Teaching With Testing:
Some Notes on Evaluation of Students¹

Peter Lehman
Department of Sociology
University of Southern Maine

Testing is one of the most important things that we do. Just ask your students! And they are right, although not necessarily for the right reasons.

For students, evaluations of their performance are crucial because such evaluations determine the bottom line—their grade for the course, their GPA and their future life chances. No matter what we may say and even believe about substantive learning, the joy of knowledge, personal growth, etc., the grade is a tangible and consequential outcome for students. We need to be careful about how we arrive at it.

For faculty, evaluation is equally crucial because it is where the goals and objectives of a course are put into concrete practice—operationalized. If we want students to think for themselves, then that's what we need to concretely assign and assess. Tests, and other forms of evaluation, tell students *how they should learn* as well as *what* they should learn.

Yet we spend more time thinking about the *what* than the *how.* We spend relatively little time on designing our evaluations of students—on deciding about appropriate forms of evaluation and their content. Most of the faculty I know quite frequently spend less time writing an examination than they spend developing a single class lecture.

The same neglect is evident in our professional discourse about evaluation, and especially traditional testing. Years go by without any articles in *Teaching Sociology* on evaluating students. The rare or article on evaluation addresses "different" or "innovative" approaches to teaching and evaluating students -- the use of journals, observation assignments, etc. These are sometimes valuable but almost all of the teachers I know, including those who use alternative approaches, nonetheless rely heavily on traditional examinations (tests).

"Traditional examinations" come in three main flavors: out-of-class essays (take-home exams), in-class essays and in-class "objective" tests. You can look a long time for articles on these -- the literature by sociologists for sociologists is extremely sparse. Even the ASA Teaching Resource Center booklet on "Methods of Evaluating Student Performance" (Turk, 1982) generally ignores the traditional examination. An exception to this neglect is a very helpful and thoughtful chapter on "Evaluating Student Achievement" in Goldsmid and Wilson (1980).

The most important thing I have learned is that *testing teaches*—however we test and whatever we do—and that the challenge is to consciously integrate and *use* this fact. The more important students think tests are, the more important tests are in shaping the

learning experience. The following essay explores some of the dimensions of teaching with testing with an emphasis on testing in the introductory course.

An Example

So you want to test your students. You whip out your trusty pencil or plunk down at your word processor and write:

Where was Albion Small born and raised?

    a. Chicago, Illinois

    b. Whitfield, Kansas

    c. Muncie, Indiana

    d. Buckfield, Maine

    e. Des Moines, Iowa

So much for testing. No mystery. This is a legitimate and important examination question on Albion Small. C. Wright Mills (1942) thought that where Albion Small was born and raised had enormous consequences for several generations of sociology.

But consider the above question again. It provides us with an excellent example of how we might think through questions before they reach our students. Every question should be examined closely by asking ourselves "***Why*** am I asking this question and what do I hope to accomplish?" First, are we really concerned about where Small was born? (It's Buckfield, Maine—I had to look it up.) The important point for Mills is that it was small town rural America, not whether it was Iowa or Maine. Second, are we really concerned with Small at all? How about:

According to Mills, where were most of the early American sociologists born?

    a. Western Europe

    b. New York City

    c. small rural towns in the United States

    d. urban areas in the Northern United States

    e. none of these

The difference between the first and the second question transcends issues of whether they are fair, too detailed, etc. The two questions ask for different things. And the difference will profoundly affect the way our students read and study *every single future word* in our course. If we ask the first question, we can't complain that students can't see the forest for the trees—we are evaluating them on the trees. The second version at least has the virtue of being a major point rather than a minor piece of evidence.

But even the second version merely tests a bit of knowledge extracted from its context of meaning. It is still a memory question, clearly and directly answered in Mills' article. We can further improve the question by asking for comprehension of Mills' point:

Mills argues that an important source of anti-urbanism in American sociology is

   a. neglect of the importance of demographic changes.

   b. the racism of early sociologists.

   c. that most early sociologists were white.

   d. the small town rural backgrounds of many early sociologists.

   e. all of these.

This version addresses student understanding of the *logic* of Mills' argument. It *teaches* students to look for *why* the author is discussing a particular point and *how* it connects to the overall argument being developed. Now we are testing for *higher level cognitive skills* rather than scattered bits of knowledge. In the process of testing for these higher level skills, we are rewarding and encouraging their mastery. In addition, when we test this way we are creating an opportunity, when we carefully review the examination in class, to discuss how to reach this sort of understanding—foreign territory for most of our introductory students. *Tests should* not only evaluate but *instruct and motivate* as well (Sanders, 1966:1).

Testing As Research

It helps to think of testing as a form of research—we are trying to find out something. Unfortunately, this insight is often followed by discussions of how to make testing instruments valid and reliable, scaling, and other such technical considerations. These issues are probably important but often miss the point. As Hedley (1978) puts it, these considerations are to ensure that "we will be measuring what we intended to measure." "Traditional approaches to disciplined inquiry in test design have tended to focus on optimizing the *measurement efficiency* of tests rather than on optimizing their *instructional efficacy*" (Nitko, 1989:448, emp. added). *First* we need to figure out what we are intending to measure -- and whether measurement is our primary objective.

Testing is evaluation research—we are attempting to find out how well goals and objectives have been met. And one of the most difficult and productive parts of evaluation research is clarifying the goals and objectives of the research—the goals and objectives that will be evaluated (Glaser, 1988). It seems to me that the most important implication of viewing testing as research is that it reminds us to carefully and systematically conceptualize what we are doing, how, and why. Indeed, testing is intimately related to sociological inquiry and an integral part of what we are trying to teach students. Goldsmid and Wilson nicely conceptualize this goal and its connection to teaching:

> The goal is reliable knowledge about the social world, sought by the best methods available. The stress, that is to say, is on the seeking, not the absorbing. We seek answers, however tentative. Answers imply questions. Questions are what testing is about. (1980:326)

Testing is not just evaluation research, it is *formative evaluation* research— research aimed at improving or changing the extent to which goals and objectives are met. Tests do not stand alone, they are part of an ongoing course and courses are part of an ongoing process of trying (sometimes desperately) to educate. All evaluation has an

effect on the activities being evaluated—influences the way things are done, the way people feel, what they value, etc. Evaluation reinforces some activities and accomplishments at the expense of others.[1] Formative evaluation makes this effect a conscious goal. So should testing.

I have already contended that our tests—regardless of how they are constructed or the form they take—convey messages and teach students. Therefore, make this conscious. Integrate testing into the course. Use it as a teaching tool in as many ways as possible. How? This depends on what we hope to accomplish.

Cognitive Goals and Objectives

In 1956, as part of a larger project by the College and University Examiners, Benjamin Bloom and others developed a taxonomy of the cognitive domain—"the cognitive area of remembering, thinking and problem solving" (1956:2). They classified educational outcomes in order to assist teachers in clarifying objectives and in thinking about how to evaluate or measure outcomes. Bloom, *et al*'s original work came complete with sample test questions illustrating the various cognitive levels. The scheme has remained extraordinarily useful in discussions of goals (e.g. Vaughan, 1980; Wagenaar, *et al*, 1982, Mosher, 1989). A summary is presented in Table 1.

Table 1:
The Cognitive Domain: Bloom et al.

Level	Description
1. Knowledge	recalling specifics and ways of dealing with specifics, remembering universals, abstractions, generalizations, and theories.
2. Comprehension	understanding material so that one can translate it, interpret it, or extrapolate from it.
3. Application	applying abstractions (ideas, rules principles, theories) in particular situations.
4. Analysis	identifying elements and parts, relationships, and organizational principles.
5. Synthesis	forming wholes from elements and parts, assembling them into a new pattern.
6. Evaluation	judging the value of material and methods in relation to a given purpose.

Source: Bloom *et al*, 1956 with descriptions
as summarized by Vaughan, 1980: 270.

The hierarchical nature of the schema is quite important in designing teaching strategies. The model builds from the "simple" to the "complex." The abilities and skills of one stage "make use of and [are] built on the behaviors found in preceding" stages (1956:18). Thus, comprehension requires knowledge, application requires knowledge and comprehension, and so on. This hierarchy provides us with clues about how to develop higher-level skills and warns us against unrealistic expectations.

The hierarchical nature of the schema doesn't imply a positivist epistemology—fact then interpretation. "Knowledge" here refers to recall of information presented in the text, in class or elsewhere. Sanders (1966) renames this category "memory" to clearly denote that the category refers to material which the student remembers, or finds, rather than creates. Knowledge, in this sense, is a lower level cognitive skill but fundamental to all other functions. For example, in the treatment of crime rates students may be presented with information about how they are constructed and their problematic nature. This information about the social construction of crime rates then becomes "knowledge" in the sense that you want them to be able to remember and recite it as reflected in the following question:

The Uniform Crime Report

   a. is based upon extensive interviews with a national sample of crime victims.

   b. includes crime events which are known to and recorded by the police.

   c. is a representative sample of all crimes committed in the United States.

   d. only includes crimes for which a person was arrested.

   e. all of these.

   f. none of these.

In the case of the crime rates, *comprehension* means the ability to translate the knowledge into their own words and/or interpret the knowledge. More concretely, a comprehension question might be:

The Uniform Crime Report produced by the FBI is titled "Crime in the United States." Write a better (more descriptive) title.

Or, in a multiple choice format:

The Uniform Crime Report produced by the FBI is titled "Crime in the United States." A better (more descriptive) title would be:

   a. "Criminal Behavior in the United States."

   b. "Reported Crimes in the United States."

   c. "Estimates of Crime in the United States."

   d. "Police Recorded Crime in the United States."

e. all of these.

Both these questions build on *knowledge* but, in addition, require *comprehension*. *Application* is the next step—taking understanding of the ideas and using them in a concrete instance. For example, consider the following question:

In August, 1990, the Portland, Maine, police released their UCR statistics for the first half of 1990. According to the news report, rapes decreased 30 percent compared to 1989—they decreased from 23 in the first half of 1989 to 16 in the first half of 1990.

a. There were fewer rapes in Portland in the first half of 1990 compared to the first half of 1989.

b. This shows that the risk of rape in Portland has declined.

c. This helps demonstrate the effectiveness of rape avoidance education programs for women in Portland.

d. This probably reflects efforts by the police department to increase responsiveness to rape victims.

e. All of these.

f. None of these.

Many of the questions we ask, and the objectives we *actually* implement and teach in the introductory course stay at the first level, knowledge. Only at the next level do Bloom *et al* consider the objectives "intellectual" and I assume that most of us agree. Certainly, our course objectives tend to go past the simple conveyance of a body of knowledge (Vaughan, 1980:268). We want students to develop critical thinking, reflective thinking or problem solving skills. But our tests often do not get that far.

Our textbooks often do not help either. As Sanders puts it:

Textbooks create a major problem for teachers concerned with composing good questions. …*the textbook is weak in that it offers little opportunity for any mental activity except remembering*. If there is an inference to be drawn, the author draws it, and if there is a significant relationship to be noted, the author points it out. … The result is that the creative process and the controversy of competing ideas are hidden from the student. (1966: 158)

We need to make sure that the questions we ask, and other assignments we may give, include the "intellectual abilities and skills" described in the Bloom typology. While remembering that knowledge is a necessary basis, questions and other assignments should emphasize higher levels in the typology—levels that require a student to make active *use* of knowledge. "This begins with a consideration of the forms of thinking which are appropriate for the course and a decision to place new emphasis on certain kinds of questions." (Sanders, 1966: 155)

This does NOT mean that our questions or assignments should focus only on these higher levels. We can, but probably should not, create synthesis and evaluation assignments which call for a knowledge base and skills not possessed by many of our

students, especially at the introductory level. Students need to *learn* to accomplish these higher level skills and we need to encourage and nurture their development. It is easy to make assignments which are more difficult than they appear—that call for skills beyond what our students possess. In this case we set ourselves up to be disappointed and set our students up to fail. Consider an assignment to

> Critically evaluate Weber's model of bureaucracy in understanding the operation of the local Burger King restaurant as you observed it.

This assignment actually calls for all six of the cognitive levels. The student is being asked to form and present a reasoned judgment about a set of ideas in relation to a standard or value developed by the student.[2] The student is being asked to take observations and not only identify elements and parts but to discern patterns and essences—to answer the question "What's going on here?" at Burger King. Answering this question calls for analysis and synthesis. It also requires that the student deal with values in deciding what is *important* in what's going on. In addition, the student is being asked to understand Weber's model and analyze how its elements apply. Finally, the student is being asked to judge how well Weber's model gets at what's going on and give reasons for her or his judgment. Overall, the assignment is probably out of the reach of most of our students and certainly out of the reach of our introductory students. Again, students need to *learn* to accomplish these higher-level skills. We need to encourage and nurture their development.

The Bloom typology can help us to structure questions or assignments to help this development process. "Teachers can lead students into all kinds of learning through careful use of questions, problems and projects." (Sanders, 1966:2) For instance, have students do an assignment in steps—knowledge, comprehension, etc. An elementary example:

> In your own words, what is a group? [comprehension] Identify a group that you belong to or have belonged to recently. [application] Give reasons why this should be considered a "group" in the sociological sense. [analysis] Give an example of a norm distinctive to this group. [application] Give reasons why you think this should be considered a "norm" in the sociological sense. [analysis]

Various types of questions can be structured. Here is an example from a recent student "Study Guide" (Mosher, 1989):

> Name some assumptions of mass society theorists. [knowledge] Explain Milgram's research. [comprehension] Show how Milgram's results did or did not support mass society theory. [analysis]

One of my favorite techniques[3] is a series of true-false questions. The following series moves from knowledge of a concept to various dimensions of comprehension of the concept, including its implications.

> The concept of a role:
>
> **T** F   refers to a collection of norms specific to a status in society. [knowledge]
>
> T **F**   refers to a collection of values. [knowledge]

T **F**   refers to a script that says how a particular person is supposed to act at all times. [comprehension]

T F   is illustrated by a statement such as "Police officers must not drink on duty." [application]

T F   helps explain why the same person may act very differently at different times of day. [analysis]

This same sort of technique can be extended to *illustrate as well as test* the reasoning process by using a concrete statement of information followed by a set of questions. This can become quite complex and sophisticated. The following questions actually model some of the steps in a structured critique:

> In class, we examined the relationship between getting drunk and where people live (urban, rural, etc.). We found that it didn't make a significant difference where they lived—the proportion reporting getting drunk is very similar.

T F   The dependent variable here is getting drunk.

T **F**   The thing we are trying to explain here is where people live.

T **F**   A sociologist would suggest that whether a person lives in a small town or a city determines their drinking behavior.

T F   Our results are not consistent with what mass society theorists would expect.

T **F**   This research tends to support Durkheim and Morselli's contention that modernization and urbanization lead to increased social isolation and deviance.

T **F**   We might be able to explain these results by suggesting that people who have drinking problems tend to become poor and homeless and move to urban areas.

Enhancing the Process

The questions we ask, then, are part of what we are teaching. But the initial asking is only one part of the process. Whatever we do, we must go over it carefully and thoroughly in class afterwards. "It is desirable to follow through on a test, using it as a means of review, as a diagnostic instrument, and as a tool for filling lacunae and remedying deficiencies" (Goldsmid and Wilson, 1980:326). The more carefully we have conceptualized what skills we are asking for, the more clearly we will be able to help students understand how to get there. This takes time, but "if we are to raise questions, as we do in examinations, they had better be worthwhile ones. And if they are, in fact, worthwhile ones, then they merit dwelling on" (Goldsmid and Wilson, 1980:326).

On closed-ended questions (no one who has ever discussed an exam with a class could ever call them "objective") I make overheads of the test and spend at least a whole class period reviewing them. This allows me to clear up confusions, explain things a different way, and to coach students in needed skills. I also try to emphasize how questions—or rather the issues and ideas *in* the questions—relate to one another.

Be responsive (rather than defensive) in discussing test items. Listening to how students thought about issues helps us correct them—or correct ourselves if the question is invalid or ambiguous. Neither answering questions nor asking them is an easy task; *this is a lesson well worth teaching*. David Heise includes a form, attached to the exam, for students to vote to drop ambiguous or especially difficult questions; he then drops items voted against by 25 percent of the class (in Goldsmid and Wilson, 1980:324). I prefer the chaos of doing this in class, in part because it pushes students (and me) to clarify and articulate the issues—a teaching/learning experience.

This process of review does more than just extend the cognitive learning process. It should, if well done, enhance reliability and validity and convey to students a sense of fairness and equity. There is good evidence that students' perceptions of evenhanded evaluation are related to enhanced student effort and that how results are communicated has a strong bearing on how students perceive the evaluations (Natriello and Dornbusch, 1984). When students feel they are being dealt with fairly, they respond. Students know respect when they encounter it.

Along these lines, don't ambush students. The clearer idea students have about what they will be asked to do, the better their preparation. Remember that their review and preparation—their trying to pull things together and make sense of them—is one of *our* teaching goals. Again, their sense of fairness—evenhandedness and validity— affects how students perceive evaluations and increases their effort.

Evaluations also need to be fairly frequent. Studying student performance and motivation, Natriello and Dornbusch found that "frequent and challenging" teacher evaluations of student work increases student effort (1980: 144). This applies to *all* students; "even students with low levels of skills seem to benefit from more frequent and challenging evaluations" (*Ibid.*). This research finding is consistent with the only good advice I got as a novice teacher: evaluate early and evaluate often. The earlier we test, the earlier we communicate our expectations to students and the earlier we begin the process of teaching through testing and feedback. The more often we test the more review we give students and the more feedback everyone gets. In addition, frequent evaluation lowers anxiety.[4]

Along with frequent evaluation, consider open book and open note tests. Why not? This is the most articulate way to say, "*Understand* what's going on rather than memorizing it." It allows us to more effectively test for understanding and application. It enhances students' perceptions of fairness. And, finally, it's nice to eliminate at least one form of cheating simply by making it legal.

Choosing the Format

With all of these things in mind, what are the best formats for evaluating? It depends, of course, upon our objectives.

Closed-ended questions (such as multiple choice, true-false, completion or matching questions) are excellent in assessing knowledge, comprehension and application, and can be very good at assessing analysis. They have the advantages of being able to more fully sample different areas of the course as well as isolating subject

specific skills from general language skills. Closed-ended questions are not really effective in assessing synthesis and evaluation (including organization and creativity).

Interestingly, closed-ended questions have been shown to be extremely effective in assessing the ability to work with novel problems (Chase, 1978; Thorndike and Hagen, 1969). The multiple choice question about local rape rates, discussed above, is an example of this potential—it presents new information and asks students to work with it using their knowledge and understanding of the materials studied.

Unfortunately, good closed-ended questions are difficult and time-consuming to write and it is easy to fall into the trap of simply testing bits and pieces of knowledge. "It requires much thought and time to construct a battery of [multiple choice] test items that adequately sample a sector of sociology" (Goldsmid and Wilson, 1980:322). However, they are much more reliable, create more sense of evenhandedness among students, are a clearer focus for review and class discussion, and they allow prompt return—all of which contribute to the learning/teaching enterprise.

Open ended questions are more effective than closed ended questions in assessing analysis, synthesis and evaluation. They can allow students to explore territory on their own and integrate diverse materials. Open-ended questions permit creativity and self-direction. But, this makes responses to open-ended questions difficult to consistently evaluate, in part because "students may not tackle the same problem—or the same aspects of a given problem" (Goldsmid and Wilson, 1980:324). As a result, they are generally not as effective for assessing basic knowledge, and weak in assessing comprehension and application (Chase, 1978; Thorndike and Hagen, 1969). As Bloom, *et al* (1971:149) put it, they "are easier to write but more difficult to score."

The weakness of open-ended questions is especially acute because general facility with language can mask lack of mastery of the specific subject. Students who write well and are facile with language will generally do well on essays even though their mastery of the material might not be superior. In this sense, essay questions can encourage bluffing and it is very difficult for the teacher to call the bluff. It is often unclear how much the student really knows and understands.

The assignment concerning the Uniform Crime Reports (above) exemplifies some of these problems.

> The Uniform Crime Report produced by the FBI is titled "Crime in the United States." Write a better (more descriptive) title.

Responses to this open-ended question will be extremely difficult to evaluate. The multiple choice version of this question explicitly polled various misunderstandings. For instance, the title "Estimates of Crime in the U.S." is an inferior *choice* in the closed-ended question but how would you evaluate it in an open-ended context? And what about the student who writes: "The FBI Does It By the Numbers"? This is a wonderful title but does the student know and understand the material?

One obvious strategy is to turn this into a short essay question by adding "Explain." at the end. This asks the student to defend their title.[5] But how would you evaluate the following response?

"The FBI Does It By the Numbers" is a good title because it emphasizes the extent to which the FBI, and other law enforcement agencies, use the statistical information in the UCR to create public concern, expand budgets and convince us that there is a serious problem with crime in the United States.

Again, does the student know and understand the material?

Another strategy might be to simply ask "What does the Uniform Crime Report count and how does it count it?" This more straightforwardly tests knowledge. It limits bluffing and allows more consistency in evaluation although it does not eliminate either of these problems. Most seriously, however, it doesn't really take advantage of the open ended format at all since it doesn't require any of the higher level intellectual skills.

Despite all these problems with essays, there are still very good reasons to use them, particularly to complement other types of evaluation. The Bloom typology and the discussion thus far has only addressed cognitive goals. There are other goals, such as improving writing skills. Writing practice is an important foundation for further courses; it is also an important tool in developing higher-level cognitive skills.

One approach is to use essays in conjunction with closed-ended questions— probably with less weight placed on the essays. This approach allows evaluation of higher-level skills while separately assessing the student's knowledge and comprehension base.

Unfortunately, many essay questions used in introductory courses aren't very appropriate to meeting these goals of teaching effective writing and higher level cognitive skills. Effective essay questions, particularly at this level, should be fairly concrete and help the student structure her/his discussion. For instance, the question might begin by asking the student to summarize or relate something ("Summarize Ofshe's research") before asking for the higher-level analytic task ("and explain how this research tells us something about the importance of attachments."). Clarifying our goals helps us to clarify the assignments.

If our goals are developing language and thinking then we need to give students time to think and write. Essays should not be hastily written. Good writing requires drafting and redrafting, clarifying ideas and thinking through issues, as well as clarity of style. This is not done on the spur of the moment in class. "Trial by ambush" essays— the traditional model for in-class exams—teach students all the wrong things about writing, including a tendency towards fluff.

One way around this dilemma without assigning take-home essays (which are simply too large and unwieldy for introductory students), is to give them the essay questions ahead of time. I do this in a number of courses. Typically, I give students two or three essay questions to prepare ahead of time. At the exam, I will choose one of these for them to write on (or actually transfer into a bluebook if they have prepared a full draft). This approach keeps the exam manageable (they still have to write the essay down in a limited time) while allowing ample opportunity not only for drafting and thinking, but for class discussion of ideas and strategies before the exam.[6]

Another approach is simply to "test" using closed-ended questions and use out-of-class assignments to promote writing and foster higher-level intellectual skills. These assignments can ask students to actively connect concepts from the course with their own observations or experiences. We need to be careful not to overwhelm the student—as in the Burger King assignment—and provide structure.

In sum, I argue that closed-ended questions are the appropriate primary **evaluation** tool at the introductory level but other formats are important to complement and expand this primary evaluation. Carefully done, closed-ended questions are the best tool for assessing knowledge, comprehension and application. They are not as good a vehicle for developing skills in analysis, synthesis and evaluation. They do not really assess or reflect capacity for imagination or creativity and, obviously, they do little to promote or develop language skills.

To address these objectives, other tools need to be used. The traditional examination essay is one possibility but, given clearly defined objectives and the clearly defined role of closed-ended questions, other approaches may be even more appealing—alternative, creative or innovative techniques such as student journals (e.g. Wagenaar, *el al*, 1982) and take-home exercises and assignments. The most successful of these are exercises which ask students to *do* sociology.

The mix of methods each of us chooses to evaluate and teach will differ. But however we do it, we need to carefully define what we are trying to accomplish and carefully evaluate whether we are choosing the right tools.

Notes

1. Glaser (1988:15) makes this point in terms of "people-changing institutions" in general. Natriello and Dornbusch (1984), in their sociological study of teacher evaluation, consider the "evaluation of students by teachers" as a "primary mechanism" for encouraging and shaping student effort.

2. The student must both develop the standards and apply them -- decide how well they are met -- for a question or assignment to be considered "evaluation." If the standard is given, then the exercise is really interpretation or application, not evaluation. "If no standards are offered in the question or if the standards are only suggestive and require refinement in order to [apply them], then the question is classified as evaluation." (Sanders, 1966: 142)

3. A technique that I first encountered in Rod Stark's introduction to the test bank in the first edition of *Sociology* (1985). The following example is taken from Stark's original test bank.

4. Probably, this can be overdone. The weekly quiz tends to focus on knowledge alone and, unless brilliantly done, tends to detract from discussion of "what's going on here?"

5. This addresses the "most complex behavior" under *Application*: stating the reasoning enployed in support of a conclusion. (Bloom, *et a*, 1971:176)

6. When I do this in an introductory level course, the essay is only a portion (typically 30%) of the exam. The rest is close-ended. This strategy allows me to be demanding of the essays without demolishing the students. The two parts also allow me to check for consistency of result and, in part, distinguish between serious writing problems and problems of basic knowledge and application.

References

Bloom, Benjamin S., M. Engelhart, E. Furst, W. Hill and D. Krathwohl. 1956. *Taxonomy of Educational Objectives, Handbook I: Cognitive Domain.* New York: David McKay.

Bloom, Benjamin S., Thomas J. Hastings and George Madeaus. 1971. *Handbook on Formative and Summative Evaluation of Student Learning.* New York: McGraw-Hill.

Canglelosi, James. 1990. *Designing Tests for Evaluation of Student Achievement.* New York: Longman.

Chase, Clinton. 1978. *Measurement for Educational Evaluation.* Beverly Hills, CA: Sage.

Glaser, Daniel. 1988. *Evaluation Research and Decision Guidance for Corrections, Addiction-Treatment, Mental Health, and Other People-Changing Agencies.* New Brunswick, NJ: Transaction.

Goldsmid, Charles and Everett K. Wilson. 1980. *Passing On Sociology: The Teaching of a Discipline.* Washington: American Sociological Association.

Hedley, R. Alan. 1978. "Measurement: Social Research Strategies and Their Relevance to Grading." *Teaching Sociology.* 6:21-29.

Mills, C. Wright. 1942. "The Professional Ideology of Social Pathologists." *American Journal of Sociology.* 49:

Mosher, Carol. 1989. *Study Guide for Sociology, Third Edition, by Rodney Stark.* Belmont, CA: Wadsworth.

Natriello, Gary and Sanford Dornbusch. 1980. *Teacher Evaluative Standards and Student Effort.* New York: Longman.

Nitko, Anthony J. 1989. "Designing Tests That Are Integrated with Instruction." Pp. 447-474 in Robert Lin, ed. *Educational Measurement*, Third Edition. New York: American Council on Education, MacMillan.

Sanders, Norris. 1966. *Classroom Questions: What Kinds?* New York: Harper and Row.

Stark, Rodney. 1992. *Sociology*, Fourth Edition, Belmont, CA: Wadsworth.

Thorndike, Robert and Elizabeth Hagen. 1970. *Measurement and Evaluation in Psychology and Education.* New York: Wiley.

Turk, Theresa G., ed. 1982. *Methods of Evaluating Student Performance.* Washington: American Sociological Association Teaching Resources Center.

Vaughan, Charlotte. 1980. "Identifying Course Goals: Domains and Levels of Learning." *Teaching Sociology.* 7:265-279.

Wagenaar, Theodore, Judy Corder-Bolz and Ingeborg Knol. 1982. "Student Journals in Sociology." pp. 1-15 in Turk, ed., 1982.