METHODS OF TESTING IN SCIENCE TEACHING

J. Verduin
University of South Dakota

The methods described in this report are the outgrowth of an attempt to improve testing in the botany courses offered at the University of South Dakota. Science courses frequently list a large array of noble objectives, such as stimulation of interest in science, increased accurateness of observation, broadening of student's contact with life, appreciation of the scientific method, and practice in thinking, but testing is likely to degenerate into sampling (and statistically inadequate at that) of a student's success in memorizing material from a textbook, or worse, from lecture notes. It seems self-evident that any objectives which do not find place in a testing program are purely decoration; an instructor can gain no estimate of his success in attaining an objective unless it is incorporated in a testing program, and the average student will not take seriously any objective toward which his achievement is not sampled. The subject of testing is too large for adequate treatment in the space allotted here, and only a few methods and principles will be discussed.

Frequency of Testing

Tests in General Botany have been given frequently—six to eight tests per semester, all carrying equal weight toward the final grade, and being cumulative in scope, that is, each test includes questions from all material covered to date. Inadequate frequency of testing is, of course, indefensible. It is statistically bad because small samples are likely to be unrepresentative, but it is psychologically bad because it encourages cramming for tests and loafing during the rest of the semester. Such study habits are deplorable, but one does not have to live in a college atmosphere very long to realize that they are generally prevalent. These habits are a product of the testing system and will not change until the system is modified. Therefore, the method outlined above was adopted: frequent testing to insure adequate sampling, equal weight to all tests and cumulative scope to encourage careful preparation of each assignment.

Questions or Problems Involving Thought

Questions and problems in the examinations were designed to involve thinking rather than memory. Students come to college with the unfortunate idea that education is a matter of memorizing what the professor is likely to ask on examinations. Questions which test memory are likely to "evade the issue" because facts can be and often are memorized without being understood or related to other known facts. For example, when studying the mosses in botany most of the students who had read the text knew that mosses lacked vacular tissue but on being questioned many admitted that they didn't know what vascular tissue was. Although they had during the previous semester studied vascular tissues for several periods they failed to associate the moss assignment with the previous semester's study. Questions which require thinking are difficult and the average or below average students resent them when they are first encountered. Student reaction is likely to be, "Why didn't someone tell me this?" or "How were we supposed to know that?" and when told that they are supposed to figure it out they consider the question "unfair". But thought questions can be doubly useful in that they can teach as well as test. If a test is made hard enough so that the best students miss some questions then all will learn something from it and the test period has not been wasted in gathering data for grading but the teaching process continues during the test session.

Practical Tests Using Laboratory Materials

More than half of the tests given in botany courses have been "practicals". Just as memory without thinking is meaningless, so learning facts without handling the materials involved is unprofitable. For example, a student may know that the phloem of a tree is in the bark, that the xylem is in the wood, and that the cambium lies between them, but if he can't identify these regions in a twig cross section he doesn't know anything about twig anatomy. In many science courses the laboratory work contributes only a small fraction of the student's grade in the course, but students spend twice as much time in laboratory as in lecture so tests over laboratory material can reasonably contribute the major part of the grading data.

Laboratory testing offers a rich range of possibilities which are commonly neglected, for example, in organic chemistry students read about the color, volatility and odor of many compounds, but in four organic chemistry courses studied by the author, students were never graded on their ability to identify compounds by these properties but were required to list physical properties in written examinations. The students who passed such examinations were not the ones who could identify compounds but those who had memorized the lists of physical properties.

Botany is without a doubt the best subject in the world for laboratory testing. Even in the dead of winter there is material out-of-doors for several tests, and a dozen potted plants in a sunny window can provide more material than one can use in an entire semester. If some materials which have not been studied in laboratory are introduced in such a test, a student's ability to apply his knowledge to a new problem can be measured. For example, if the anatomy of the corn stem has been studied and a test contains a slide of wheat stem, the section will look different enough so the students will realize that it is foreign and a discouraging number of them will guess haphazardly at the answer, but some will apply what they have learned about corn to the wheat section and will come up with the right answer. Such questions are not unfair since no one has seen the section before, and it is impossible to test a student's acuteness of observation or thinking ability unless he is faced with a situation that is new to him.

Open Book Objective Tests

In tests covering textbook material, open books were always used, and tests were objective (the laboratory practicals were also objective tests). The gravest objection to closed book testing is naturally the unfortunate emphasis on memory, but there is at least one other important objection, that is, the limited scope dictated by the closed book. Many instructors present a digest of the textbook in lecture, and then test over the "more important parts" of that digest, thus narrowing their course hopelessly; when open book tests are used an instructor can insist that students understand everything in the text, and additional material can be presented in lecture, then the student's understanding of this much wider range of material can be tested in open book style.

In writing open book tests, true-false questions were rejected as being too easy and too wasteful; guessing will yield fifty per cent correct so half of the questions written are wasted. Instead a three-way choice was always present. In this way the "yield" is increased to sixty-seven per cent and the problems are sufficiently complicated to make careful consideration of each one essential.

Some Sample Problems

The instructions for one type of open book problem are reproduced below:

"Each of the sentences below is made up of a statement followed by a reason introduced by the word **because.** If a statement is false, cross out an (f) on your answer sheet. If a statement is true and is well supported by the reason cross out a (t). If a statement is true but the reason given is inadequate or irrelevant cross out a (c)."

1. The bacteria are considered to be relatives of the blue-green algae because they lack nuclei and multiply by fission.
10. The slime molds are classed as plants rather than animals because their energy requirements (respiration) are similar to those of plants.

12. A plant-like **Fucus** in figure 288, page 381 (Holman and Robbins) is properly called heteroecious because male and female gametangia are produced on different individuals.

Statement 1 is a (t) statement since the lack of nuclei and multiplication by fission are possessed in common by groups and suggest relationship. Statement 10 is a (c) statement since respiration in plants and animals is essentially similar and cannot be used as a characteristic distinguishing between groups. Statement 12 is obvioulsy false.

Problems of this sort measure a student's ability to find information in the text plus his ability to apply informatios to questions which are not directly answered in the text. Factors which contribute to a student's success are: care in reading statements, familiarity with the text, and good judgment in evaluating the "supporting clauses".

The instructions for a second type of problem are: "The following statements refer to the carbon cycle diagram shown on page 402 (Holman and Robbins). If a statement is false cross out an (f). If a statement is probably true but is not supported by the diagrams, cross out a (t). If a statement is true and is confirmed by the diagram, cross out a (c).

1. If photosynthesis were completely arrested the carbon cycle would grind to a halt after the carbohydrates lying around had been changed to carbon dioxide.
2. If all animals were destroyed the carbon cycle would soon be at a standstill.
11. Adding carbon dioxide to the air in a greenhouse would probably improve plant growth.

Statement 1 is a (c) statement—true and confirmed by the diagram, 2 is false (f), and 11 is a (t) statement—true but not confirmed by the diagram. This type of problem tests a student's ability to interpret diagrams. Some students find this sort of problem especially difficult, they are apparently unable to picture things mentally. Such students certainly find textbook diagrams practically useless as teaching aids.

A third type of problem is sampled below:

"A man wanted to know whether his poor radish seed germination was caused by a pathogenic organism in the soil or by improper storage temperature, so he treated five samples of seed as indicated in the table below and got the results shown in the right hand column.

Sample No.	Storage Temperature	Planted in	Germ-ination
1.	Normal (10° C.)	Ordinary soil	47%
2.	Normal (10° C.)	Steam sterilized soil	88%
3.	0° C	Ordinary soil	50%
4.	15° C.	Ordinary soil	49%
5.	25° C.	Ordinary soil	15%

All other conditions uniform for all samples.

Read the statements below: If a statement is supported by the data in the table cross out an (s) on your sheet. If a statement is contradicted by the data cross out a (c) on on your answer sheet. If a statement goes beyond the data and is neither supported nor contradicted cross out an (n).

4. Plants grow better when the soil is warm, about 25 degrees.

5. The data indicate that the cause of poor germination was largely removed by steam-sterilizing the soil.

14. The experiment did not shed any light on the man's problem.

Number 4 is an (n) statement—probably true, but there is no evidence for or against it in the table. Number 5 is an (s) statement and 14 is a (c) example. A problem of this type tests a student's ability to evaluate data. In a course which presumes to provide acquaintance with the scientific method such problems are well worth while. Sometimes there are tables in the text which can be used in such questions but it is usually advisable to construct a hypothetical experiment for the sake of simplicity.

Problems of this type usually contained fifteen units (questions or statements) and four problems constituted an entire examination. Such problems lend themselves well to open book testing, and numerous other types will suggest themselves—the familiar multiple choice completion type among them.

In open book objective examinations of the kind described above, the range of errors made by individual students has fallen between 50-60% for the poorest students and 4-8% for the best students. Each of the tests given during the semester yielded a fairly normal curve, and the composite grades at the semester's end yielded a normal curve without revision or redistribution, indicating that more students "stayed in their class"—that is, the "D" students seldom wrote a "C" test, or if they did, it was balanced by an "F" on some other test, and similiarly with the other grade ranks.

A list of the representative grades is shown below:

									Semester Grade
1.	79	80	*71	83	66	70	75	*76	C
2.	61	60	78	63	64	68	66	64	D
3.	99	94	100	98	98	100	98	100	A
4.	89	86	88	85	91	98	90	95	A (low)
5.	84	60	60	70	72	95	74	66	C
6.	76	62	80	77	74	74	74	82	C
7.	89	90	83	83	79	70	82	83	B
8.	61	66	66	70	71	51	66	72	D
9.	79	82	78	88	88	75	82	79	B
10.	66	56	51	61	55	59	58	66	F

*These columns represent open book objective examinations, the others are laboratory "practicals".

The close agreement of grades on different tests by the same individual is conspicuous, but there are enough examples of rather wide deviation on a single test to emphasize the danger of inadequate sampling. The students numbered 5 and 6 exhibit an interesting reaction to the open book objective tests: Student number 5 obtained grades in the sixties on both of these tests while his laboratory practicals showed grades usually in the seventies or better, student number 6 obtained grades of eighty or better on the open book tests while the laboratory practicals were consistently in the seventies. Students who show such variation in response to the two types of test are not uncommon, but most students show the same general rank in either type of test. The close agreement of open book test scores with other test scores (in which memory is an important factor) indicates that the advantages provided by open book test are obtained without sacrificing any accuracy or efficiency in grading.

SUMMARY

The results obtained in two years of experimentation with laboratory practical tests and open book objective tests show that a variety of student skills can be measured, that such tests can serve as teaching devices as well as grading devices, and that undesirable over-emphasis on memory is avoided.