewey made a sharp distinction between the process of reasoning and the study of formal logic. Logical forms are not

used in actual thinking, but merely set forth the results of thinking.

[Logical] forms apply not to teaching conclusions, not to arriving at beliefs and knowledge, but to the most effective way in which to set forth what has already been concluded, so as to convince others . . . of the soundness of the result [1933: 74].

According to Dewey, “Actual thinking has its own logic; it is orderly, reasonable, reflective” (1933: 75). He believes that every thinking experience begins with “a perplexed, troubled, or confused situation”; it will lead to a “cleared-up, unified, resolved situation at the close.” Coining the term “reflective thinking,” Dewey summarized five phases of this cognitive process (1933: 107):

- (1) *suggestions*, in which the mind leaps forward to a possible solution;
- (2) an intellectualization of the difficulty or perplexity that has been *felt* (directly experienced) into a *problem* to be solved, a question for which the answer must be sought;
- (3) the use of one suggestion after another as a leading idea, or *hypothesis*, to initiate and guide observation and other operations in collection of factual material;
- (4) the mental elaboration of the idea or supposition (*reasoning*, in the sense in which reasoning is a part, not the whole, of inference); and,
- (5) *testing* the hypothesis *by* overt or imaginative *action*.

These five phases are not necessarily adopted in a fixed order. Different problems will involve different patterns with some phases overlapping, recurring, expanding in importance, or diminishing in significance. Dewey certainly did not intend to create a mechanical method of thinking which could be uniformly divided into five phases.

The irony of Dewey’s philosophy is the ease with which his pragmatic ideas became textbook dogma in the potboiler literature produced for college students planning careers in American

public schools. Thousands of students memorized the five stages of reflective thought as gospel truth. But Dewey's model acted also as a stimulus and guide for college teachers committed to the educational aims of general education (Dressel and Mayhew, 1954).

Many philosophers and educators believe Dewey too extreme in stressing the contrast between formal logic and reflective thinking. But they still accept his concept of critical thinking as a cognitive process. Critical thinkers approach a given problem or statement with a coherent set of steps or stages which can be employed flexibly as a unified system of logical operations. Elements of this cognitive process are mobilized to attack a statement or problem as a total thinking experience. The problem or statement is seen as a whole unit of intellectual inquiry. Critical thinking involves a relatively short list of mental activities essential to understanding the thoughts of others and to constructing a thoughtful position of one's own (Dressel and Mayhew, 1954; Angell, 1964; Scriven, 1976).

A second approach to defining the mental skills of critical thinking is more complex and elaborate. Philosophers of this orientation are not interested in quarreling about various schemes of reasoning or reflexive thinking. Rather, they extend the definition of critical thinking by adding numerous topics of logical analysis. Max Black's widely-adopted text, *Critical Thinking* (1952), presents an array of topics including deductive logic, language usage, induction, and the scientific method. Black's work is instructive, but the taxonomy on critical thinking constructed by Robert Ennis is the most comprehensive and systematic statement extant. Defining critical thinking as "correct assessing of statements," Ennis (1962: 84) identifies twelve aspects of this elusive process:

- (1) Grasping the meaning of a statement

Judging whether:

- (2) there is ambiguity in a line of reasoning
- (3) certain statements contradict each other

- (4) a conclusion follows necessarily
- (5) a statement is specific enough
- (6) a statement is actually the application of a certain principle
- (7) an observation statement is reliable
- (8) an inductive conclusion is warranted
- (9) the problem has been identified
- (10) something is an assumption
- (11) a definition is adequate
- (12) a statement made by an alleged authority is acceptable.

This list does not exhaust all aspects of sophisticated thinking. For example, Ennis excludes by definition any discussion of creative thinking; and he excludes temporarily the judging of value statements because the topic is too complicated and controversial. Ennis recognizes that topics in the above list are not mutually exclusive. And finally, he stresses the importance of pragmatic judgment in the assessment of any statement. The pragmatic dimension “requires the admission that complete criteria can not be established for critical thinking. An element of intelligent judgment is usually required in addition to applying criteria and knowing the meaning” (1962: 85).

During the past seventeen years, Ennis has refined, extended, and tested empirically many items in this original taxonomy (see especially, 1964, 1969, 1973, 1975). He is currently preparing a revised statement on critical thinking more extensive than the original. Renaming his area of inquiry rational thinking, Ennis elaborates on his original twelve topics. Ennis (1979: 5-6) outlines briefly characteristics of rational thinkers as persons who:

A. are *proficient* at:

- (1) *observing*,
- (2) *inferring* an explanation, the point, the meaning, and so forth,
- (3) *generalizing*,
- (4) *conceiving and stating* assumptions, alternatives, plans, predictions, definitions, and so forth,
- (5) *offering* a well-organized and well-formulated *line of reasoning*,

- (6) evaluating authoritative-sounding statements, deductive reasoning, inductive reasoning, explanations, value statements, and definitions, and
 - (7) *detecting standard problems and realizing appropriate action*; and
- B. have the *tendency* to:
- (1) *exercise the proficiency* they possess,
 - (2) take into account the *total situation*,
 - (3) be *well-informed*,
 - (4) demand *as much precision* as the subject matter permits,
 - (5) deal with the parts of a *complex situation in an orderly fashion*,
 - (6) consider seriously *other points of view* than one's own,
 - (7) withhold judgment when the evidence and/or reasons are insufficient,
 - (8) *take a position* (and *change the position*) when the evidence and reasons are sufficient to warrant so doing, and
 - (9) accept the necessity of *exercising informed judgment*; and
- C. *exercise good judgment*.

In 1962, Ennis defined critical thinking as a distinctly *reactive endeavor* in which the critical thinker responded with rigorous scrutiny to the thoughts of others. Ennis has shifted his position in recent years, presently defining critical thinkers as more positive and assertive persons. They are expected not only to detect a variety of problems, but to take appropriate action. Critical thinkers actively seek to become well informed, and do not shy away from value questions.

SOME PERENNIAL ISSUES AND NEW TERMINOLOGY IN TESTING

Sociology teachers can benefit from closer examination of various concepts of critical thinking. They also can gain new insight into the craft of teaching by becoming familiar with current literature about evaluating learning outcomes. During the past two decades, conventional ideas of testing have been challenged among educational researchers. The controversy has centered on the purposes for which tests are intended. One such

debate distinguishes between criterion-referenced tests and norm-referenced tests. Other scholars identify the functions of testing by using the contrasting terms of normative and summative evaluation. I will review briefly these basic types.

During the past half century, testing technicians have developed discriminating tests which rank people according to group norms (the group may be a classroom, school, state, or the nation). Tests such as the SAT, ACT, and GRE are designed to make comparative judgments among people. They serve as sorting devices to determine placement in various learning settings (e.g., “slow” classrooms, elite universities). While critics often have denounced such tests, an alternative measurement theory did not surface until 1963. In a landmark article, Robert Glaser (1963: 519) introduced a new dichotomy. “What I shall call criterion-referenced measures depends upon an absolute standard of quality, while what I term norm-referenced measures depends upon a relative standard.” Norm-referenced tests determine a student’s relative standing (e.g., 90th percentile); criterion-referenced tests determine a student’s level of proficiency along a continuum of achievement (e.g., highly proficient). Criterion-referenced tests “provide information as to the degree of competence attained by a particular student which is independent of reference to the performance of others” (1963: 519). The issue for Glaser is not to weight the relative worth of one’s scores, but to demonstrate the mastery of a specified content area.

Glaser’s concern is hardly new. Criterion-referenced measures—e.g., tests for a driver’s license, a lifeguard certificate, Boy Scout Badges—have been institutionalized for many years. But his insistence on using criterion-referenced measures in classrooms seemed revolutionary to many technicians in the testing industry.

Educational researchers such as Glaser and Popham are persuaded that the full range of academic subjects in schools is best evaluated by criterion-referenced tests. Others disagree, causing the topic of testing to be highly controversial (Ebel, 1978; Popham, 1978). But the education literature suggests a clear trend favoring future development of sophisticated criterion-refer-

enced tests (Hambleton et al., 1978). In many respects, the debate is not over types of tests but rather the functions tests serve. Tomko and Ennis have clarified this issue with an important recommendation (1979: 11):

Since it appears that the same test can be used both as a criterion-referenced test and a norm-referenced test, we propose to change the labels slightly and talk of criterion-referenced *testing* and norm-referenced *testing*. This labeling explicitly recognizes the dependence of the distinction upon purpose and interpretation in the given situation. This relabeling relieves us of the burden of classifying every test as one or the other type, a task we found in practice to be impossible.

This second, revised classification focuses on appropriate occasions of testing during the instructional cycle. Coining the terms “formative” and “summative” evaluation, Michael Scriven (1967) clarified different uses tests can serve during different time periods. Evaluation with feedback opportunities during the *developmental* process is strikingly different from evaluation as the *final product* of this process. His insight was amplified by Bloom et al., *Handbook on Formative and Summative Evaluation of Student Evaluation* (1971). These authors added a third type of test—diagnostic—to Scriven’s original typology.

Diagnostic testing assesses the student’s level of knowledge and learning capacities either prior to or at the beginning of instruction. Such tests serve several purposes. They may record previous achievement, determine presence or absence of prerequisite skills, or classify the student according to various characteristics considered relevant for alternative modes of instruction. In some cases, diagnostic instruments serve as pretest data, establishing a baseline to measure the intervention effects of an instructional experiment. Finally, diagnostic tests can be shared with students as an early feedback system which helps them assess strengths and weaknesses in their own learning. Diagnostic tests serve both formative and summative evaluation purposes.

The main purpose of formative evaluation is to help students and teachers assess the learning outcomes while the class is still in

the midst of its work. "The purpose is not to grade or certify the learner; it is to help both the learner and the teacher focus upon the particular learning necessary for movement toward mastery" (Bloom et al., 1971: 61). Summative evaluation, on the other hand, provides the final overall assessment of the students' learning achievements. It is the sum total of the student's recorded accomplishment for a given unit of instruction.

Glaser's dichotomy between criterion-referenced tests and norm-referenced tests is not logically inconsistent with Scriven's distinction between formative and summative evaluation. Formative evaluation is always criterion-referenced measurement; summative evaluation may be either criterion- or norm-referenced. This brief excursion into the jargon of educational testing will be useful during later discussion of various tests of critical thinking. The functions of testing can best be understood through a concrete assessment of empirical instances in which teachers and researchers have articulated explicitly the educational goals of critical thinking. It is time to move beyond conceptual discussion into an examination of empirical studies about critical thinking and the teaching/learning process.

EMPIRICAL STUDIES OF CRITICAL THINKING

Empirical scholarship on the topic of critical thinking first emerged in professional literature during the 1930s. Some of these early studies still merit careful scrutiny. To this day the pioneering investigations of Edward Glaser (1941) and Ralph Tyler (1936) offer something of value to sociologists. During this same period, sociologists investigated questions of critical thinking by studying educational methods capable of counteracting propaganda. Ironically, their educational research failed to generate empirical studies of enduring value. As researchers of human behavior, the educators seemed more sophisticated.

Relevant Studies of the High School Curriculum. As early as 1932, Ralph Tyler began to systematically measure the learning outcomes of college instruction. Guided by his interest in higher

mental processes, he studied correlations between the students' ability to recall information and their abilities to apply principles to new situations and to draw inferences from data not presented earlier. Tyler's (1936: 17) statistical analysis led to convincing conclusions: "Memorization of facts frequently fails to result in the development of higher mental processes. If the higher mental processes of application of principles and inference are really to be cultivated, learning conditions appropriate for their cultivation are necessary."

During the Depression, more young people remained in school, forcing administrators to cope with a more heterogeneous student body. In response to the new problems facing the high school curriculum, The Progressive Education Association launched a major eight-year experimental project with thirty schools throughout the United States. Tyler directed the evaluation staff for the eight-year project. His early empirical work with college students made him ideally suited to develop new learning assessment instruments for an expanding educational mission. Among Tyler's many achievements in this project was the creation of several critical thinking tests: Test 1.41—Social Problems; Test 2.51—Interpretation of Data; Test 5.11—Nature of Proof (Smith and Tyler, 1942). Since the Interpretation of Data test was used later by sociologists studying the effects of sociology instruction (Cook and Koeninger, 1939; Cook, 1950), it will be described briefly.

An important intellectual skill promoted in formal schooling is the ability to interpret various types of data such as news items, advertisements, scientific reports, and public opinion polls. Tyler and his associates reasoned that students can appraise an information item with one of four possible responses: (1) an accurate interpretation, (2) an overgeneralization, (3) an undergeneralization, and (4) crude errors. These responses were used as a guide to study data presented in various forms such as tables, prose, charts, and graphs. A jury of experts constructed fifteen interpretive statements for each information item. Any given statement could be scored as true, probably true, false, probably false, or indeterminate because of insufficient data. By matching

student responses with expert opinion, Tyler constructed a complex tally sheet with two global scores (general accuracy; crude errors) and five subscores (agreement with experts on true/false; probably true/probably false; insufficient data; complex measures of overgeneralization; and, undergeneralization). The Interpretation of Data scheme is imaginative. But the scores cannot be reduced to simple numerical indices of critical thinking abilities. Furthermore, Tyler (1942: 60) warns that it is “inadvisable to interpret scores on this test in terms of national norms.”

The second classical empirical study of critical thinking is Edward Glaser’s doctoral dissertation, *An Experiment in the Development of Critical Thinking* (1941). In collaboration with Professor Goodwin Watson, Glaser developed a battery of critical thinking tests which were administered as pre- and post-test instruments in four experimental classrooms and in four control classrooms. Some of the curriculum materials used in the experimental classes were developed by the Institute for Propaganda Analysis. Glaser’s measuring instrument was an extensive revision of the Watson tests of fair-mindedness published in 1925. The Watson-Glaser Tests of Critical Thinking contained several measuring devices: Test A1-A2: A Survey of Opinions; Test B-AR: General Logical Reasoning; Test 3: Applied Logical Reasoning Test; Test C: Inference Test; Test D: Generalization Test; Test E: Discrimination of Arguments; Test F: Evaluation of Arguments. Glaser used other tests such as the Nelson-Denny Reading Test, and a variety of observation techniques. In Denzin’s terms (1970: 26), the investigation is an example of triangulation. With elaborate data and cautious optimism, Glaser reported that a self-consciously constructed curriculum could enhance the learning of critical thinking skills. The experiment was successful.

During the 1930s, sociologists and educators were alarmed by the persuasive power of propaganda. In collaboration with the Institute for Propaganda Analysis, Alfred McClung Lee and Elizabeth Briant Lee (1939: 132) made extensive studies of propaganda techniques; they were convinced that new kinds of educational literature could help students “become more alert to the tricks of the propagandist’s trade.” They urged the reader of

The Fine Art of Propaganda to "Beware of your own prejudices. Suspend your judgments until more sides of the issue are presented. Analyze them" (1939: 134). Public schools often used institute literature, and teachers offered testimony regarding its effectiveness. Despite these testimonies, numerous empirical studies produced no conclusive results (Doob, 1950: 450). Many of the studies concentrated on attitude changes, and none of them carefully defined the nature of critical thinking. The most rigorous study using a quasi-experimental design failed to demonstrate the effectiveness of antipropaganda instruction (Osborn, 1939: 1-17). Many sociologists and social psychologists wrote extensively about fascism, authoritarian personalities, and propaganda, yet a review of this literature fails to yield any new insights about critical thinking, how it might be taught and learned, or how it might be measured.

Following Sputnik and the publication of Bruner's *Process of Education* (1961), the intellectual ferment in American public schools reached new levels of intensity. During the 1960s and early 1970s, educators and social scientists led a curricular reform movement known as "The New Social Studies" (Haas, 1977). As a result, new teaching/learning materials were produced from over one hundred special curricular projects. Many of these projects paid special attention to issues of critical thinking. Given such extensive interest in more sophisticated learning opportunities, it seems logical to ask: Did this multimillion dollar educational movement generate any new knowledge about learning and about appraising critical thinking? After extensive review of this literature, I have concluded that little new knowledge was gained. But *little* is better than *none*, and two of these projects deserve brief mention.

Under the direction of Robert Angell and his staff, many prominent sociologists helped to develop an extensive high school sociology curriculum entitled "Sociological Resources for the Social Studies" (SRSS). They produced twenty-three "Episodes in Social Inquiry" which dealt with such diverse topics as "Roles of Modern Women," "Delinquency," "Simulating Social Conflict," and "Social Mobility in the United States." Everett K.

Wilson (1970: 120) articulated the project's basic rationale for inquiry:

I've been suggesting that a stress on inductive learning (1) promotes participation, (2) constrains us to consider the relevance of problems posed, and (3) promotes a commitment to the task not so readily engendered by conventional teaching patterns. Participation, relevance, and task commitment, in turn speed learning and lengthen retention span.

This problem-solving approach, with a strong emphasis upon inductive reasoning, was brought to its logical conclusion on occasions of evaluating student performance. To quote from the instructor's manual: "SRSS does not advocate" essay tests or objective tests. These evaluation devices "too often suggest that knowledge is at a standstill" (Sjoberg, 1973: 52). The preferred methods of evaluation include problem-solving exercises, miniature empirical studies, and group discussion.

During the developmental phase of the project, extensive testing as formative evaluation was conducted by F. Lincoln Grahls. Despite considerable pressure from the publisher, Angell refused to allow these formative instruments to go on the market as summative devices. SRSS Episodes contain neither glossaries of sociology terms nor test items which measure the students' capacity to memorize such terms. The SRSS project was committed to producing high quality curricular materials which would not compromise the principles of inquiry. Unfortunately, no research findings on the inquiry model were ever published from the project. It is impossible to know if the learning outcomes hypothetically anticipated could ever be empirically demonstrated.

The SRSS materials cannot be dismissed as elementary high school literature. They demand far more rigor than the typical basic college course and are the most systematic and comprehensive sociology curriculum ever constructed on behalf of the inductive approach. In my judgment the major drawback to the SRSS material, as college literature, is its labor-intensive character. Most teachers of the basic curriculum in higher education do

not want to work that hard with ambiguous inquiry techniques.

The intellectual achievement of the Harvard Social Studies Project is also impressive. Promoting a new concept of citizenship education, Oliver, Shaver, and Newmann developed a sophisticated critical thinking model. They assume that the United States is a diverse society with numerous value conflicts (e.g., individual autonomy vs. community welfare, majority rule vs. minority rights). A series of public issues such as school desegregation is assessed analytically in terms of various value conflicts. Labeling this teaching/learning method the jurisprudential approach, the authors attempted to help students rationally evaluate public controversies. Their framework is not a superficial discussion of current events in which one person's opinion is as good as another's. Emphasizing inquiry and dialogue, the Harvard materials demanded highly disciplined skills of critical thinking.

The Harvard Project had both a developmental and research orientation. For purposes of this article, the most important aspect of the Harvard Project is an empirical evaluation of critical thinking tests. Oliver and Shaver found inadequate all available tests of critical thinking. Their chief criticism is the validity of existing instruments to measure complex learning goals of citizenship education. The structured nature of multiple-choice tests often appears unrealistic as they tend to measure only fragments of critical thinking. Current critical thinking tests give little consideration to "the components of the total decision-making process in public controversy" (1966: 183). "In real life we don't sit at a desk checking or writing responses, but are often engaged, as part of the political process, in dialogue with others" (1964: 192). However, current paper-and-pencil tests measure only specified logical operations, without placing these mental skills in a societal context in which issues are debated. "In short, the tests tend to be unlike the context in which students are likely to apply the conceptual framework required to clarify and judge public issues" (1966: 184).

The validity problem must be explored in terms of evidence on the usefulness of tests in predicting critical thinking skills outside the testing situation. Is it possible that a student would score

highly on a paper-and-pencil test, “yet be totally inept at performing the same operations in a ‘real’ setting?” (Oliver and Shaver, 1966: 185). Oliver and Shaver attempted to answer this question by first constructing a reliable evaluation instrument to measure critical thinking competence in discussion settings. Just over 100 students participated in the “crucial experiment”; scores generated in the discussion setting were compared to scores obtained from pencil-and-paper tests. The correlations were extremely low, indicating that “there may be little or no relationship between competence to defend one’s point of view in public and competence to do well on available ‘critical thinking’ tests” (1966: 225). While this study of 100 students hardly constitutes a major educational experiment, it does point in the direction of much needed research. In a small way, the study substantiates Oliver and Shaver’s reservations about standardized critical thinking tests which claim to measure significant learning outcomes in citizenship education.

Space does not permit elaborate discussion of the Harvard Project’s analytical model of instruction or the experimental design used to test its educational effectiveness. Suffice it to say that the researchers used pre- and posttest data, and experimental and control groups. Research results from this quasi-experimental format supported their claim of significant gains for students using the Harvard Project materials. In contrast to Glaser’s findings (1941), students in the experimental classes did not show significant improvement on the Watson–Glaser test. It is difficult to interpret research results in educational experiments on critical thinking. After reviewing several studies, I agree with Oliver and Shaver: generalized tests of critical thinking intended to measure transfer of learning skills may show few or no significant differences between experimental and control groups. But when students are coached to develop specific types of reasoning skills represented in the tests, they respond to the coaching by yielding significant experimental results. Essentially, the teaching/learning experiment is little more than coaching for the exam. Research literature on critical thinking is dominated primarily by studies of high school students.

Relevant Studies at the College Level. I have discussed some of the more important studies in this area. Attention will now turn to the work of sociologists at the college level. Published literature on this topic is exceedingly sparse. During the past forty years, only two sociologists have attempted to assess learning outcomes in terms of critical thinking skills.

Using Tyler's "Interpretation of Facts Test" as a measure of critical thinking, Cook and Koeninger (1939) collected pre- and posttest data on sixteen elementary sociology classes at Ohio State University. The average net learning gain for all seven subscores was negligible. Disappointed in these test results, Cook conducted a second study with students in Sociology of Education classes. Two groups of students were matched on five counts, and the same instructor taught both the experimental and the control group. The experiment employed a "flexible, group-work teaching plan" in which students engaged in various problem-solving discussions. End of semester results indicated that the experimental group made significant improvement (1950: 34). It is worth noting that Cook is the only sociologist ever to examine empirically the issue of critical thinking with the technical assistance of an educational researcher.

In 1976, Charles Logan attempted to answer the question: "Do sociologists teach students to think more critically?" His search for an answer led to a research project and an experimental class in social problems. Logan measured critical thinking by constructing twenty short-statement items which "contained some example of uncritical or nonscientific thinking about social phenomena" (1976: 32). This test was used to measure both inclination and ability to think critically. Using a quasi-experimental design, he obtained pre- and posttest data on students at all levels of instruction (elementary undergraduate, advanced undergraduate, and graduate). With the exception of his own experimental class in which the skills of critical thinking were stressed, he found low scores in all classes. Logan's conclusions are in direct contrast to prevailing professional myths about the sociology curriculum (1976: 37):

Neither inclination nor ability to think critically or scientifically about statements on social issues is associated directly or substan-

tially with increasing levels of sociological training in the ordinary course of the curriculum.

Logan explains his negative findings in the following terms (1976: 41):

The professed concern among sociologists with teaching students to think more rationally and scientifically about social phenomena may be to a considerable degree lip service that masks a hidden curriculum. Sociology professors may in fact be more concerned with teaching students *what* to think than *how* to think. That is, they may be more concerned with instilling certain ideas, beliefs, values, and ideologies in their students than with teaching them how to critically and scientifically evaluate those ideas.

This provocative study deserves much attention, but it contains a central flaw: lack of a clear definition of critical thinking. The measuring instrument does not specify those aspects of critical thinking which are under investigation. Testing items did identify various logical fallacies, problems of overgeneralization, and circular reasoning, but these items were not classified in any systematic fashion.

Logan's success with a teaching strategy to promote critical thinking is consistent with empirical studies reviewed in this article. Students can be coached to learn specific critical thinking skills found on tests administered at the end of the semester (Glaser, 1941; Lyle, 1958; Oliver and Shaver, 1966; Cook and Cook, 1950; Bloom and Rakow, 1969). These investigations confirm Glaser's (1941: 69-70) assessment of 25 studies conducted in the 1920s and 1930s. To quote from his synopsis of the pioneering literature:

There is no evidence that students acquire skill in critical thinking as a necessary by-product of the study of any given subject. On the other hand, almost any subject or project can be so taught as to put pupils on guard against hasty generalization, contradictory assertions, and the uncritical acceptance of authority. In general the research indicates that if the objective is to develop in pupils an attitude of "reasonableness" and regard for the weight of evidence and to develop ability to think critically about controversial problems, then the component attitudes and abilities involved in

thinking critically about such problems must be set up as definite goals of instruction.

In the spirit of a new Columbus, Logan may have rediscovered what an educational researcher had clarified from primary sources 35 years ago. Nevertheless, Logan still made an important contribution by placing critical thinking back on the agenda of sociology teaching and research. To paraphrase Andre Gide, perhaps it was all said before, but since no one was listening, it must be said again.

Discussion of *knowledge available* about critical thinking has emphasized various conceptual taxonomies and empirical investigations of learning outcomes in the classroom. This review of literature does not exhaust the research topics about critical thinking by any means. Four significant areas of inquiry are not covered. First, not discussed is the thorny issue of the relationship between intelligence (as measured by I.Q. tests) and critical thinking (as measured by critical thinking tests). What is the meaning of a moderately high correlation between these two types of tests? Educators differ in the interpretations (Morse and McCune, 1964: 9, vs. Oliver and Shaver, 1964: 192; see also Rust et al., 1962; Furst, 1950). Second, no mention is made of the relationship between critical thinking abilities and hierarchical stages of human development. In recent years, educators have devoted much attention to various typologies of human development (e.g., Piaget, Kohlberg, Loevinger, Erikson). These stages seem to parallel ideas of higher mental processes which place critical thinking near the top of the lofty peaks of cerebral performance. Yet the stages lack precision and the match with critical thinking skills is far from perfect (Ennis, 1975). What can be said about various stages of development and the capacity to succeed with various critical thinking tasks? Third, I have side-stepped the ambiguous literature on critical thinking as an attitude. Clearly, critical thinking is more than proficient use of logic; as early as 1910 Dewey recognized the importance of personal attitude in solving problems as well as the mental rigor necessary to carry out the task. Notions of open and closed minds

for both teachers and students merit further research commitment (Mahan, 1955; Ehrlich, 1961; Kemp, 1962). Finally, I have not discussed classroom conditions in which critical thinking might be optimized. Does the stimulation of critical thinking necessarily require small classes of homogeneously bright students who enjoy a Socratic dialogue with the teacher? Further inquiry into the relationship between critical thinking processes and group dynamics is in order (Oliver and Shaver, 1966; Rice, 1978; Baker and Jones, 1979). Some knowledge is *available* on all four topics, but more research is *needed* to advance the improvement of instruction. I will not explore these topics further. This article will continue in another direction by exploring knowledge needed to improve the testing of critical thinking outcomes.

KNOWLEDGE NEEDED TO IMPROVE ASSESSMENT OF CRITICAL THINKING

Three strategies are suggested to improve the sociology teacher's capacity to assess critical thinking. The first approach involves a survey of existing instruments which have been constructed by experts. The second strategy calls for the development of analytical tools which can be used to assess existing banks of sociology test items. And last, knowledge is needed to help teachers intelligently construct their own evaluation instruments. These three strategies are designated, respectively, as inductive, critical, and constructive. They involve more than the identification or construction of test items. It is equally important to understand the purposes for which the tests are intended. The purposes of testing, in turn, are related to the educational goals of instruction. In short, one must always consider the techniques of testing in the broader context of teaching and learning.

INDUCTIVE STRATEGY: SURVEY OF EXISTING INSTRUMENTS

During the past half century there have been many attempts to construct reliable and valid tests of critical thinking. The Illinois

TABLE 1
Instruments to Assess Critical Thinking
(Tests Designed by Experts and Currently Available
or Being Developed)

Author	Name of Instrument	Testing Format of Critical Thinking Instrument	Aspects of Critical Thinking Being Assessed
1. Goodwin Watson & Edward Glaser (1964)	Watson-Glaser Critical Thinking Appraisal (Forms Zn and Yn)	Series of relatively short passages with one hundred objective multiple choice items	1. Ability to make inferences on basis of given data; 2. Ability to recognize assumptions; 3. Deduction; ability to decide whether a statement follows necessarily from premises; 4. Interpretation; ability to weigh evidence; 5. Ability to evaluate strength or weakness of arguments
2. Robert Ennis & Jason Millman (1971)	Cornell Critical Thinking Test (Level 2)	Series of relatively short passages and a chart from a scientific experiment are used as complex stimuli for 52 objective multiple-choice items	Determine student's proficiency in judging whether: 1. a statement follows from the premises, 2. something is an assumption, 3. an observation statement is reliable, 4. an alleged authority is reliable, 5. a simple generalization is warranted, 6. a hypothesis is warranted, 7. an argument depends upon an ambiguity, 8. a statement is overvague or over specific 9. a reason is relevant (Tomko & Ennis, p. 4)
3. McBer & Company	Test of Thematic Analysis	Open ended essay responses which compare and contrast two groups of stories	"The ability to form and articulate complex, abstract concepts which will order and arrange a confusing array of inchoate data" (1978: 3)
4. McBer & Company	Analysis of Argument	Two open-ended essays responding to a controversial statement; first essay refutes statement while the second essay supports it	"The capacity to think clearly and dispassionately while seeing elements of truth in all sides of a heated controversy" (1978: 3)
5. American College Testing Program	College Outcomes Measures Project (COMP) Measurement Battery	Sixty open-ended questions clustered around 15 activities	Ability to communicate, to solve problems, to clarify values
6. American College Testing Program	COMP - Objective Test	Sixty-three multiple choice questions requiring application of general knowledge and skill to realistic situations	Ability to communicate, to solve problems, to clarify values

Critical Thinking Project has reviewed 38 tests (Stewart, 1978; Tomko and Ennis, 1979), and I have located additional evaluation instruments. Given the limitations of space, I will mention some of the most important tests classified in two broad categories: (1) expertly-designed tests which have maximized reasoning skills and minimized the importance of content (see Table 1); (2) published tests which provide useful formats to

TABLE 2
Instruments to Assess Critical Thinking
(Published Tests Relevant but Not Readily Available
for Classroom Use)

Author	Name of Instrument	Testing Format of Critical Thinking Instrument	Aspects of Critical Thinking Being Assessed
7. Paul Baker & Janet Jones (1979) Baker (1979)	Creative Writing on a Social Problem	Forty-five minute essay about a specific social problem (e.g., poverty, crime)	Ability to: (1) define the problem, (2) present supporting evidence, (3) articulate cause-effect relationships, (4) offer a logically consistent statement regarding a solution (or non-solution) to problem
8. Charles Brown (1950)	Name not given	A short passage with 5 objective multiple-choice items to be ranked from 1 (best) to 5 (worst) reasons for a given conclusion	Ability to choose rational evidence to support a given conclusion
9. Paul Dressel & Lewis Mayhew (1954)	A Test of Critical Thinking in the Social Sciences	Passages, graphs and charts presented as complex stimuli; series of objective multiple-choice items follow each passage or chart	Ability to: (1) identify central issues, (2) recognize underlying assumptions, (3) evaluate evidence or authority, (4) recognize limitations of data, (5) establish relationships, (6) draw warranted conclusions
10. Paul Dressel & Lewis Mayhew (1954)	A Test of Critical Thinking (form G)	Passages and short paragraphs presented as complex stimuli; series of objective multiple-choice items follow each passage	Ability to: (1) define a problem, (2) select pertinent information, (3) recognize assumptions, (4) formulate relevant hypothesis, (5) draw conclusions validly
11. Albert Levi (1948)	Inventory of Social Understanding I	150 short statements about social issues; 1 of 5 responses solicited: (1) Statement which can be proved to be true (2) Statement can be proved false (3) Statement for which there is a preference though it cannot be proved true or false (4) Statement is rejected though it cannot be proved false (5) Statement falls in none of the above four classes	Ability to discriminate between statements of social fact and statements of social value
12. Charles Logan (1976)	Name not given	Open ended responses to 20 short statements	Inclination to recognize and ability to adequately criticize various logical fallacies and non-scientific assertions
13. Horace Morse & George McCune (1964)	Testing of Study Skills and Critical Thinking: Section II	Series of short statements requiring respondent to classify as <u>fact</u> or <u>opinion</u>	Ability to distinguish between facts and opinions or between facts and an author's interpretation of facts
14. Horace Morse & George McCune (1964)	Testing of Study Skills and Critical Thinking: Section III	Series of short statements requiring students to classify as primary sources or secondary accounts	Ability to distinguish between two kinds of statements about evidence: primary sources and secondary accounts
15. Horace Morse & George McCune (1964)	Testing of Study Skills and Critical Thinking: Section V	A group of key references are listed and a series of questions are given; students are asked to match a specific reference with a given question	The ability to find and critically assess relevant information on a given social issue
16. Horace Morse & George McCune (1964)	Testing of Study Skills and Critical Thinking: Section VI "Exerciae in Open-Mindedness"	Series of short statements requiring student to respond in terms of <u>All</u> , <u>Most</u> , <u>Many</u> , <u>Some</u> , <u>No</u>	Ability to identify qualified generalizations and thereby avoid unqualified assertions of <u>all</u> or <u>none</u>

(Table 2 Continued)

TABLE 2 (Continued)

17. Horace Morse & George McCune (1964)	Testing of Study Skills and Critical Thinking: Section XIII "Exercise in Consistency"	A paragraph is presented and it is followed by a series of questions; student classified each question as consistent or inconsistent with ideas found in the paragraph	Ability to recognize questions which are pertinent to a particular line of reasoning
18. Horace Morse & George McCune (1964)	Testing of Study Skills and Critical Thinking: Section XV	Series of short statements requiring students to classify each statement as <u>thoughtful</u> or <u>emotional</u>	Ability to distinguish between biased and objective statements
19. Horace Morse & George McCune (1964)	Testing of Study Skills and Critical Thinking: Section XVI	A generalization is presented; it is followed by a series of statements to be classified as supporting the generalization or irrelevant to the generalization	Ability to recognize empirical statements which support generalizations. "Given a general statement and a body of data, what are the data which support the general statement?" (1964: 80)
20. Horace Morse & George McCune (1964)	Testing of Study Skills and Critical Thinking: Section XVII	A short statement is presented; it is followed by a series of questions to be classified as <u>meaningful</u> or <u>not meaningful</u>	Ability to recognize pertinent questions which extend empirical knowledge about a given statement
21. Donald Oliver & James Shaver (1966)	Social Issues Analysis Test I (SIAT #1)	An argumentative dialogue is presented and followed by a series of objective multiple-choice items	"Assess how well students can identify selected intellectual operations occurring in an argumentative dialogue" (1966: 191). Intellectual operations are defined as: (1) Problem identification and differentiation; public controversy requires differentiating: (a) empirical or factual controversy, (b) value controversy, (c) definitional controversy; (2) Making explicit important assumptions, (3) Identifying and using appropriate strategies for dealing with different types of problems, (4) Identifying and using common dialectical operations: (a) generalizing, (b) specifying, (c) qualifying, (5) Identifying relevant problems
22. Donald Oliver & James Shaver (1966)	Social Issues Analysis Test 2 (SIAT #2)	An argumentative dialogue is presented and followed by a series of objective multiple-choice items	"Assess both how well the student can identify the substance of an argumentative dialogue (in contrast to the intellectual operations occurring in it) and the student's ability to select the best rebuttals which might be used to counter statements made in the dialogue" (1966: 194)
23. Donald Oliver & James Shaver (1966)	Social Issues Analysis Test 3 (SIAT #3)	Interview format. Interviewer reads aloud a dialogue, while student follows on his own copy. Student is asked open-ended questions about types of problems and appropriate strategies for solution	"Requires the student to differentiate and identify various types of disagreements as well as suggest how the disagreements might best be clarified or resolved through different analytical strategies" (1966: 204)
24. Donald Oliver & James Shaver (1966)	Social Issues Analysis Test 4 (SIAT #4)	Free discussion of 12 students who respond to a controversial case. Discussion is scored for each individual participant who expresses one or more "items of thought"	Two scoring systems are superimposed on each other: (1) static categories--(a) value judgments, (b) legal claims, (c) factual claims, (d) definitional claims, (e) relevance, (f) debate strategy; (2) dynamic categories--(a) consistency-inconsistency, (b) specification-generalization, (c) qualification
25. David Simon (1977)	Name not given	Students select various magazine articles by American social critics and score the articles according to the following ideological orientations: (1) pluralist, (2) radical liberal, (3) conflict, (4) other	Ability to analyze journalistic articles according to problem loci, problem solution, and problem outcome; this analytical model provides the means to determine ideological categories of pluralist, radical liberal, and conflict

(Continued)

TABLE 2 (Continued)

26. Eugene Smith Ralph Tyler & Evaluation of the Eight Year Study (1942)	Interpretation of Data Test	Passages, charts, tables, graphs presented as complex stimulus; series of statements follow and respondent judges on the basis of data provided, whether the statement is: (1) true (2) probably true (3) not sufficient to indicate whether there is any degree of truth or falsity (4) probably false (5) false	Test explores two related questions: (1) "To what extent does the student recognize the limitations of the data?" "How accurately does the student perceive various types of relationships in the data?" (1942: 56-58)
--	-----------------------------	--	--

assess learning outcomes. The existing content in most of these tests, however, is either obsolete or irrelevant for sociologists (see Table 2).

According to Tomko and Ennis (1979: 6) the Watson-Glaser Critical Thinking Appraisal and the Cornell Critical Thinking Test are "the only general tests of critical thinking currently available" in the market of expertly-designed tests. Both tests include manuals with discussion of validity and reliability data. The designers of these tests have specified carefully the skills of logical reasoning they attempt to assess. They are constructed with carefully devised multiple-choice objective items. Since these tests measure generic reasoning skills, it is difficult to know how potentially relevant they are to sociology. Oliver and Shaver (1966: 184) see little value in such pencil-and-paper tests for civic education, but this negative opinion needs to be further examined empirically by sociologists in their classrooms.

In recent years considerable attention has been given to issues of academic accountability and the educational value of a liberal arts education. One aspect of this soul searching is a renewed effort to measure the lofty claims of general education. Responding to the need to know more about learning outcomes, two testing firms are developing new critical thinking tests. Since these efforts are in the developmental stage, they will be discussed as tentative projects with a future yet to be written in the annals of the testing enterprise.

McBer and Company (1978: 1) is developing a Comprehensive Cognitive Assessment Battery (CCAB) which probes those basic competencies essential for sophisticated adult roles. They are

examining dimensions of higher education which lead students "to become critical and discerning thinkers, competent problem solvers, and persons who will take the initiative in finding solutions to problems." CCAB consists of four tests (Test of Thematic Analysis, Analysis of Argument, Learning Style Inventory, and Self-Definition), but the first two instruments more directly concern issues of critical thinking and sociology. The Test of Thematic Analysis is described as "a measure of critical thinking ability." "This test assesses the process of forming, articulating, and using concepts when comparing, contrasting, and rearranging information in order to draw conclusions" (1978: 2; see also Whitla, 1977: 9). Analysis of Argument measures intellectual flexibility. "This test reflects the capacity to think clearly and dispassionately when dealing with viewpoints that involve controversial and emotional issues" (Mcber, 1978: 2). Designed to be content free, both tests require students to respond in writing to various stories and controversial statements.

In collaboration with twelve colleges and universities, The American College Testing Program (ACT) recently launched a major ten-year project to develop new methods of assessing intended learning outcomes of general education. The College Outcome Measures Project (COMP) evaluates six types of cognitive activities: three content-related domains (Functioning within Social Institutions, Using Science and Technology, and Using the Arts) and three process-related domains (Communicating, Solving Problems, and Clarifying Values). The three content-related domains are crosscut by the three process-related domains creating a basic three by three matrix (Forrest, 1978: 3-8). ACT also makes use of various real life materials drawn from the print and electronic media. The first phase of test development includes (1) a Measurement Battery, (2) an Objective Test, and (3) an Activity Inventory. The Measurement Battery consists of several tests in which students respond to various stimulus items by written essays or audiotaped statements. "The Objective Test is the product of an effort to create proxy measures for all components of the Measurement Battery" (1978: 15). Finally, the

Activity Inventory uses a multiple-choice format to assess the quality and quantity of participation in out-of-class activities related to the six outcome areas (1978: 4).

McBer and ACT seem more concerned with the pragmatic claims of critical thinking than some earlier designers of critical thinking tests. Their instruments are more sensitive to criterion-referenced testing which seeks to relate critical thinking to successful adult role performances. This new interest in competency-based assessment may appear sloppy, but nonetheless offers promise of a new breakthrough in testing strategies.⁴

Sociology teachers need reliable knowledge about the critical thinking capacities of their students. The expertly-designed tests mentioned above provide some of the empirical tools essential to begin such a task. But these available tests do not exhaust the range of empirical instruments which might be used to assess critical thinking. Table 2 identifies several tests also relevant to the teaching of sociology. In many cases the original instruments use content items which are obsolete or of minimal value to sociology. But the testing formats and suggested strategies of inquiry are highly relevant. Sociologists can benefit from these earlier efforts as they revise and adapt instruments to their own specific questions.

The 26 tests of critical thinking mentioned in Tables 1 and 2 are highly diverse in form, degree of complexity, range of topics, and function. I will review briefly some of the salient aspects of this diversity.

Three types of testing formats are used. The first type is an open-ended essay form in which students are required to articulate their best thoughts with minimal direction from the test itself (Tests 3, 4, 5, 7, 12, and 23). Clearly the most extreme open-ended test reported in this review of literature is Oliver and Shaver's Social Issues Analysis Test #4 in which discussion groups grapple with a controversial issue (Test 24). The second type of format is the presentation of paragraphs, charts, tables, or graphs as complex stimuli; a series of structured questions (usually multiple-choice items) follow each complex stimulus (Test 1, 2, 6, 8, 9, 10, 15, 17, 19, 20, 21, 22, and 26). Simon's test

(Test 25) is both open-ended and prestructured. Students are asked to analyze a wide variety of magazine articles with a particular classification scheme. But this scheme is not a totally closed system and on some occasions, students may reject the adequacy of the scheme altogether. The third type of format is the least complicated for students, but sometimes frustratingly complex for test analysts (Levi, 1948). A series of one-sentence statements requires the student to respond according to a predetermined classification scheme (Tests 11, 13, 14, 16, and 18). Expected judgments are often dichotomized (e.g., a statement is either fact or opinion), but occasionally responses may be more elaborate (e.g., Test 11).

Earlier discussion of various definitions of critical thinking revealed numerous complex components in the term. It is not surprising to rediscover this complexity when one examines the many aspects of critical thinking being assessed by the 26 tests. The diverse topics cannot be classified easily by a typology of mutually exclusive and exhaustive categories. The tests are designed with various degrees of analytical rigor, and in some cases information is limited to a brief discussion and a few sample items in a research article. These limitations do not preclude a few general comments about various aspects of critical thinking being scrutinized by the test makers. Some tests are comprehensive measures of critical thinking (Tests 1, 2, 5, 6, 7, 9, 10, 21, 22, 23, and 24); other tests cover a more narrow range of topics (Tests 3, 4, 8, 11, 13, 14, 15, 16, 17, 18, 19, 20, 25, and 26). The most comprehensive tests are the Cornell Critical Thinking Tests and Ennis and Millman who attempt to cover more specific aspects of critical thinking in one 50-minute test than any other instrument in the literature. At the other extreme Morse and McCune provide an extensive battery of single-focused tests of critical thinking. Tables 1 and 2 also provide some indication of the topics most frequently covered by tests of critical thinking. Several tests attempt to measure the ability to make inferences from a given body of data, find assumptions, develop a consistent line of reasoning, define problems, and clarify value judgments. It seems highly probable that virtually all of the topics mentioned in

the 26 tests are found in Ennis's (1979) recent comprehensive statement of rational thinking.

Finally, a few comments about the functions of the 26 tests. In some instances researchers have used their instruments for norm-referenced testing (Tests 1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 21, and 26). But these same instruments can also be used for criterion-referenced testing. Another group of tests is intended exclusively for criterion-referenced testing purposes (Tests 7, 13, 14, 15, 16, 17, 18, 19, 20, and 25). Several tests have been used at one time or another as diagnostic instruments (Tests 1, 2, 7, 11, 13, 14, 15, 16, 17, 18, 19, 20, and 26). Baker has constructed a test for formative evaluation (Test 7) and many others can be used in this fashion. Finally, several tests are intended for summative evaluation (Tests 1, 2, 3, 4, 5, 6, 7, 9, 10, 12, 21, 22, 23, 24, and 26). Researchers have used several tests as pre/post test devices to measure the effectiveness of various instructional innovations (Tests 1, 9, 10, 12, 21, and 26). On the other hand, many tests are not sufficiently rigorous at this point for such research purposes. Given this diversity of formats, topics, and functions, sociology teachers must weigh thoughtfully the facets of critical thinking which concern them and the purposes of assessment which guide their inquiry.

**CRITICAL STRATEGY:
EVALUATING EXISTING TEST ITEMS**

Tables 1 and 2 do not exhaust the battery of tests available to sociology teachers. With monotonous regularity the mail brings new textbooks with accompanying test banks of quiz items. What do sociologists know about items in these test banks? Do these test items measure various aspects of critical thinking? What conceptual guidelines are needed to assess the critical thinking capacities of these items? Such questions deserve much attention. But educational researchers have not developed analytical schemes or operational techniques to systematically assess the existing test items. One of the few scholarly works to explore this problem is Richard Anderson's (1972) pioneering work on assessing tests of

comprehension. Using Bloom's taxonomy as a point of departure, Anderson appraises various techniques of pencil-and-paper tests essential to measure the first two levels of knowing: knowledge and comprehension. He points out that questions requiring responses (or even rearrangement of words) can never measure comprehension. Assessing comprehension requires such methods as paraphrasing questions, substituting particular terms for superordinate terms, substituting general terms for specific terms, identifying new instances which discriminate from non-instances. Using the scheme outlined by Anderson, I examined 100 items from Ian Robertson's (1977a) test bank. Robertson has authored the leading text on the market today, but on issues of critical thinking it is a dismal failure. The vast majority of items ask for verbatim responses; some items are misleading, and few used the range of testing possibilities outlined by Anderson. Robertson's test bank did not test students at the second rung on Bloom's ladder of higher mental processes (1977b).

Teachers of sociology need to develop a theoretical framework to assess the adequacy of critical thinking instruments. Guidelines are needed to evaluate existing sociology test items as to their usefulness in revealing capacity for critical thinking. Everett K. Wilson raises some helpful questions for such an endeavor (1979, personal correspondence). First, does the test item get at students' ability to take things apart, to see the elements (analytic skill)? Second, does the item measure students' ability to see connections, systemic properties (synthetic skill)? Third, does the test item get at students' ability to extend beyond the instant case or situation? That is, does it test skills in detecting analytically isomorphic populations or situations? Wilson's three questions are a valuable beginning; the analytical work of Ennis and others could be used to extend the list.

**CONSTRUCTIVE STRATEGY:
GUIDELINES FOR TEACHER-MADE TESTS**

I have now reviewed various tests which claim to assess critical thinking. I have also suggested that sociologists need to develop

an evaluation system to determine the extent to which commercial test banks assess critical thinking. A third strategy is suggested to improve the instruction of sociology. It is the need to develop theoretical criteria and operational techniques to construct teacher-made tests of critical thinking. I know of very little literature, except Tomko and Ennis (1979), which deals with this topic directly, but many books and articles have been written about teacher-made tests (McKeachie, 1969; Green, 1963; Gronland, 1968; Berg, 1965). This "how to" literature is not grounded in careful research; furthermore, it perpetuates a widely-accepted dichotomy among educators. It seems that both amateur teachers and testing experts insist on dividing all tests into two mutually exclusive categories: essay tests and objective tests.

Conventional discussion of teacher-made testing often seems to boil down to a basic dilemma. Essay tests are easy to construct, but hard to score. Objective tests, on the other hand, are hard to construct but easy to score. In McKeachie's words (1968: 127), "It appears to be true that examinations which are easiest to construct are the most difficult to grade and vice versa." Unfortunately, this mentality pervades the practical advice to teachers who need to construct their own tests. This is a false dichotomy which fails to acknowledge that the construction of meaningful tests is *always* difficult regardless of the formats used. Likewise, the generation of scores *never* escapes the thorny question of validity. Given the administrative purposes of testing in formal education settings and the constraints of time and resources for testing any given group of students, problems of determining a meaningful sample of worthwhile questions will always be on the teacher's desk. Teachers sample a small segment of a person's life experience and attempt to generalize to the outer limits of their data. Teachers' work is complicated further by attempts to measure the higher mental processes of critical thinking. This is the practical agenda of all teachers who design their tests. The issues are not clarified by either/or decisions about essay versus objective exams.

Knowledge needed to improve teacher-made testing must begin with the teacher's instructional objectives. If sociology

teachers claim to develop critical thinking in their classrooms, then they must devise some system of assessing learning outcomes commensurate with these claims. They also must determine what aspects of critical thinking they are promoting. Some aspects of critical thinking are more amenable to objective exams than others. For example, any teacher who is interested in the topic of inference can find many working models in the current collection of published critical thinking tests. Inference questions have two basic statements: one statement takes the form of a conclusion and the other takes the form of supporting evidence. Any number of objective testing possibilities can be designed around these two components. On the other hand, critical thinking tests which attempt to relate complex theories to novel problem-solving tasks are probably best assessed by essay questions.

Deliberate consideration of instructional objectives often leads to a second basic issue. Sociology teachers frequently claim that knowledge from textbooks and classroom lectures is relevant and transferable to nontextbook literature and to settings outside the classroom. In solemn voice, the clichés come easily to sociology teachers on the first day of class: "This textbook and my lectures will help you better understand events reported in the daily newspaper." If instructional objectives intend to provide useful learning opportunities for life experiences beyond the classroom, then testing devices should extend to those life experiences thought to benefit from formal education. If sociologists believe that their instruction will benefit all persons who read newspapers, then testing devices are needed to assess the student's newly-acquired capacities to better understand social knowledge found in newspapers. Such tests can be either objective or essay. For example, Robert Ennis has designed an objective critical thinking test which uses a simulated newspaper specimen (Moorburg Letter Test); I have designed several essay tests which make extensive use of journalistic sources (Baker, 1979). Testing learning outcomes which stretch beyond the immediate boundaries of textbook or lecture requires careful selection (or imaginative creation) of relevant case materials (Newmann and Oliver, 1967).

A case can be made against objective testing; Hugh Petrie (1979) has made the best case I have ever read. But his line of reasoning does not lead necessarily to an exclusive use of essay tests. There are other possibilities such as using both objective and essay formats in deliberate tension with each other. I have been using such a hybrid arrangement for several years. My rationale and testing device are pragmatic and straightforward. Prestructured answers leave no room for inquiry and often create the misleading notion that successful academic work is little more than agreeing with the Answer Man. Fools are consistently suckers for false answers while the wise student agrees with the test maker's choice. But somewhere between foolishness and wisdom there is a gray zone of intelligent doubt about specific options. According to Ennis, critical thinking is the ability to assess the adequacy of statements; it seems reasonable to extend Ennis's point by suggesting that the mark of a critically thinking student is the capacity to assess the adequacy of the teacher's question with its accompanying options. Perhaps for logically sound reasons, none of the options fits the question, or perhaps two options can be defended as logically sound choices.

Rather than insisting on an either/or position, I have developed a multiple-choice system which permits an open-ended response to the entire question/answer complex. After careful reflection on the question and the options, students may choose to offer a short written reaction which takes exception to the prearranged claim of measuring some principle of sound reasoning. In some cases these written responses may demonstrate incompetence on the part of the student; on other occasions the written comments call attention to the inadequacy of the test maker's line of reasoning or choice of words. In Dewey's terms, students are confronted with an obstacle, and are given the opportunity to intelligently overcome it.

Developing new teacher-made tests which probe issues of critical thinking may require a radical reassessment of conventional teaching/learning expectations. Sociology teachers may need to reexamine some of the basic premises of their instruction.

One would hope that such reassessment will lead to meaningful innovation. I am currently developing an alternative teaching/learning system for basic courses in sociology (Baker, 1975, 1979). One aspect of this work is the creation of several critical thinking tests. These tests are designed to assess two fundamental cognitive skills: (1) the ability to assess the adequacy of others' statements about social reality, (2) the ability to create logically sound statements about social reality. Mastery of these generic skills allows students to go beyond the immediate text literature of sociology. I require students to examine sociology critically by comparing its truth claims with those of common sense and journalism.

Students are never asked to accept sociological knowledge as a worthy end in its own right. They are never tested on their ability to memorize or comprehend conventional textbook knowledge. As rational thinkers students must scrutinize sociological ideas with the same critical tools used for all other domains of social thought. Sociology is not immune from critical assessment in the sociology classroom. In short, critical thinking begins when professional knowledge is denied the special privilege of passive, uncritical acceptance.

CONCLUSION

I would like to suggest a simple syllogism. Since assessment of learning outcomes should coincide with instructional objectives, and since the instructional objectives of many (if not most) sociologists include commitment to critical thinking, it seems reasonable to conclude that conscientious teachers would be ever mindful of the task of evaluating critical thinking in the classroom. Yet such is not the case. Sociologists talk a good line, but they come up short at delivery time. They seem satisfied with teaching/learning arrangements which demand little more of students than the replication of favorite concepts, famous names, and club-house quarrels among the elite.

Critical thinking is one of those buzz words which seems to mean just about anything to anybody. It comes in handy for rhetorical justification of sociology as a curricular enterprise, but the term rarely is examined as an analytical concept capable of empirical scrutiny. When one reviews the conceptual literature, it is clear that the best researchers and educators have been probing a complex set of logical operations and interdependent mental processes. Different aspects of critical thinking have been measured successfully and these measuring instruments get at crucial dependent variables in numerous educational experiments. Many educators have demonstrated repeatedly that instructional arrangements can be designed to promote the learning of critical thinking. These successful learning outcomes have been measured for the short run, but little is known about the long-run educational processes that make for critical thinking as a life-long intellectual orientation. Much work remains to be done. It will begin when teachers of sociology become conscious of the gap between what they say and what they do.

APPENDIX A

ADDRESSES OF CRITICAL THINKING TESTS CURRENTLY AVAILABLE

Cornell Critical Thinking Test, Level Z
 Robert Ennis
 Rational Thinking Project
 Education Building
 University of Illinois
 Urbana, Illinois 61801

Watson-Glaser Critical Thinking Appraisal
 Harcourt Brace Jovanovich
 New York, New York

College Outcome Measures Project (COMP)
Audrey Forrest, Director
The American College Testing Program
P.O. Box 168
Iowa City, Iowa 52240

Comprehensive Cognitive Assessment Battery (CCAB)
McBer and Company
137 Newbury Street
Boston, Massachusetts 02116

NOTES

1. In a provocative paper, Gregory Bateson (1972: 279-308) identifies three levels of learning, and suggests that this hierarchy is part of the larger scheme of all human communication.
2. This point is worth emphasizing. Critical thinking is not a residual category which includes all intellectual abilities above the level of rote memorization.
3. I say rarely because Max Black includes elementary symbolic logic in his influential book, *Critical Thinking* 1952.
4. Since most sociology teachers may be unaware of these expertly-designed tests of critical thinking, Appendix A provides addresses for further inquiry.

REFERENCES

- ANDERSON, R. A. (1972) "How to construct achievement tests to assess comprehension." *Review of Educ. Research* 42: 145-170.
- ANGELL, R. (1964) *Reasoning and Logic*. New York: Prentice-Hall.
- AUSUBEL, D. (1961) "Reception versus discovery learning in classroom instruction." *Educ. Theory* 11: 21-24.
- BAKER, P. (1979) "Dialectical teaching: an instructor's manual to social awareness modules." (unpublished)
- (1975) "Social awareness project." *Teaching Sociology* 3: 74-80.
- BAKER, P. and J. JONES (1979) "A case study of benign disruption in the classroom." *Teaching Sociology* 7: 27-44.
- BATESON, G. (1972) *Steps to an Ecology of Mind*. New York: Random House.
- BENDIX, R. (1970) *Embattled Reason: Essays on Social Knowledge*. New York: Oxford Univ. Press.

- BERG, H. [ed.] (1965) *Evaluation in Social Studies, Thirty-Fifth Yearbook of the National Council for The Social Studies*. Washington, DC: National Council for the Social Studies.
- BEYER, B. (1971) *Inquiry in the Social Studies Classroom*. Columbus: Charles E. Merrill.
- BIGGE, M. (1976) *Learning Theories for Teachers*. New York: Harper & Row.
- BLACK, M. (1952) *Critical Thinking*. Englewood Cliffs, NJ: Prentice-Hall.
- BLOOM, B. and E. RAKOW (1969) "Higher mental processes," pp. 594-601 in R. Ebel (ed.) *Encyclopedia of Educational Research*. New York: Macmillan.
- BLOOM, B., J. HASTINGS, and G. MADAUS (1971) *Handbook on Formative and Summative Evaluation of Student Learning*. New York: McGraw-Hill.
- BRADSHAW, T. and S. McPHERRON (1978) *Data Report for Issues and Resources in the Undergraduate Sociology Curriculum (ASA Projects on Teaching Undergraduate Sociology)*. Washington, DC: ASA Teaching Resources Center.
- BROWN, C. (1950) "An experimental diagnosis of thinking on controversial issues." *Speech Monographs* 17: 370-377.
- BRUNNER, J. (1961) *The Process of Education*. Cambridge, MA: Harvard Univ. Press.
- CHICKERING, A. (1976) "Developmental change as a major outcome," pp. 62-107 in M. Keeton (ed.) *Experiential Learning*. San Francisco: Jossey-Bass Publishers.
- COOK, A. and E. COOK (1950) *A Sociological Approach to Education*. New York: McGraw-Hill.
- COOK, A. and R. KOENINGER (1939) "Measuring learning outcomes in introductory sociology." *J. of Educ. Sociology* 13: 208-225.
- DENZIN, N. (1970) *The Research Act*. Chicago: AVC.
- DEWEY, J. (1933) *How We Think*. Lexington, MA: D.C. Heath.
- DOOB, L. (1950) "Propagandist vs. propagandees," pp. 439-458 in A. Gouldner (ed.) *Studies in Leadership*. New York: Harper & Row.
- DRESSEL, P. and L. MAYHEW (1954) *General Education: Explorations in Evaluation*. Washington, DC: American Council on Education.
- EBEL, R. (1978) "The case for norm-referenced measurements." *Educ. Researcher* 7: 3-5.
- EHRlich, H. (1961) "Dogmatism and learning." *J. of Abnormal and Social Psychology* 62: 148-149.
- ENNIS, R. (1979) "Logic, rational thinking, and education," *Philosophy of Education 1979: Proceedings of the Thirty-Fifth Annual Meeting of the Philosophy of Education Society Jerrod Coombs* (ed.) *Philosophy of Education Society*.
- (1975) "Children's ability to handle Piaget's propositional logic: a conceptual critique." *Review of Educ. Research* 45: 1-41.
- (1973) "On causality." *Educ. Research* 2: 4-11.
- (1969) *Logic in Teaching*. Englewood Cliffs, NJ: Prentice-Hall.
- (1964) "Operational definitions." *Amer. Educ. Research J.* 1: 183-201.
- (1962) "A concept of critical thinking." *Harvard Educ. Rev.* 32: 81-111.
- and J. MILLMEN (1971) *Manual for Cornell Critical Thinking Test, Levels X and Z*. Urbana, IL: University of Illinois.
- FORREST, A. (1978) *Comp: Annual Report 1978*. Iowa City, IA: American College Testing Program.
- (1977) *Comp: College Outcome Measures Project*. Iowa City, IA: American College Testing Program.
- FURST, E. (1950) "Relationship between tests of intelligence and tests of critical thinking and of knowledge." *J. of Educ. Research* 43: 614-25.

- GARDNER, W. (1970) "The current scene," pp. 1-10 in W. Gardner and F. Johnson (eds.) *Social Studies in the Secondary Schools*. Boston: Allyn & Bacon.
- GIROUX, H. (1978) "Writing and critical thinking in the social studies." *Curriculum Inquiry* 8: 291-310.
- GLASER, E. (1963) "Instructional technology and the measurement of learning outcomes." *Amer. Psychologists* 18: 519-521.
- (1941) *An Experiment in the Development of Critical Thinking*. New York: Teachers College, Columbia University.
- GREEN, J. (1963) *Teacher-Made Tests*. New York: Harper & Row.
- GRONLUND, N. (1968) *Constructing Achievement Tests*. Englewood, NJ: Prentice-Hall.
- HAAS, J. (1977) *The Era of the New Social Studies*. Boulder, CO: Social Science Education Consortium.
- HAMBLETON, R., H. SWAMINATHAN, J. ALGINA, and D. COULSON (1978) "Criterion-referenced testing and measurement: a review of technical issues and developments." *Rev. of Education Research* 48: 1-47.
- KEMP, C. (1962) "Critical thinking: open and closed minds." *Amer. Behavioral Scientist* 5: 10-15.
- LEE, A. and E. LEE (1939) *The Fine Art of Propaganda*. New York: Harcourt Brace Jovanovich.
- LEVI, A. (1948) *General Education in the Social Studies*. Washington, DC: American Council of Education.
- LOGAN, C. (1976) "Do sociologists teach students to think more critically?" *Teaching Sociology* 4: 29-48.
- LYLE, E. (1958) "An exploration in the teaching of critical thinking in general psychology." *J. of Educ. Research* 52: 129-133.
- MAHAN, T. (1955) "Human judgment: can the classroom improve it?" *J. of Educ. Research* 49: 161-169.
- McBer and Company (1978) *Users Guide for the Comprehensive Cognitive Assessment Battery (CCAB)*. Boston: Author.
- McKEACHIE, W. (1969) *Teaching Tips: A Guide for the Beginning College Teacher*. Lexington, MA: D.C. Heath.
- MORSE, T. and G. McCUNE (1964) *Selected Items for the Testing of Study Skills and Critical Thinking*. Washington, DC: National Council of the Social Studies.
- NEWMANN, F. and D. OLIVER (1967) "Case study approaches in social studies." *Social Education* 31: 108-113.
- OLIVER, D. and J. SHAVER (1966) *Teaching Public Issues in the High School*. Boston: Houghton Mifflin.
- (1964) "Teaching students to analyze public controversy: a curriculum project report." *Social Education* 28: 191-195.
- OSBORN, W. (1939) "An experiment in teaching resistance to propaganda." *J. of Experimental Education* 8: 1-17.
- PETRIE, H. (1979) "Against 'objective' tests: a note on the epistemology underlying current testing dogma." (unpublished)
- POPHAM, W. (1978) "The case for criterion-referenced measurements." *Educ. Researcher* 7: 6-10.
- RICE, T. (1978) "Cognitive mapping: small groups technique in teaching." *Teaching Sociology* 5: 259-274.

- ROBERTSON, I. (1977a) *Sociology*. New York: Worth Publishers, Inc.
- (1977b) *Test Bank to Accompany Robertson Sociology*. New York: Worth.
- RUSSELL, D. (1960) "Higher mental processes," pp. 645-661 in C. Harris (ed.) *Encyclopedia of Educational Research 3rd Edition*. New York: Macmillan.
- RUST, V., R. JONES, and H. KAISER (1962) "A factor analytic study of critical thinking." *J. of Educ. Research* 55: 257-265.
- SCRIVEN, M. (1976) *Reasoning*. New York: McGraw-Hill.
- (1967) "Methodology of evaluation," pp. 40-66 in R. Tyler, R. Gagil and M. Scriven (eds.) *Perspectives of Curriculum Evaluation*. Chicago: Rand McNally.
- SIMON, D. (1977) "The ideologies of American social critics." *J. of Communication* 3: 40-50.
- SJOBERG, G. (1973) *Instructor's Guide in Cities: Where People Live and Why*. Boston: Allyn & Bacon.
- SMITH, E. and R. TYLER (1942) *Appraising and Recording Student Progress*. New York: Harper and Row.
- STEWART, B. (1978) *Reviews of Selected Tests in Critical Thinking Test File*. Urbana, IL: Illinois Critical Thinking Project.
- TOMKO, T. and R. ENNIS (1979) "Evaluation of informal logic competence." (unpublished)
- TURNER, S. (1979) "The concept of face validity." *Quality and Quantity* 13: 85-90.
- TYLER, R. (1936) "The relation between recall and higher mental processes," pp. 6-17 in C. Judd (ed.) *Education as the Cultivation of the Higher Mental Processes*. New York: Macmillan.
- University of Illinois (1979) College of Arts and Sciences "Final report on the requirement in general education."
- WHITLA, D. (1977) *Value Added: Measuring Outcomes of Undergraduate Education*. Cambridge, MA: Office of Instructional Research and Evaluation, Harvard University.
- WILSON, E. (1979) Personal letter to author (July 1979).
- (1970) "Inductive methods in teaching sociology," pp. 113-120 in M. Krug, J. Poster and W. Gillies (eds.) *The New Social Studies*. Itasca, IL: Peacock.