# The Test as a Teaching and Learning Tool

*VINCENT J. CUSIMANO*

WHAT DOES A GRADE on an examination really mean? Unfortunately this question is asked more frequently by students than by teachers. It is time we as educators asked the same question.

The pressures of mass education have forced teachers to use tests simply as marking tools—as a measure of what the student knows. Most examinations rely on knowledge, the lowest level of a student's cognitive development. Although this is probably the easiest type of examination to construct, the educational growth of a student is not dependent on knowledge alone but rather on the comprehension and ability of a student to apply, interpret, and analyze. This is especially true in science education. In testing only knowledge, we fail to make the examination an integral part of the total learning experience of the individual. If an examination is based on all cognitive levels, the grade takes on meaning. If we further analyze each question on the examination, we can even diagnose failure and prescribe appropriate remedies, making the test a powerful teaching and learning tool.

## *Grouping Major Concepts*

In constructing an examination of this type, the material is divided into major concept areas. For example, for an examination on respiration for a life science class (a high school general biology course), the learned concepts may be divided into four main groups:

1. Burning a substance produces carbon dioxide, water, and energy.
2. Burning and respiration are similar processes.
3. The rate of respiration, and therefore product formation, depends on the activity of the organism.
4. Different foods vary in their energy (calorie) content.

Within each group, list the facts that the student should know to properly understand the concept. For example, a student must know the definition of calorie to understand concept 4. This procedure is followed for each concept until all the important facts are tested under each concept. In the test I constructed there were a total of 23 fact questions for the four concepts. These questins comprise that part of the examination called knowledge.

The next part of the examination can be described as comprehension. Each concept is tested by a student's ability to interpret or extrapolate from his knowledge. As an example, the student should be able to predict the decreased formation of carbon dioxide following a slowing of the respiratory rate (concept 3).

A third part of the examination includes a number of questions testing application of these concepts. By application, a student will be able to use these concepts in a new situation that may differ from previous experiences to a greater or lesser extent. Having previously calculated the amount of carbon dioxide produced during a walking activity, the student might be required to predict the amount of carbon dioxide produced during skiing (concept 3). This would constitute a new situation with a small variance from a previously experienced situation. A widely varying situation might be posed by having a student suggest a reason for the greater condensation of water vapor on the windows of a gymnasium when students are playing basketball than when they are sitting down. The degree of application is often governed by the level of sophistication of the student. For the students in my life science group, I would not go much further than the former question, reserving the latter for a more sophisticated group of students.

I usually conclude the examination with a summary question which requires the student to either analyze, synthesize, or evaluate those concepts being tested. A question of this type requires the student to use all the concepts as an integrated whole and to examine the interrelationship of these concepts. This question generally takes the form of a problem situation in which all concepts are present. The student is required to explore these interrelationships and sometimes make evaluative judgments.

An examination of this type is extremely time-consuming in construction and can be of questionable value unless the results are carefully analyzed. If these general construction principles and the following analysis methods are carefully applied, the value of any one examination will no longer be questioned. A grade on a test will be a true indicator of a student's ability, or lack of ability, in any area of any subject.

There are four steps in analyzing the results: (i) tallying each question; (ii) conversion of the tally to per-

The author teaches biology at Susan E. Wagner High School, 1200 Manor Rd., Staten Island, N.Y. 10314. He received a B.S. in biology from Fordham University in 1965 and an M.S. in biochemistry from Long Island University in 1971. Cusimano carried out research on the biochemistry of serotonin at Queens Childrens Hospital and Brooklyn's Veterans Administration Hospital from 1972 to 1974. He sponsors the high school research club, and his hobby is building harpsichords.

**Table 1. Number of correct responses to each question within a group.**

| Question Number | Group A (n=20) | Group B (n=20) | Group C (n=20) |
|---|---|---|---|
| 1 | 18 | 16 | 12 |
| 2 | 17 | 11 | 6 |
| 3 | 16 | 10 | 9 |
| 4 | 6 | 3 | 5 |
| 5 | 16 | 16 | 17 |

centages for difficulty and discrimination value; (iii) an analysis of the problem questions; and (iv) diagnosis.

### *The Tally*

The examinations are graded and placed in order from the highest to the lowest score. They are then divided into three equal groups. Group A consists of the top one-third, group B the middle one-third, and group C the lowest third. If there are not enough examinations to make equal groups, the highest and, if necessary, the lowest scoring tests are eliminated.

For each question, a mark is recorded if the student answered the question correctly. This is followed for each test within each group. The results are then totalled for each question within each group as in table 1. This number represents the number of students within each group that answered a question correctly.

### *Conversion to Percentages*

Two important parameters can now be examined: degree of difficulty of a question; and the ability of a question to discriminate between the high-scoring and low-scoring groups. The sample size for each group in table 1 was 20. To calculate degree of difficulty of each question, the number of correct responses to each question is totalled and divided by the total number of test papers. In the example, question 1 has a difficulty of 46/60, or 0.77. For question 1, 77% of the students answered the question correctly. Question 3 has a difficulty of 23%.

To determine the discrimination value of a question, the number of correct responses to a question in the low-scoring group is subtracted from the high-scoring group. If this value is divided by the total number of examinations in a group (in this example, 20), a percentage value is obtained. If the high-scoring group on question 2 recorded 17 correct responses, and 6 students in group C answered the question correctly, the discrimination value is $\frac{17-6}{20}$, or 0.55. A value of 1.00 or 100% would indicate a perfect discrimination question. This information is recorded in table 2.

**Table 2. Difficulty and discrimination values for five sample questions.**

| Question number | Total correct responses | Group A minus Group C | Difficulty | Discrimination |
|---|---|---|---|---|
| 1 | 46 | +6 | 0.77 | 0.30 |
| 2 | 34 | +11 | 0.57 | 0.55 |
| 3 | 35 | +7 | 0.58 | 0.35 |
| 4* | 14 | +1 | 0.23 | 0.05 |
| 5* | 49 | -1 | 0.82 | -0.05 |

*Questions that should be classified as doubtful in value.

### *Analysis and Diagnosis*

Unfortunately, there are no simple rules that will describe the value of a question in educational parameters. In general, a question that is answered correctly by a large percentage of the students and has low discrimination would be classified as a question of dubious value and should be further examined. It should not be eliminated immediately since any easy question may indicate a strong point in the students' general ability or a well-taught idea or concept. On the other hand, there may be a cue in the wording of the question which makes the correct choice the only possible answer. In this case the question should be reworded. Further discussion with the class on this question will quickly reveal the reason.

A question which is difficult for all groups should also be examined. This may indicate that there was very little understanding of a concept or that the question was poorly worded, leading to confusion. Again, a class discussion will indicate whether the question should be reworded or the concept retaught.

**Table 3. Analysis of doubtful questions.**

| Question Number | Choices A | B | C* | D |
|---|---|---|---|---|
| 4 | 11 | 30 | 14 | 5 |
| 5 | 0 | 11 | 49 | 0 |

*Choice C was the correct answer to questions 4 and 5.

There is an inherent danger in this type of analysis. An examination may be so purified that all concepts are not adequately tested. Questions should not be eliminated simply because they are too easy or poorly worded. There is a reason for their ease or difficulty, and it is the teachers' job to diagnose the cause, with the class, and apply the appropriate remedy. Very often I include a question which does have a cue in the wording so that most students will answer it correctly. This serves to bolster the confidence of the weak students so they do not quickly concede to an examination.

What can be done with the doubtful questions? A further analysis can be carried out, before class discussion, to determine possible causes of poor discrimination, ease or difficulty. This involves tallying individual responses to each doubtful question. The results are shown below for questions 4 and 5:

For question 4, the majority of students found choice

**(Concluded on p. 181)**

While the use of the population sampling board was only one segment of our biometrics course, we feel it was a very succesful vehicle for introducing some of the basic concepts of statistics. In fact, the students constructed their own sampling boards and will be using them in their various colleges in their own courses.

**References**

ELLIOTT, J. M. 1973. *Some methods for the statistical analysis of benthic invertebrates.* The Ferry House, Ambleside, Westmoreland, Great Britain.

SCHLEFLER, W. C. 1969. *Statistics for the biological sciences.* Addison-Wesley Publishing Co., Reading, Mass.

*Ronald F. Barnes*
Department of Mathematics
State University of New York
Brockport, N.Y. 14420

# *Short-Loop Living . . .*

**from p. 175**

lization and diversity would become accepted design concepts.

### *Tomorrow Can Be Better*

The implications of short-loop living are far reaching. One impact could be the realization that the size and complexity of networks are not the only important criteria. The amount of flow is also important. Thus using less could become a critical value for the short-loop society.

Short circuiting networks could also contribute to the creation of shorter loops. For example, instead of eating meat which is now the end product of a long food chain involving petroleum, tractor factories, soybeans, grass, cattle food processing, butchering, and packing, the protein could be consumed in more direct forms, such as soybeans. This short-loop consumption might also serve to release food and resources for consumption elsewhere.

The most significant implication of the short-loop society could be surprising. To understand this impact it is important to look at what has happened to other visions of decentralized communities. Some of the turn of the century dreams of satellite communities, such as Sir Ebenezer Howard's Garden Cities outside of English cities like London, offered us a decentralized model. Several were actually built in England and although they were only able to provide residences for a small number of people, they did make a major contribution. In addition to providing a strong influence on urban planning, those early dreams offered a positive vision of a desirable tomorrow with people living in harmony with nature.

Given the current rampant pessimism about the prospects for the future, it may not be important that an urban greenhouse or a basement fish tank can't overnight provide us with a solution to either the world's food problems or urban instability. The fact that they exist could provide a promise that tomorrow can be better. That idea alone could be an important contribution.

# *The Test . . .*

**from p. 177**

B more appealing than the correct choice. The wording of choice B may have misled the students or, more importantly, a concept or idea was incorrectly learned or taught. At times I have found that students actually learned the opposite of what I thought they were learning! This type of analysis will prevent such catastrophes from being promulgated. In cases of questions which are too easy, further analysis, as for question 5, will not yield much information.

### *The Test: an Integrating Factor*

For an examination to be a good teaching, learning, and diagnostic tool, construction must highlight levels of the cognitive domain: knowledge, comprehension, application and analysis, synthesis, and evaluation. All levels can be adequately tested by means of an objective examination. It is time that we move away from a purely knowledge examination to one which allows a student to show his true ability, not his memorizing capacity. An examination is not a marking tool, separate and distinct from a student's learning experience. With proper application of these construction and analysis techniques, we can truly test ability and understanding. A grade on an examination will become a meaningful measure of both teaching and learning, and, in the end, a means for improving both.

# NABT Convention Calendar

The following are dates of seminars and conventions sponsored by the National Association of Biology Teachers. For additional information, write to NABT, 1420 N Street, N.W., Washington, D.C. 20005.

May 12–14, 1975. NABT-ICFAR Seminar, Indianapolis, Indiana.

October 23–26, 1975. NABT National Convention, The Portland Hilton, Portland, Oregon.

October 14–17, 1976. NABT National Convention, The Regency, Denver, Colorado.

October 20–24, 1977. NABT National Convention, The Anaheim Convention Center, Anaheim, California.