SOME EVALUATION TECHNIQUES IN TENTH-GRADE BIOLOGY

H. W. CRALL
Western Illinois State College, Macomb

Evaluation techniques for a 10th grade biology class at Western Illinois State College Laboratory School, 1949-1950, included:

1. Questionnaires concerning interests and backgrounds.
2. Interest and aptitude tests.
3. "Opinionnaires" and tests to measure progress.
4. Interview forms.
5. Forms for parents.
6. Forms for other teachers.
7. Behavior checking forms.
8. Guide sheets.
9. Anecdotal records.

The questionnaire was modelled after one found in *Methods and Materials for Teaching Biological Sciences,* by Miller and Blaydes, 1938.

The aptitude test used was the California Mental Maturity Test.

The interest test was "The Occupational Inventory" of Lee and Thorpe, 1946.

The "opinionnaires" were modelled after those constructed by Everote in his Ph.D. dissertation, "Agricultural Science to Serve Youth," Columbia University, 1943. They were constructed with the cooperation of seven methods students and five practice teachers of biology at Western Illinois State College. Some degree of reliability was given these instruments by having them scored by advanced biology students. From an analysis of these results the "correct" answers and the "best" reasons for them were set up. The opinionnaires consisted of questions which were to be answered "yes," "no," and "undecided." A reason for each choice was asked for. One opinionnaire of 30 questions was constructed to measure tolerance and "open-mindedness." An example of a question from this test is:

Do you agree that Indians and Orientals should have equal citizenship rights with others in this country?

The other opinionnaire of 28 questions was devised to measure sensitivity to the social implications of science. An example of a question from this test is:

Because of their potentialities for world wide destruction, do you think the production of atomic bombs should be internationally controlled?

Tests to measure the ability of students to make choices and decisions were constructed by these same advanced biology students. Some degree of reliability was given this test by having these students score it. An example of a problem from it is:

Tell, in a few sentences, how you would study the life history of the housefly in a biological laboratory, and indicate what equipment and materials you would use.

In making a test to measure laboratory skills and techniques the same

group of students cooperated, and a degree of reliability was given to the test in a manner similar to that used with the other tests. An example of a problem from it is:

Study a prepared slide under high power and under low power, and make sketches of what you see.

Tests to measure abilities to interpret data logically, and to distinguish facts from assumptions in making conclusions, were constructed by this same group of students based on tests described in "Science in General Education," by the Commission on Secondary School Curriculum, Committee on the Function of Science in General Education, Progressive Education Association, 1938, and in the bulletins used in "Evaluation in the Eight Year Study," P.E.A., 1936-1938. Some of the test questions were taken intact from these sources. Others were modelled after questions found therein. A degree of reliability was given these tests in the same manner it was given the others.

A test of ability to apply biological principles was constructed by the same group of students by following the models set up by Raths, Frutchey, Tyler, and Zechiel in their bulletins of "Evaluation in the Eight Year Study," P.E.A., 1936-1938. Reliability of a degree was again established by similar methods.

A test over knowledge of biological factual information was constructed by these advanced students and a degree of reliability was established by them.

This same group of students cooperated in constructing interview forms, form letters for parents, forms for other teachers, guide sheets, and so-called "instructional" tests designed to measure progress of students toward the various objectives at one time. Reliability of a degree was given interview forms by having two interviewers record an interview simultaneously on separate forms.

A behavior check form was constructed with their cooperation modeled after, but not identical to, forms devised by West, Wrightstone, Goodenough, and Olson in various publications, including a doctoral dissertation, 1934 to 1943. Degrees of validity and reliability were given this instrument by determining sample consistencies of agreement between the behavior checkers, in the consistencies of five-day samples of frequency of observation, and the determination during the first week of school of the types of behavior to be observed as being "positive" with respect to the different abilities to be learned.

Anecdotal record taking was modeled after Frutchey and Tyler, "Examinations in Natural Sciences," *The Construction and Use of Achievement Examinations*, 1936. To insure some degree of reliability, two observers made an anecdotal record of a student simultaneously whenever possible.

These instruments were used as follows:

1. Pre and final questionnaires, interest and aptitude tests were given.
2. Special pre and final opinionnaires and tests were given to measure progress for the year. These included the two types of opinionnaires, nature-of-proof

tests, interpretation-of-data tests, application-of-principles tests, and factual information tests.

3. Each student was interviewed about twice every six weeks.
4. Other opinionnaires and tests were given bi-weekly. These were the so-called "instructional" tests.
5. Every six weeks parents were informed of student progress, and the parents were asked to evaluate progress each quarter.
6. Other teachers having these same students evaluated student progress every six weeks.
7. Each student's classroom behavior was checked about twice a week, using the behavior check form.
8. Guide sheets accompanied class experiences, and these sheets were designed to facilitate growth toward the objectives.
9. Anecdotal records were made about twice a quarter for each student.

Results at the end of the year:

1. A direct relationship existed between reading ability, I.Q., growth in subject-matter interest, and in subject-matter learning.
2. Behavior checking was the most reliable index of progress, and anecdotal records were least reliable.
3. Results of different evaluation techniques were related, were a composite reliable index of student progress, and were utilized in grading students for achievement.