R E P O R T   R E S U M E S

ED 011 841 SE 000 756

EVALUATION IN ELEMENTARY SCHOOL SCIENCE.

BY- TANNENBAUM, HAROLD E. AND OTHERS

OFFICE OF EDUCATION, WASHINGTON, D.C.

REPORT NUMBER CIRC-757 PUB DATE 64

EDRS PRICE MF-$0.18 HC-$3.32 83P.

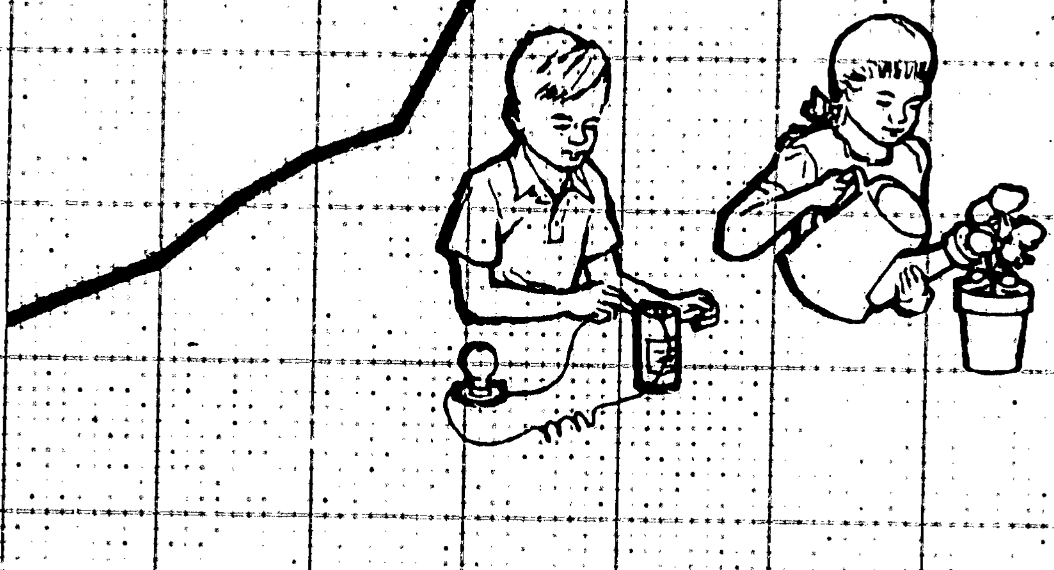
DESCRIPTORS- *CURRICULUM EVALUATION, *EVALUATION, *ELEMENTARY SCHOOL SCIENCE, *EDUCATIONAL OBJECTIVES, INSTRUCTION, INSERVICE TEACHER EDUCATION, SCIENCE TESTS, SUPERVISION, SELF EVALUATION, STUDENT EVALUATION, TEST CONSTRUCTION, TEACHER EVALUATION, U.S. OFFICE OF EDUCATION, EFFECTIVE TEACHING, TEXTBOOK EVALUATION, DISTRICT OF COLUMBIA

GUIDELINES FOR THE EVALUATION OF ELEMENTARY SCHOOL SCIENCE PROGRAMS AND ELEMENTARY SCHOOL SCIENCE STUDENTS ARE PRESENTED. CRITERIA ARE INCLUDED FOR (1) ESTABLISHED CURRICULUMS, (2) LIBRARY FACILITIES, (3) EQUIPMENT AND SUPPLIES, (4) FIELD EXPERIENCES AND AUDIOVISUAL AIDS, (5) TEACHER-MADE TESTS, AND (6) CLASSROOM TEACHING TECHNIQUES. SUBDIVISIONS OF THESE MAJOR CATEGORIES INCLUDE SUGGESTIONS FOR PROGRAM ANALYSIS. SUGGESTIONS FOR THE DEVELOPMENT OF INSTRUMENTS FOR CURRICULUM EVALUATION ARE ALSO INCLUDED. SUGGESTIONS FOR STUDENT EVALUATION ARE BASED ON BEHAVIORAL GOALS. CONSIDERATION IS GIVEN TO (1) SELECTION OF DEVICES FOR EVALUATION, (2) VALIDITY AND RELIABILITY OF EVALUATION TECHNIQUES, AND (3) TYPES OF EVALUATION INSTRUMENTS. MANY SAMPLE TEST ITEMS FOR DIFFERENT CATEGORIES OF TESTS ARE PRESENTED. CHECK LISTS FOR (1) EVALUATION OF TEACHERS, (2) LESSON PLANS, (3) INSERVICE COURSES, (4) SCIENCE BOOKS, (5) FACILITIES, AND (6) TEACHER UTILIZATION OF TELEVISION SCIENCE BROADCASTS ARE INCLUDED. THIS DOCUMENT IS AVAILABLE AS FS-5-229--29057 FROM THE SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE, WASHINGTON, D.C. 20402, FOR $0.45. (AG)

756

OE-29057

EVALUATION In Elementary School Science



U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

OE-29057
Circular No. 757

EVALUATION In Elementary School Science

by HAROLD E. TANNENBAUM
Professor of Science Education
Hunter College,
The City University of New York

and NATHAN STILLMAN
Professor of Education
Yeshiva University

in collaboration with ALBERT PILTZ
Specialist, Science
Instructional Resources Branch
U.S. Office of Education

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Office of Education

Superintendent of Documents Catalog No. FS 5.229:29057

U.S. GOVERNMENT PRINTING OFFICE
WASHINGTON : 1964

**For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C., 20402 - Price 45 cents**

CONTENTS

FOREWORD

Thousands of man-hours and millions of dollars are being spent for the improvement of science instruction in American schools. Since 1958, title III of the National Defense Education Act has provided financial assistance to States, and through them to individual school systems, for strengthening instruction in science, mathematics, and modern foreign languages. It not only provides matching funds for the purchase of materials and equipment, but also allocates funds for the expansion and improvement of State supervisory services.

This bulletin is intended to serve a number of different purposes in the area of elementary school science. First and foremost, it should be helpful to supervisors and administrators on the State and local levels in evaluating the effectiveness of the total elementary science program. Second, it should enable teachers to do a more effective job of evaluating the growth of their students toward the many-faceted goals of the science program. Third, it should provide supervisors with materials for inservice teacher education programs in evaluation. Fourth, since effective teaching is reflected in student achievement, the bulletin should aid supervisors by giving them some guiding principles in the area of evaluation which they can use for measuring the effectiveness of the teachers.

The bulletin was prepared by Professors Harold E. Tannenbaum and Nathan Stillman, pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. It represents the joint efforts of the authors and of Dr. Albert Piltz of the Office of Education staff.

INTRODUCTION

As presently viewed, elementary school science is part of a total science curriculum which extends from kindergarten through grade 12 or 14. Science educators, supervisors, and classroom teachers generally are agreed upon the following long-term goals for this total curriculum:

> The primary goals of science education should be intellectual. What is required is student involvement in an exploration of important ideas of science. ...Science is one of man's major intellectual accomplishments, a product of the mind which can be enjoyed - not for its material fruits alone but for the sense which it provides of order in our universe.[1/]

In the elementary school, the curriculum should introduce pupils to those experiences which lead to the realization of these long-term goals. Thus, elementary school science should foster in each pupil:

1. Increasing ability to comprehend and deal with his natural environment;

2. Understanding of the methods, techniques, and attitudes employed in the scientific enterprise so that he may begin to use this rational approach to the solution of his current and future problems;

3. Recognition of the interrelationship of science with all other human experiences.

The science program should be designed to promote these curricular goals. The facts that pupils are taught, the concepts they are to acquire, the skills they are to develop, the methods through which instruction is carried out and the evaluation of progress are all part of a total program built to lead pupils toward rational thinking and an intellectual appreciation of the potentials and limitations of science.

Although the words "goals" and "objectives" are often used interchangeably, in this bulletin the following distinction is made:

> Goals are the long-term and broad-scoped aims of the science curriculum. Thus, a goal influences the total science curriculum.

1/ "The NSTA Position on Curriculum Development in Science," <u>The Science Teacher</u>, Vol. 29, No. 8, page 35.

> Objectives are the short-term and limited aims of a single unit or portion of the curriculum. Objectives are derived from goals and the fulfillment of related objectives leads towards achievement of a goal.

This bulletin has been prepared for a diverse professional audience. The first section should serve as a guide for supervisory personnel, both State and local, when they examine and appraise elementary school science programs. The second section should assist classroom teachers in designing satisfactory measuring instruments and in using these instruments effectively. This section also will be useful to local supervisors in appraising individual lessons and in assisting or training classroom teachers to use evaluative techniques.

SECTION I. EVALUATION BY STATE AND LOCAL SUPERVISORS

Significant Aspects of Effective Science Programs

Local control of American schools has resulted in diverse interpretations and implementations of the generally accepted objectives of universal education. Educational procedures vary from school system to school system, but in spite of such diversity, any program has significant aspects which are common to all. These may be used as guides by supervisors and other educational personnel in the evaluation of specific science programs.

Personnel responsible for appraising the total science program of a school, a school system, or even an entire State should be very selective in the use of evaluative devices. The tremendous amount of paperwork involved in the jobs of such people makes it important that only key and distinctly significant data be gathered. From six major aspects of information supervisors can gain significant, though not conclusive, impressions of the program: 1) Established curriculums; 2) library facilities; 3) equipment and supplies; 4) field experiences and audiovisual aids; 5) teacher-made tests; and 6) classroom teaching techniques. These aspects will be considered seriatim, and questions useful to supervisors will be suggested for each.

1. Established curriculums

Officials of school systems differ in the ways they plan, organize, and record their total curriculums. Some feel that only the barest outlines of a curriculum should be fostered by the central administration; others prepare suggested guides; still others mandate specific and detailed programs for each subject for each class in every school. In many school systems in the United States, elementary school science is not separately recorded in the curriculum. Some kind of formal curriculum is the first essential for any sound program in elementary school science.

Whether or not a school system has a written science curriculum is an important measure of its concern for teaching in this area. When there is no written curriculum, the system probably has a haphazard and catch-as-catch-can science program. The following questions may suggest guidelines either for examining newly prepared curriculums or for evaluating those in use.

a. Does the curriculum provide a scope and sequence in keeping with the general educational objectives of the State and/or local community?

748-546 O - 64 - 2

Current thinking in curriculum construction makes a strong case for an elementary school science curriculum which is, on the one hand, a part of a kindergarten-through-12th grade total science program, and, on the other hand, a unit integrated with the total educational activity of the elementary school. Devising a plan which satisfies these two demanding requirements is no mean task. Furthermore, each State and each community has its own special needs and consequent objectives. The supervisor can examine a curriculum to determine the:

(1) Number of units included in the total elementary science program;

(2) Range of science topics included in the program;

(3) Organization for repetition of these areas (e.g., loose spiral, tight spiral);

(4) Articulation of science curriculums among elementary school, junior high school, and senior high school; and

(5) Plans for variations in experiences to provide for the needs of children of differing abilities.

b. Who prepared the elementary school science curriculum, and how was it prepared?

As increased numbers of teachers truly share in preparation of the curriculum, it is more likely to be used effectively; a curriculum for any subject is only as good as its implementation. On the other hand, just a list of names - supposed sponsors and authors of the curriculum - is of little value. The supervisor can examine the composition of the curriculum committee to find out whether:

(1) The committee has a representative range of the elementary school teachers in the system;

(2) Junior and senior high school science teachers are participating members of the committee;

(3) The committee is a standing committee;

(4) Replacements are appointed to the committee by some regular procedure.

An elementary science committee profits from the services of special personnel in addition to regular teachers, and from appropriate resource materials. The supervisor can determine whether:

(1) An elementary science specialist is attached to the committee in some capacity;

(2) An elementary school curriculum coordinator is available to the committee;

(3) The committee has access to a curriculum library.

c. Have arrangements been made for review, evaluation, and revision of the curriculum?

Recent changes in educational practice have outmoded many curriculums which were prepared five or more years ago. Alteration of content, stress upon individualized experiences, emphasis upon science processes such as formulating and testing hypotheses and upon accurate observation and quantitative measurement, and the demand that fewer topics be studied, but studied in much greater depth--all have necessitated re-examination of existing curriculums. The supervisor can discover whether:

(1) A regular review, with feedback from teachers to the curriculum committee, is an organized procedure;

(2) The curriculum takes into account the teaching of newer topics;

(3) The curriculum includes the teaching of the processes of science.

d. Are there provisions for inservice education of the staff members who use the curriculum?

Since the science background of many elementary school teachers is limited, a comprehensive inservice education program designed in relation to a new or revised elementary science curriculum is an indication to the supervisor that a school is working to upgrade its program in science. The supervisor can investigate:

(1) What inservice education programs in elementary school science exist; and

(2) How the inservice programs are related to the curriculum.

e. Have equipment, supplies, published materials, and resources kept up with the changes in curriculum?

Simply changing the curriculum without providing the necessary facilities for implementing these changes is not very useful. In evaluating a curriculum, the supervisor can examine the lists of new materials and supplies that have been acquired as a result of changes instituted by the curriculum revision to determine:

(1) What percentage of new supplies and equipment has been purchased since the curriculum was revised;

(2) Whether these new materials and supplies are needed for specific experiences included in the new program.

2. Library facilities

Of great importance in determining the nature of an elementary school science program is an inventory of the science books in the school library, as well as information about their circulation. Much more than the number of science titles alone should be determined; it is wise, for example, to know the distribution by subject and level of difficulty of the titles available. The supervisor can use the following questions and suggestions to evaluate phases of the science program through information about the library, its resources, and how they are used.

a. Is there a balanced distribution of titles and an adequate reference collection?

There should be materials to satisfy the range of interests and abilities of the children in the school from the least mature to the most capable. The supervisor can determine whether there are:

(1) Books in the physical sciences as well as the biological sciences;

(2) A range of reading materials with primary, intermediate, and upper-grade books;

(3) Materials at the high school and even at the adult level as well as materials with high interest and low reading level.

The library materials should fit into the curriculum design of the school. The supervisor can find out if:

(1) There are materials to satisfy the curriculum plans for each of the grades;

(2) There are materials on given topics at various reading levels so that children of different reading abilities can find information about the same topic;

(3) New library materials have been acquired as the curriculum has been changed.

An important objective of many elementary school science programs is teaching the use of reference materials. The supervisor can look for:

(1) A broad range of children's reference materials in the library (e.g., children's encyclopedias, children's dictionaries, atlases, vertical files of pamphlets and pictures, files of elementary science periodicals, standard field guides); and

(2) An adult encyclopedia and an unabridged dictionary for use by teachers and more mature children.

b. How are science materials evaluated and acquired?

Some indications of the use of library resources can be obtained by knowing the ways in which library materials are acquired, i.e.:

(1) Whether one person has responsibility for coordinating all purchases and acquisitions;

(2) How information concerning the needs of individual classes and teachers is directed to the coordinator;

(3) How teachers are informed about new publications.

By examining the process used to evaluate science materials in the library, the supervisor can find out:

(1) Whether there are forms prepared for reviewing books and other materials;

(2) What procedure, if any, is used for evaluating new materials once they have been purchased;

(3) Whether children are asked to evaluate the materials they use.

Another significant aspect is the recency of the science materials and their relationship to the total library collection. Science is a rapidly changing field and a library collection requires constant revision to stay up-to-date. The supervisor can ascertain:

(1) What percentage of the total collection is science material;

(2) What percentage of the science collection is less than 5 years old;

(3) What the replacement budget for the science collection is and what percentage of the total budget this represents;

(4) What funds are provided specifically for new science acquisitions.

c. What kinds of science materials are used most frequently by the children?

Specific information about circulation is very revealing in judging which parts of the science program are most effective. The supervisor can inquire whether:

(1) More books about the physical sciences are being circulated than books about the biological sciences;

(2) More primary than intermediate grade science books are being circulated;

(3) The picture files are used frequently.

3. Equipment and supplies

Information about equipment and supplies also can provide pertinent data for the supervisor. This is an area which needs particular attention. In the past, very little money was allocated for purchasing elementary school science supplies. With the passage of the National Defense Education Act, funds became much more readily available for such purposes. Depending upon the nature of the local elementary science program, the funds can be expended in many ways, and the supervisor has the responsibility of evaluating them. The following suggestions and questions are offered to aid in the assessment of elementary science equipment inventories and the formulation of priorities for new purchases.

a. Are there equipment and supplies which lend themselves to curricular experiences and experiments planned for the program?

The supervisor can determine the appropriateness and adequacy of equipment and supplies by finding out if:

(1) There are necessary individual supplies available in sufficient quantities to allow for the planned individual experiences of the children;

(2) There are appropriate demonstration facilities available so that the planned demonstrations can be carried out;

(3) The necessary supplies and equipment are specified in the curriculum so that teachers and other concerned personnel know what materials to use and where they are available.

b. Are repair and storage facilities provided for the science materials?

Supplies, materials, and equipment (even those inventoried as permanent) wear out and are used up. The supervisor can see whether:

(1) There is a regular budget for repair of science equipment;

(2) There is a regular budget for replacement of expendable items;

(3) The maintenance of science equipment is part of the regular assignment of a particular member of the staff.

Storage facilities for science equipment is another possible determinant of how well a science program operates in a school. The supervisor can check if:

(1) There is a special location for materials and supplies which need to be shared by various classes;

(2) There are coordinated arrangements for borrowing from the central pool;

(3) One person is in charge of the central supplies.

c. Is there an organized procedure for requisitioning new materials and for evaluating them?

Just as in the case of acquiring library materials, there are many possible ways to acquire new science equipment and supplies. The supervisor can examine how such procedures have been organized to determine:

(1) Who, if anyone, is responsible for the coordinated purchase of equipment and supplies;

(2) How priorities for purchase are determined;

(3) If purchases are justified in terms of the curriculum of a given grade or class.

4. Field experiences and audiovisual aids

Field experiences and audiovisual aids are important teaching tools and learning aids for all education. They are particularly valuable for teaching science since many natural phenomena are understood better when they are seen. Furthermore, many phenomena do not lend themselves to personal experience, and films, tapes, and similar devices are often valuable substitutes, if properly used. The following questions and suggestions can assist the supervisor in evaluating this aspect of the elementary school science program.

a. Is an organized and planned program of field and audiovisual experiences a regular part of the curriculum?

To evaluate this phase of the program, the supervisor can decide whether:

(1) Clear reasons are given in the curriculum for the use of trips and audiovisual aids;

(2) Budget is provided for renting or purchasing needed audiovisual materials;

(3) School procedures are established for taking field trips;

(4) Needed transportation facilities are available and funds for such transportation are provided.

Often many people in the community can serve as resource persons in elementary school science. The supervisor can find out if:

(1) There are plans for use of such persons;

(2) A file of community resource people is available to the teachers;

(3) The services of public and governmental agencies (such as the Board of Health or the Department of Water Supply) are used in the elementary science curriculum.

b. Are needed facilities and equipment for audiovisual and field experiences provided?

Without adequate budgetary support and without care, audiovisual equipment deteriorates. The supervisor can check if:

(1) The necessary items of equipment for the plans included in the elementary science curriculum are available (e.g., sound-movie projector, slide projector, opaque projector);

(2) The equipment is controlled for shared use;

(3) Repair facilities are available and budgeted;

(4) Acquisitions are made through a centralized plan;

(5) Teachers are kept informed about newly prepared or acquired materials.

5. Teacher-made tests

An examination of a representative group of teacher-made tests can provide basic information about the science program. The following questions may be of help to the supervisor.

a. What aspects of the science curriculum are stressed in the tests?

Regardless of what appears in the curriculum, a teacher's tests reveal the objectives which he considers of importance in his teaching. The supervisor can determine which of the following are stressed:

(1) Science generalizations;

(2) Application of generalizations to specific problems;

(3) Information about specific items of subject matter;

(4) Interpretation of science data.

b. What kinds of test items are included?

The choice of test items reflects the teacher's approach to the curriculum. The supervisor can examine the tests to see how the following items are used:

(1) Essay tests;

(2) Objective tests;

(3) Performance tests.

6. Teaching techniques

For obtaining accurate and specific information about a program, observation of teachers as they work with their students is most revealing. In most cases, supervisors of large districts or of States do not have the time to visit many classrooms in action, but random sampling of classroom operations, when possible, will convey some feeling for the character of the teaching being carried out in the system. Questions that are appropriate for this phase of evaluation include the following:

a. Is the teacher presenting his lesson to stress the over-all objectives of the science curriculum?

It has been argued that elementary school science should help teach children to solve problems through rational thinking. The supervisor can observe whether:

748-546 O - 64 - 3

(1) The class is conducted so as to foster problem-solving techniques;

(2) The teacher encourages careful descriptive statements of phenomena;

(3) The teacher encourages students to formulate hypotheses and to plan for testing such hypotheses;

(4) The teacher asks questions which encourage exploration of ideas.

If a major objective of the program is the teaching of specific information about science phenomena, the supervisor might want to observe whether:

(1) The statements of information are accurately presented;

(2) The teacher explains the information so that all the students have a chance to understand what has been presented.

b. Is the classroom appropriately arranged for teaching elementary school science?

Again, the objectives of the science program determine the kinds of questions the supervisor tries to answer. Where individualized experiences are stressed, the supervisor might look for the following conditions:

(1) The room is set up so that individual children can participate in the desired experiences;

(2) The room is set up so that the teacher can supervise such things as safety and care of materials and supplies;

(3) Facilities are available in the room for special individual projects.

If demonstrations are important in a given curriculum, the supervisor can determine whether the demonstration can be seen and followed by all the children.

Good science teaching in the elementary school pervades the entire class environment. The supervisor can observe whether:

(1) Science bulletin boards and science displays related to the materials currently being studied are on display in the room;

(2) There are evidences of science projects in the room;

(3) There are well-cared-for living materials, such as plants and aquariums, in the classroom.

Devices for Evaluating Science Programs

The supervisor, as he prepares to evaluate the science program of a community, must establish a model against which to measure what he finds. A State supervisor can prepare a series of generalized models appropriate to the various types of communities in his State. Then, when he is considering a specific community, he can use the appropriate general model which fits the local situation. Employing the criteria which he considers appropriate, he can determine the soundness of the curriculum design, the value of the books that the system owns, the suitability of the equipment available in the schools, the adequateness of the field experiences and audio-visual aids, the relevance of the teacher-made tests, and the effectiveness of the classroom teaching techniques.

With his model as a measure, the supervisor can do one of two things. He can visit the school himself, and, using a pertinent checklist based on his established criteria, he can rate the school after having observed it. Or, if the supervisor cannot make the necessary personal visit to evaluate through observation, he can have local personnel complete a questionnaire appropriate for gathering the data needed.

1. Suggestions for preparing checklists

Checklists can be used to record observations of all phases of the school which have bearing on the science program - administrative practices, supervisory procedures, curriculum, equipment and supplies, teacher competency, and the like.

a. A separate checklist should be prepared for each phase of the school program that is to be evaluated. A series of short, discrete lists is more easily used than a single, long, comprehensive list.

b. Checklists should be prepared in the light of the criteria which have been established. Each item or statement on the list should be included only if it provides relevant data for evaluating one of the criteria.

c. Checklists should include all the information needed, but no more; brevity is a desirable attribute of a checklist.

A sample of a list of items with regard to administrative practices might look like this:

(1) Science is scheduled by administrative mandate for:

each grade ______________

only grades 4-6 ______________

no grade ______________

(2) In grades K-3, the weekly time allotment for science is:

none ___________

20-40 minutes ___________

40-60 minutes ___________

60-90 minutes ___________

more than 90 minutes ___________

at the teacher's discretion ___________

(3) In grades 4-6, the weekly allotment for science is:

none ___________

20-40 minutes ___________

40-60 minutes ___________

60-90 minutes ___________

more than 90 minutes ___________

at the teacher's discretion___________

(4) Science equipment and supplies comprise a separate item in the school budget: Yes_________ No_________

(5) The per student budget for science equipment and supplies for the current year is:

none ___________

less than $1 ___________

$1-$2 ___________

$2-$5 ___________

$5-$10 ___________

over $10 ___________

2. Suggestions for preparing questionnaires

The questionnaire is a convenient and useful device for gathering data about various aspects of the science program. Questionnaires generally are classified as either structured or unstructured. Structured questionnaires restrict the response to a word or phrase, and are fairly simple to complete and easy to tabulate and summarize. Unstructured questionnaires, on the other hand, permit a free response in the individual's own words. Although unstructured questionnaires sometimes can reveal more significant information, the responses frequently are difficult to interpret and tabulate. A questionnaire may include both structured and unstructured items. The supervisor should determine the type of response he wants for each item and use the appropriate forms to achieve his purposes. When he is not certain of exactly what he wants, he is more likely not to ask the right questions. A common difficulty involved in constructing questionnaires is that the items frequently are ambiguous and, as a result, are misinterpreted by the respondents.

The following are suggestions for helping supervisors improve the clarity of questionnaire items:

a. Define terms that are not clear or precise.

Poor: Are problem solving techniques employed in teaching science?
Yes_______ No_______

The meaning of the phrase "problem solving" in relation to teaching science must be made clear. It could mean solving a problem on a test, or solving a problem in the laboratory, or the approach used by the teacher in introducing a unit.

Improved: Is the problem-solving approach employed in presenting new materials?
Yes__________ No__________

b. Do not use such terms as "frequently," "sometimes," "often," or "rarely." Since they have only relative meaning, they are difficult to interpret. Instead, indicate some stated frequency such as "minutes per day," "times per week," or "hours per week."

Poor: How often does the supervisor visit each class?
Frequently_______ Sometimes_______
Rarely_______

Improved: How often does the supervisor visit each class?
Weekly______ Monthly______
2 to 5 times a year______
Once a year______ Never______

c. Construct items that deal only with a single idea.

Poor: Are superior students placed in special class groups and assigned to special laboratory sections?
Yes______ No______

Two questions are obviously needed to elicit an accurate response.

Improved: Are superior students placed in special class groups?
Yes______ No______

Are superior students assigned to special laboratory sections?
Yes______ No______

d. Provide a sufficient number of alternatives when alternatives are required.

Poor: Did you study science in college?
Yes______ No______

The alternatives supplied for this item are completely inadequate. If accurate information is desired about preservice science background, a sufficient number of alternatives relating to courses and credit-hours must be provided.

Improved: How many semester-hours of each of the following sciences did you complete at college?

Course	Semester-Hours
Biological Sciences	______
Physics	______
Chemistry	______
Astronomy	______
Geology	______
Other (please list):	
______________	______
______________	______
______________	______

e. Avoid making assumptions that are not justified.

Poor: Have the students benefited from science demonstrations?
Yes_______ No_______

A "no" answer could mean that demonstrations were not performed or that students did not derive any value from the experience. In addition, this item is unsatisfactory because a respondent would not be sure whether it referred to all students, or a majority of the students, or a single student. This can be improved by breaking it down into several questions:

Improved: How many demonstrations were performed during the year? _______

How many of these will be repeated next year? _______

List the three most effective demonstrations performed.

____
____
____
____

Why do you believe these were effective with the students?

____
____
____
____

List the three least effective demonstrations performed.

____
____
____

Why do you believe these were not effective with the students?

____
____
____

These suggestions will not make the questionnaire foolproof, but they will help in removing a good deal of the usual ambiguity. Questions that deal with a single idea, that are worded as simply as possible, and that have important terms defined, permit a respondent to provide accurate and relevant information.

3. Available checklists and questionnaires

Several checklists and questionnaires have been prepared for general use. Some of them may serve as formulated; others may be adapted to specific needs. The following titles issued by the U.S. Department of Health, Education, and Welfare, Office of Education are available:

Obourn, Ellsworth S. Science as a Way of Life. (OE-29023) Washington, D. C.: U.S. Department of Health, Education and Welfare, Office of Education, 1961. Washington, D.C.: Superintendent of Documents, Government Printing Office, 1961. 27 p. 15¢

U.S. Department of Health, Education, and Welfare, Office of Education. A Suggested Checklist for a Science Program. Selected Science Services, April 1962. (OE-29034) Washington, D.C.: The Office, 1962. Washington, D.C.: Superintendent of Documents, Government Printing Office, 1962. 11 p. 15¢

Piltz, Albert. Five Point Checklist for a Science Lesson. June 1960. Distributed by U.S. Office of Education.

Piltz, Albert. Ten Point Checklist for a Science Lesson. June 1960. Distributed by U.S. Office of Education.

The National Science Teachers Association has published A Checklist for Self-Evaluation, by Joseph Brzeinski (Elementary School Science Bulletin, No. 40) and Twenty Point Checklist for a Science Room (Elementary School Science Bulletin, No. 9). These and other checklists may be found in the Appendix to the bulletin.

SECTION II. EVALUATION BY CLASSROOM TEACHERS AND THEIR SUPERVISORS

Nature and Function of Evaluation of Student Progress

"What," "how," and "how well" are key words introducing three of the major questions faced by every supervisor and classroom teacher in the formal education of students. The first, "What should children learn?" is concerned with both the long-term goals and the consequent immediate objectives of the educational enterprise. The second, "How should they learn?" related to the problem of method or of determining the most effective means of helping students achieve the objectives. The third, "How well have they learned?" involves ascertaining the degree to which the objectives have been achieved. It is this careful appraisal of where a pupil is and how well he is progressing that comprises "student evaluation."

Teachers generally and rightly consider evaluation as one of the most complex problems in teaching and one for which they have been inadquately prepared. This is especially true in the area of elementary school science, where curriculums are undergoing extensive revisions not only in relation to content but also with respect to scientific attitudes and problem solving approaches. Standardized tests in elementary school science which would be appropriate for most schools are still nonexistent. Merely using a test labeled "science" could well result only in obtaining erroneous information about students and would be worse than not using tests at all. Thus, both the classroom teacher and the local personnel charged with the supervision of the elementary school science program are faced with the responsibility for learning how to develop, use, and interpret a variety of appraisal techniques for the purpose of furthering the teaching-learning process.

Measuring the outcomes of the teaching-learning process requires a great variety of evaluative techniques. For some purposes, testing devices such as objective tests or essay tests are not only satisfactory but even necessary. For example, objective tests can be used to determine if students can define such terms as "compound," "mixture," and "element"; essay tests can be used to determine if students are able to use these terms in formulating chemical explanations of natural or man-made phenomena.

Other kinds of achievement can be evaluated by nontesting techniques such as the rating of pupil-made products, or observation of a pupil's classroom performance, or the nature and extent of an individual child's behavior during selected science activities.

748-546 O - 64 - 4

Consider science fairs and the children's exhibits as an example. In addition to being excellent motivators, such activities can furnish the supervisor and teacher with many opportunities for evaluating how well given pupils are progressing towards pertinent goals of the science curriculum. Does the exhibit present scientific information clearly? Are the details of the exhibit relevant to the main concept? Can the student explain his exhibit so that it is obvious that he understands the principles involved? Such nontesting techniques are valuable in judging the progress of students and the effectiveness of programs.

Dependable evaluation is recognized as fundamental to both effective teaching and effective learning. Thus, evaluation procedures used by a teacher with his class can serve such functions as:

(1) Providing feedback information for the teacher and helping him decide how effective a given amount of teaching has been.

(2) Supplying diagnostic information concerning individual pupils.

(3) Providing motivation for students. As pupils see progress toward established goals, they can be motivated to further learning.

(4) Affording sound evidence for differentiating among pupils.

1. Defining objectives

The first and most important step in evaluating teaching and learning is to define the objectives that are to be attained, and questions like the following may prove useful:

a. What problem-solving abilities is the student expected to achieve? Can he describe what he saw when he watched a candle being snuffed as a jar covered it? Can he formulate hypotheses about the observed phenomena? Can he plan experiments to test his hypotheses?

b. What information is the student expected to acquire? Has he learned the components of air? Does he know about changes in percentages of these components depending upon altitude or other variables?

c. What skills is he supposed to display? Can he measure amounts of air that are consumed by a candle burned in a jar? Can he manipulate science equipment so that he can perform experiments with air?

d. How is he expected to reveal his attitudes? Does he defer judgment when he observes a demonstration? Does he consider the reliability of a source of information?

e. What applications of his knowledge is the student expected to make? Can he apply what he learned about candles burning in a jar to other problems involving combustion? Does he understand that rusting of iron is another form of oxidation?

All school programs have objectives, but too often these are stated in very general terms and are vague and unclear. Before a supervisor or teacher can determine whether a student has reached certain objectives, he must be able to identify specifically what the student was supposed to achieve. For example, "to help students develop a wholesome attitude toward science" or "to help students appreciate the methods of science" are purposes that have little meaning and would not be useful as guides for evaluation. Objectives must be clearly and specifically defined in terms of pupil behavior, and the performance expected of a student if the objective has been achieved must be specified. If a teacher is unable to list the characteristic behaviors of a student who has reached a particular objective, it is meaningless.

The job of defining objectives in terms of student behavior is recognized as an extremely difficult one, but it is basic to effective evaluation. For example, consider the long-term goal "to help children learn some techniques of problem-solving." Among the specific pupil behaviors that might be expected when this goal is achieved are the abilities to:

(1) State a variety of hypotheses concerning the problem,

(2) Plan experiments for testing these hypotheses,

(3) Report observed phenomena, and

(4) Generalize from what has been observed.

As a second example, the long-term science goal "to help children develop rational attitudes toward the world around them" may be cited. When this is specified in terms of pupil behavior, it might require that the child be able:

(1) To identify certain superstitions, and

(2) To identify sources of information and consider their dependability.

In every science unit, certain of the attitudes, skills, and techniques that have been defined generally must be spelled out so that they relate to the science to be taught. Thus, in a study of weather, growing out of the goal of identifying superstitions might be the following immediate objectives. The pupil can:

(1) Distinguish between "weather superstitions" and "weather facts," and

(2) Offer explanations for the origins of some "weather superstitions."

In a similar manner, the goal of identifying sources of information and considering their dependability might be formulated as the following objective. The pupil can:

> Compare the reliability of weather information from such varied sources as the United States Weather Bureau and the Farmer's Almanac.

The examples cited above suggest only a few of the kinds of behavior which would indicate achievement of the desired goals. There are, in every case, many different kinds of behavior which could be examined. In fact, there are far too many behavioral characteristics for adequate consideration by teachers in elementary school classrooms, and some of the desired behaviors, while very important, do not lend themselves to evaluation in classroom situations. What the teacher must do is to choose those particular behavior patterns which he feels are important and measurable within the framework of his classroom, and use them as his criteria for evaluation.

2. Assigning relative emphasis to objectives

Every unit in elementary school science has a variety of objectives which students are expected to attain. However, all of the objectives are not of equal importance. If an objective has been assigned major emphasis, it should be allotted a greater proportion of the class time; other objectives, which have been assigned lesser emphasis, should be allotted relatively smaller amounts of time. Therefore, to simplify the process of planning a unit of study, the weight assigned to each objective should be specified. This can be easily accomplished by using the numerical scale of 100 and distributing numerical weights to each objective. Higher weights should be assigned to major objectives and lower weights to minor objectives. If the objectives are all determined to be of equal importance, equal weight should be assigned to each.

This simple system of assigning weights to objectives has real value for the teacher in planning both the content and the evaluation. He now has a measure for determining how much class time should be allotted to the achievement of each objective, as well as an index of the degree to which the objective should be emphasized in the evaluation process. For example, if major emphasis is placed on helping students learn to use the microscope, this should be emphasized to a similar extent in the total evaluation of the student's achievement. If identifying the parts of the microscope is considered a minor objective, it should receive relatively little emphasis in the appraisal process.

The procedure of assigning weights to objectives does not have to follow a rigid formula but may be as flexible as a teacher desires. The emphasis allotted in a classroom quiz may differ from the emphasis assigned in a weekly test. In addition, the distribution of emphasis may be modified from group to group, depending on such factors as the students' readiness for achieving certain objectives, their current interests, and the availability of necessary materials and equipment for developing particular areas of the elementary science program. But regardless of what evaluation procedure is used, the important factor is that the procedure be planned so that it achieves the purpose for which it was intended.

3. Outlining the content

Once the objectives have been clearly specified and defined, the next important step in the evaluation process is outlining the content. The content of the unit or area of study becomes the actual means of achieving the objectives. The term "curriculum" frequently is defined as a means to behavioral ends. The content of a unit is the curriculum for attaining the specific behavioral objectives of that unit. By outlining the content, the teacher is able to relate the particular objective to the method of achieving the objective. Frequently, the same content can be utilized in attaining several objectives; sometimes a variety of methods have to be developed for achieving a single objective. For example, if an objective of a science unit is to help children learn to use measuring instruments such as therometers, the content might include activities like reading daily temperatures in and outside of the classroom, measuring termperatures in sun and shade, or measuring temperatures of hot and cold liquids. This content also could be used for achieving other objectives relating to heat absorption, heat reflection, heat transfer, or insulation.

Another way in which outlining the content can have value for the teacher is in predetermining the specific materials and equipment which must be available for achieving certain objectives. If children are to use thermometers, such instruments must be available in sufficient quantity so that each child may have several opportunities for using the instrument. If skill in doing reference work about geologic periods is in the objective, a variety of appropriate books on historical geology must be available so that each child may read from various sources and compare the information obtained.

Although this may appear to be a time-consuming task for the teacher, it actually can simplify the arduous job of preparing conventional lesson plans and, in addition, specify the evaluation procedure to be used in determining student progress toward particular objectives. A simple way of arranging objectives, content, and evaluation for a unit is to use a three-column table. The first column would contain the objectives; the second column would indicate the methods and materials necessary for achieving each objective; and the third column would specify the type of evaluation procedure to be used in determining student achievement of each objective.

In the first table one objective requires two procedures for evaluation:

Objective	Methods and materials	Type of evaluation
The ability to record and interpret temperature data from information gathered using outdoor thermometers	Make simple bar graphs of the daily temperature as found at noon in the shade next to the building. Use the graphs to find significant temperature information; make comparisons among the days studied and also among reading made by different children. Have sufficient supply of outdoor thermometers so that variations in readings can be observed. Have necessary graph grids prepared so that each child has appropriate equipment.	Examine the charts prepared by each child. Does the chart show that the child can make an accurate and understandable table of data. Devise written questions to be answered by individual children, each using his own chart: Which day was hotest? Which day was coldest? On which two days was the temperature about the same at noon.

In the second table, two objectives may be evaluated through a single procedure:

Objective	Methods and materials	Type of evaluation
The ability to use library resources to do reference work	Introduce the use of the card catalog and encyclopedias, to show a) How to search out appropriate references; b) How to make summaries; c) How to make a bibliography.	Examine the summaries for accuracy of reporting, comparison of material found in different sources, completeness of bibliography, comprehension of content.
The ability to prepare a summary of information found on the last glacial period in North America	Have each child find at least four appropriate references. Make sure there are sufficient references for the children's use. Make sure there are reading materials at the appropriate reading level for each of the children.	

Evaluation Techniques

The teacher who is concerned with evaluation will recognize that he must select appropriate evaluation methods for each of his educational objectives. For certain objectives, the overt behavior of the student - in the classroom situation, in the library, in the laboratory - would yield the most appropriate information. For other objectives, where opportunities for students to demonstrate their actual behavior are very limited or do not exist, pencil-and-paper tests would be more satisfactory. Some objectives are more difficult to measure than are others; for some, appropriate measuring devices have not yet been developed. However, because of the intimate relationship which exists between objectives and evaluation, objectives must be stated in concrete and specific ways if evaluative results are to lend themselves to precise statements; objectives that are stated in vague and general terms result in evaluation methods that yield incomplete and inaccurate results.

The various kinds of evaluation -- observation of student behavior, appraisal of student projects, and the variety of paper-and-pencil tests -- yield a wide range of information. But unless sufficient use is made of the results of such observations, appraisals, and measurements, they become a waste of time both for the supervisor or teacher and for students. Such evaluative techniques have three basic uses. In the first place, they can provide a means for assessing the growth of individual students. Secondly, such devices can help each student know his strengths and weaknesses. Finally, through the use of these techniques, a supervisor or teacher can learn how well the objectives of a program are being met.

It must be remembered, however, that the supervisor or teacher is the ultimate evaluator. A test or other evaluative device can gather information about a student or a program. But only the evaluator can look at this information, weight it, add and subtract related data, and come up with an appraisal of where a student is in relation to his potential, or how well a program is attaining its stated objectives.

1. Selection of appropriate devices

There is no single evaluation procedure which is best for judging student achievement. Each method has its advantages and limitations, depending on the situation in which it is used. The key to selecting the most appropriate technique is the behavioral objective the teacher wishes to measure. How should a teacher appraise a child's ability to work cooperatively with others on a science problem? A written test for this kind of appraisal would be pointless. Observation of the students in a variety of classroom situations would seem more appropriate in determining progress toward this objective, but even "observation" has its limitations and must not be counted on for evaluating too many objectives. Two criteria for the effectiveness of observation as an evaluative technique are suggested:

(1) Can the desired behavior be evoked in the classroom?

(2) Has the teacher sufficient opportunity to observe and record what he sees?

Take another example: How should skill in setting up laboratory apparatus be measured? Often, teachers wishing to make such a determination give a written test and make their evaluations on the basis of such test results. A moment's consideration, however, shows that such tests are nowhere near as appropriate as having the student actually carry out the desired laboratory procedure. In fact, it is quite conceivable that students could answer correctly all the questions relating to the ways in which a Bunsen burner should be lit, but, placed in the actual laboratory situation, be unable to light the burner correctly and safely. Again, the kind of test that is given must be determined by the kind of behavior to be observed and evaluated.

Under what conditions, then, are pencil-and-paper tests appropriate? What can such tests indicate? If the teacher is concerned with determining how well students can explain certain phenomena, compare materials from several sources, make inferences, draw conclusions, or select significant factors, pencil-and-paper tests are appropriate vehicles.

It is also necessary to recognize that for measuring certain desired outcomes present devices for evaluating student behavior are still inadequate. Results of attempts to evaluate such behaviors are highly unreliable and must be treated with this fact in mind. One such behavior pattern relates to the use of science information in the choice of a well-balanced diet. Obviously, young children are not in a position to purchase and prepare their own meals. Thus, the desired behavioral objective cannot be measured adequately at this time. In short, each teacher must recognize what he is trying to measure and to what extent his measuring device is effective.

2. Validity and reliability of evaluation techniques

There are two important factors that teachers must consider before using any procedure or instrument for determining student growth. The first and foremost factor is validity, the second is reliability. An instrument is valid insofar as it measures what it is supposed to measure. An instrument is reliable insofar as it measures accurately. A thermometer can be perfectly reliable in that it measures temperature accurately, but the thermometer alone hardly would yield valid results in measuring the caloric content of a substance.

The distinction between reliability and validity can be seen clearly when one examines the stated objectives of a given science program and then considers almost any of the currently available achievement tests in elementary school science. The tests all report high reliability. In general, these reports are true; the tests do measure accurately what they set out to measure. But an inspection of the content of these tests and a comparison of the items in the test to the stated objectives of a particular science class readily show that there is little relationship between the items found in the test and the objectives for which elementary science is being taught. Actually, the validity of these tests for measuring what is being taught in science classes is extremely low.

748-546 O - 64 - 5

The ideal evaluation technique must serve the purposes for which it is intended and it must yield information that is accurate.

Since validity is so important in evaluation, teachers must be very much concerned with this factor both in selecting commercial testing instruments for measuring the work of the students in their classes and in preparing their own instruments. The validity of any teacher-made instrument will depend on the degree of correspondence between the behavior that is to be appraised and the objectives that have been set for given instruction. It is for this reason that so much emphasis has been placed on identifying the behavioral objectives as the first step in the evaluation process. If the teacher can spell out the way in which a student will act after he has learned a given item or successfully mastered a given unit of work, he can then construct valid evaluative instruments by developing situations in which the student will be able to indicate how effectively he has mastered the desired behavior. For example, if it is desired that student know the parts of an electric motor, a pencil-and-paper test which requires the student to describe these parts could be satisfactory.

However, suppose the stated objective of a unit is to develop an understanding of the interrelationships of the parts of an electric motor. A test which asks students only to identify or describe the parts of the motor would not be valid in relation to the stated objective. Some kind of test situation which determined the extent to which the student understood the interrelationships among the parts would be necessary. A valid test might ask the student to explain the relationship of the armature to the brushes. Or he might be asked to explain the relationship between the stator and the rotor. For the test situation to be valid, it must elicit responses that are very similar or identical to the objectives of the unit being studied.

Another important factor which teachers should be aware of both in selecting and in making valid instruments is the vocabulary used for the test situation. If a test or an item in a test is worded so that it can be understood by only a few students in the class, such an item is not valid for measuring the learning outcomes of the other members of the class. This is especially true in the primary grades, but also holds true for other groups in which there are wide variations in reading ability and comprehension.

Reliability of commercially prepared testing materials is generally not a problem. Such materials usually are carefully designed and controlled for reliability. But the reliability of teacher-made classroom tests depends upon a few special considerations.

The usual and conventional methods of developing reliable tests of the standardized variety generally are impractical for teacher-made tests. However, since the major reasons for the low reliability of teacher-made tests are that such tests are either too short or too difficult, remedies can be found which will help teachers prepare more reliable tests.

First, teachers should include large numbers of items in the test; second, each test should include an adequate number of items which most students will be able to answer. For example, if a test is to be constructed to determine the ability of students to classify foods as proteins, carbohydrates, or fats, listing merely three foods - nuts, dried beans, and apples - would make a test that has little reliability and, incidentally, little validity. To improve the reliability of the test, large numbers of foods of each of the three categories would need to be included. Furthermore, among the items included would need to be simpler ones such as meat, butter, and sugar. Adequate samplings must be included if an evaluation device is to have high reliability and high validity.

Finally, adequate samplings are needed so that the teacher can determine how well - to what extent - a student is meeting the objectives which have been established for him and for the class as a whole. It is the student's performance on the samples included in a test that enables the teacher to judge the student's progress in the total area which has been studied. To obtain sound results for such evaluations, test samples must be sufficiently large and sufficiently varied in difficulty.

Categories of Classroom Tests

The classroom test is still the foundation of the day-to-day evaluation program of most schools. Yet many of these tests have poorly framed items, confusing directions, ambiguous statements, and other flaws that seriously impair the usefulness of the scores they yield. Some of the basic principles of test construction are summarized below so that teachers will be able to derive full value from the use of these instruments in evaluating student growth.

Classroom tests generally are classified into two broad categories, essay tests and objective tests. The essay or free-response examination permits the student to compose and express his answer in his own words. The response may range from a few sentences to several pages, and the accuracy and quality of the response is judged subjectively by a person who is competent in the field.

Objective tests restrict the student's response to a symbol, word, or phrase, and subjective judgment is practically eliminated in determining the accuracy of the answer. The term "objective" as applied to objective tests refers to the scoring of the response and not to the choice of the content. Objective-test items generally are classified as supply or selection items. In responding to a supply item, the student provides the necessary word, phrase, or symbol. In responding to a selection item, the student chooses a response from among those presented to him.

1. Essay tests

Essay tests have certain advantages that cannot be matched by any other form of evaluation. The chief merit of the essay item is that it provides the student with an opportunity to demonstrate the degree to which he can analyze a problem, select relevant information, present evidence, and organize his answers logically and effectively. In addition, since no answer need be completely right or completely wrong, it is possible for a teacher to determine the degree of correctness of a student's response.

Despite these distinct advantages, essay tests have come under considerable attack in recent years because of certain glaring weaknesses. Many essay tests, as they are currently used in various schools, measure nothing more than the ability to reproduce information. Merely phrasing a question in essay form does not automatically guarantee that progress toward such goals as recognizing causal relationships, applying principles, or making generalizations will be assessed. In addition to poor design of questions, the essay test frequently suffers from inadequate sampling, from highly subjective and inconsistent scoring, and from the influence of such extraneous and irrelevant factors as literary skill and handwriting. However, these are weaknesses that can be overcome, and the following guidelines are suggested to help teachers develop greater skill in designing appropriate items and in improving their methods of evaluating student answers.

a. Limit the use of essay items to those objectives that are measured most efficiently by the essay format. For example, the question:

What is the accepted composition of air at sea level?
tests only for specific facts, namely, the percentages of various components of air. It illustrates inefficient use of an essay item. The essay item should be reserved for evaluating progress toward more complex educational goals than merely reproducing information. The following example, based on the same general topic, is more appropriate for an essay item:

> Compare the composition of the air of a large community and a small farm community, and account for the differences.

In answering this question, the student would have to demonstrate his ability to use information to interpret data.

b. Improve effectiveness of essay items by requiring the student to use knowledge in situations that have not been discussed directly in class. Thus, if students have studied the use of the lever, inclined plane, and wheel, the following instruction might be appropriate:

> There is a large stone in the playground which is too heavy to lift and carry away. Explain how you could use simple machines to remove the stone.

In this situation, the student would be using information learned in class to solve a problem that has not been discussed previously.

Another example providing the student with an opportunity to demonstrate his ability to apply principles might result from a study of the properties of metals. In this case, the student could be asked to:

> Explain three ways in which you could test the "lead" in a pencil to determine if it is a metal.

Similarly, after a study of experimental procedures, the student could be presented with the following situation:

> John heard that weak tea makes plants grow better than tap water. He set up an experiment to find out if this is true. First, he obtained two identical plants. He kept one plant on the shelf along the wall and the other plant on the window sill. He watered each plant daily. He used tap water on one plant and a weak tea solution on the other plant. He found that the plant which was given the tap water grew better than did the plant that was given a weak tea solution.

The student then could be asked any number of questions relating to this situation, such as:

> On the basis of this experiment, should John conclude that tap water is better for plants than weak tea? Why?

or:

> How would you perform the experiment to determine if weak tea is better for plants than tap water?

748-546 O - 64 - 6

Naturally, the quality of the response would depend upon the grade level of the students, their familiarity with experimental methods, and their ability to make accurate inferences.

c. Frame essay items that measure ability to apply principles, recognize relationships, or make generalizations more effectively by starting with such phrases as "Explain how," "Explain why," "Compare," "Interpret," "Show the relationships," and "Give reasons for." Essay items which start with such words as "what," "who," or "list" generally require that the student merely reproduce certain facts. Essay items that start with such phrases as "What is your opinion of" or "What do you think of" are usually inappropriate for measuring various facets of educational achievement. Frequently, the teacher who uses this type of phrase actually is concerned with the student's ability to analyze a situation or support a particular position and not with the giving of a personal opinion. Therefore, it would be preferable for the teacher to rephrase the question so that the desired response could be elicited.

d. Word essay questions clearly so that the answers which students give will be limited to the specific objectives which are being measured. Too often essay items are so vague and ill-defined that pupils are forced to guess what the teacher wanted. If a student guesses wrong through no fault of his own, or if he interprets a question one way while the teacher wants a different interpretation, the responses become impossible to score, and the advantage of using essay items is lost. Thus, the question:

> What effect will atomic energy have on the world?

is much too broad and vague. The response could be limited to the destructive properties of atomic energy, or the constructive uses to which it can be set, or both. To be sure that each student will interpret it the same way, the following statement is better:

> Plans are now being made for the peacetime use of atomic energy. Give two examples of how atomic energy could be used in agriculture.

In the rephrased statement, there is no uncertainty about what is wanted and about the specific areas to be discussed.

e. Allow sufficient time for students to answer essay questions. Since essay items are used to evaluate the more complex educational goals which require a good deal of thought, the student must have adequate time for analyzing the question, organizing his answer, and then writing it. When students are pressed for time, their responses frequently show careless thinking and sloppy writing.

f. Score every objective that is to be measured by the essay question independently. The grading of correct factual information should be judged separately from the grading of organization of material. If grammar, spelling, or writing style are included in the objectives of the unit, these areas should also be scored, but scored separately from the other educational objectives. If only a single score is given for an essay response, the student has no way of knowing how well he has progressed toward each of the objectives established for the particular science unit.

g. Prepare scoring guides in advance. By so doing, judging essay responses can be made more reliable. One of the chief disadvantages of using essay questions has been the inconsistency of the scoring methods. Not only have different teachers reading the same answer come up with divergent scores, but the same teacher reading the same answer has reported different scores on different days. To eliminate such inconsistencies, teachers can prepare a model answer in advance, indicating the factors that should be covered and the credits assigned to each factor. This guide can provide a more uniform basis for evaluating the written responses of each student.

h. Administer several essay tests during the school year to increase the sampling of subject matter. Generally, adequate subject matter sampling can be obtained more satisfactorily through the use of other measuring devices. However, where the essay test is the best instrument for measuring progress toward a goal, it should be used. In essay tests, adequate sampling of subject matter can only be provided for by increasing the number of essay items used. This is not feasible because of limitations of class time; therefore, several tests are necessary.

2. Objective tests

The objective test was introduced into the classroom to overcome some of the weaknesses of the essay test. One obvious advantage of the objective test is that it permits extensive sampling of the topics covered, whereas the essay test tends to limit the amount of subject matter that can be sampled. Another advantage of the objective test is that the answers can be scored quickly and objectively. In the essay test, scoring is generally time-consuming and sometimes unreliable. The major complaint made against the use of the objective item is that it tends to measure bits of superficial and random information rather than broad understandings and more complex abilities. But this limitation, when examined carefully, seems to be more the fault of the person constructing the test items than of the inherent nature of the test itself.

Items can be constructed that test not only for knowledge but also for the more complex abilities of understanding and reasoning. However, designing such items for an objective test is far more difficult than preparing similar items for an essay test.

It is no longer a question of which kind of test to use because both essay and objective tests can be used to advantage in the classroom. The most important factor is how well an item is constructed. A poorly constructed test fails to achieve its purpose and actually can interfere with the learning process. Therefore, certain guiding principles are offered here to assist the teacher in constructing objective tests so that greater benefits will be derived from the classroom evaluation program.

a. Supply items

One major type of objective-test item is the supply item. In a supply-item test, the student is required to provide information, usually in the form of a word or a phrase. Generally speaking, there are two kinds of supply items, the short-answer and the completion item. If the problem is presented in question form, it is a short-answer item. If the problem is presented as an incomplete statement, it is a completion item. The following examples show how the same information can be elicited from both forms:

(1) Short Answer: What is the source of energy in a flashlight?

Completion: The source of energy in a flashlight is the_______________.

(2) Short Answer: What is the atmospheric pressure at sea level in pounds per square inch?

Completion: The atmospheric pressure at sea level in pounds per square inch is________.

or: The atmospheric pressure at sea level is______________ pounds per square inch.

(3) Short Answer: What is the chemical formula for hydrochloric acid?

Completion: The chemical formula for hydrochloric acid is_________________.

Supply items emphasize recall of information and are satisfactory for measuring knowledge of specific facts, names, dates, and simple computations. In addition, supply items allow the teacher to sample a large body of subject matter in a relatively brief period of time.

In a supply test, the probability of a student guessing the correct answer is reduced to a minimum. However, these items are not well suited for measuring the more complex abilities of understanding and reasoning.

Suggestions for Constructing Supply Items

(1) Design items that avoid misinterpretation and require one correct response. The following shows how a poorly stated item can be improved:

Poor: The two most common gases in the air are____________and____________.

Although the teacher expects students to respond with "oxygen" and "nitrogen," it would not be surprising to find some pupils responding with "invisible" and "important" or any two other qualities. It would be difficult to score such an answer since it is actually correct even though it is not the answer desired.

Improved: The names of the two most common gases in the air are__________and__________.

(2) Design items that require only one or two completions to be made in a statement. When statements are interrupted by many blanks, the meaning of the item is destroyed, and students are forced to resort to guessing.

Poor: A____________is an ____________for measuring the____________of the air.

Improved: A barometer is an instrument for measuring the____________of the air.

(3) Place blanks near or at the end of a statement. When blanks are placed at or near the beginning of statement, the student generally must read the statement twice before being able to supply the answer.

Poor: A(n)__________measures the speed of the wind.

Improved: The speed of the wind is measured by a(n) __________.

(4) Do not provide clues to the correct answer. In the previous example, the article is listed as "a" or "an" so that the student who does not have accurate information cannot guess at the correct response. In addition, the length of the blank should not offer the student a clue to the length of the word omitted. It is a good policy to make all blanks a uniform length, but long enough so that the child has room to write his answer.

(5) Specify the units in which a numerical answer is to be given.

Poor: The freezing point of distilled water is ______________ degrees.

Improved: The freezing point of distilled water is ______________degrees Fahrenheit.

b. Selection items

Another major group of objective tests is the selection item. In a selection test, the student is required to select a response from among those presented to him. Selection items are also referred to as recognition items and include true-false, multiple choice, and matching items.

(1) True-false items

The true-false test is perhaps the most widely used of all selection tests. It generally consists of a simple declarative statement to be judged true or false, such as:

True. False. It is the oxygen in the air which supports combustion.

A variation of the traditional true-false test is sometimes employed which requires the student to correct the item if it is false. The student must supply the correct answer in the blank provided, for example:

The sun is a planet. True. (False.) Star

This modified true-false item helps to reduce guessing, and thus provides the teacher with more valid information about the student's knowledge.

The chief advantage in using true-false tests is that the teacher can sample a large body of subject matter in a short period of time. However, the tests are appropriate only for measuring specific pieces of information, rather than broad understandings. True-false items are also very difficult to construct because they have to be limited to statements that are either completely true or absolutely false. As a result, many of the items that are seen on true-false tests are ambiguous and pose difficult problems, especially for the bright student. Another weakness of true-false tests is that they encourage guessing.

Suggestions for Constructing True-False Items

(a) Use statements that are completely true or absolutely false. One of the glaring weaknesses of true-false items is that the capable student generally can think of certain exceptions, and thus finds them difficult to answer, as in this example:

Poor: T. F. The boiling point of water is 212° Fahrenheit.

Improved: T. F. The boiling point of distilled water at sea level is 212° Fahrenheit.

In the first example, certain information related to atmospheric pressure and kind of water is omitted, which might pose a real problem for these students who see the need for additional qualifications before answering the statement as true. The second example takes these factors into account.

(b) Avoid using specific determiners that give students clues to the probable answer. Words that tend to identify a statement containing them as true or false are called specific determiners. For example, such words as "always," "never," "none," and "all" are found in statements that are likely to be false. On the other hand, words such as "sometimes," "may," "usually," and "could" are found in statements likely to be true, as for example:

T. F. Evaporation always takes place more rapidly in summer than in winter.

T.F. Evaporation sometimes takes place more rapidly in summer than in winter.

In the above examples, a student without specific knowledge in the area could probably answer these statements correctly by using the specific determiners "always" and "sometimes" as clues.

(c) Avoid using negative statements. Negatives tend either to confuse students by complicating the meaning of the statement or to cause careless errors when students overlook them, as in the following:

T. F. Mercury is not a metal.

(d) Avoid lengthy and involved statements. On the one hand, statements that are lengthy are frequently true. On the other hand, they needlessly prevent the pupil from readily recognizing the important factor in the item, for example:

T. F. When a person moves from New York City to Denver, he finds that the reduced atmospheric pressure at higher altitudes alters the forces exerted on water molecules, and as a result, water changes to steam at a lower temperature than along the seacoast.

(e) Avoid using statements that are partly true and partly false. This again leads to confusion and obscures the real purpose of the test item.

T. F. Oxygen, a gas which supports combustion, was discovered by Newton in 1736.

(2) Multiple-choice items

A second type of selection item is the multiple-choice test. In a multiple-choice item, the student is given an introductory statement, called the stem, and several alternative answers from which he must select the one that is most appropriate. The introductory statement or stem may be in the form of a question or an incomplete statement as follows:

Question form: Which of the following foods is the best source of vitamin C?
(a) raisins (b) grapefruit
(b) pears (d) cherries

Incomplete statement: We get the most iron from a normal serving of (a) fish (b) veal (c) liver (d) ham

Incomplete statement: We get the highest caloric value from one ounce of (a) lean beef (b) banana (c) white bread (d) butter

In the above examples, four options are included for each stem. There is no fixed rule regarding the number of options used, but generally four or five possible responses are listed becuase guessing is then reduced to a minimum. With younger children, however, fewer options can be given without destroying the effectiveness of the item.

The multiple-choice item is considered the most valuable and most flexible of all objective items. It can be used to measure the degree to which a student can recall factual knowledge as well as measure the degree to which he can use the more complex abilities of understanding and reasoning. Many content areas can be sampled adequately even though the amount of time needed for answering multiple-choice items is greater than for true-false items.

The most serious drawback of multiple-choice items is that plausible distractors are difficult to construct. As a result, teachers sometimes use options that are obviously incorrect or resort to such alternatives as "all of these" or "none of these," which are more often wrong answers rather than right answers. It would be preferable for teachers to use fewer options than to weaken the multiple-choice item by presenting distractors that do not seem plausible to the student. Another limitation of the multiple-choice item is that it cannot measure the ability of pupils to organize and present their ideas.

Suggestions for Constructing Multiple-Choice Items

(a) Word the stem clearly and meaningfully. The stem should present a single problem adequately. Teachers who have had little experience in constructing multiple-choice items probably will find that it is easier to state the central problem when the stem is in the form of a question than when it is in the form of an incomplete statement. When an incomplete statement does not present a specific problem, the alternatives merely become a series of independent true-false statements with the student deciding which one is more correct than the others, for example:

Poor: A study of plants tells us that
1) green plants store food only in leaves and stems
2) some green plants grow from bulbs
3) green plants need only air, heat, and water to stay alive
4) green plants and animals do not have common needs.

It is rather obvious that this item does not present a definite problem. Instead of the student being asked to select the best of four choices concerning a single problem, he actually is involved in deciding which of four somewhat related true-false statements is more true than the others. One suggestion which has been made for determining whether there is a central problem in the stem of a multiple-choice item is to cover the alternatives and see whether the stem, standing by itself, points to a definite problem. This would not be the case in the stem illustrated above. However, it could be improved as follows:

Improved: A study of the ways in which green plants react to sunlight shows . . .

(b) Include in the stem as much of the item as is possible and especially any words that would otherwise have to be repeated in each alternative. Thus, items are improved because after reading the stem, the student knows exactly what to look for before he examines the alternatives, as in the following:

> The temperature of the water for sterilizing baby bottles at home should be 1) 112° F., 2) 212° F., 3) 100° F., 4) 312° F.

(c) Design distractors that are plausible to students. The distractors should appear to be reasonable answers to students who do not have the knowledge required by the item. When some alternatives are obviously incorrect, students with inadequate understanding of the material can arrive at the correct response by the process of elimination, for example:

Poor: The process of nuclear fission normally is started in a nuclear reactor by

1) neutrons hitting atomic nuclei

2) earthquakes

3) releasing electrons

4) volcanic explosions

This item can be improved by substituting for the implausible responses 2) and 4) new responses that are more closely related to the others, such as:

Improved: The process of nuclear fission normally is started in a nuclear reactor by

1) neutrons hitting atomic nuclei

2) uniting atomic nuclei

3) releasing electrons

4) neutralizing protons

(d) State the problem in positive form. The use of negatives tends to confuse the student and causes careless errors.

Poor: Which of the following is not an element?
1) mercury 2) oxygen 3) salt 4) hydrogen

Improved: Which one of the following is an element?
1) mercury 2) peroxide 3) salt 4) hydrocarbon

(e) Construct responses that are **grammatically** consistent with the stem. A correct sentence should be formed when each alternative is attached to the incomplete statement. Cues resulting from grammatical inconsistencies should be avoided.

Poor: The voltage in an alternating current circuit can be stepped down by a

1) transformer

2) induction coil

3) oscillator

4) alternator

Improved: The voltage in an alternating current circuit can be stepped down by a

1) transformer

2) rectifier

3) magneto

4) condensor

The grammatical inconsistency in the first example could also be remedied by removing the article "a" from the stem and using the appropriate article with each option.

(f) Use situations that the student has not previously encountered in class when designing items to measure such abilities as reasoning, problem solving, or any of the other higher mental processes. If students are presented with items that have already been used in the text or discussed in the classroom, the teacher may be measuring only rote memory rather than thinking ability.

(3) Matching items

A third type of selection test is the matching-item test. Typically, such a test consists of two columns of items which are to be associated on some directed basis. The first column is called a list of premises and the second column a list of responses. In the simplest form, the two columns have the same number of items, but the matching test can be made more complex by increasing the number of responses or requiring the use of more than one response item for some items in the list of premises. For most elementary school programs, however, the simpler test is more appropriate:

Directions: In the space next to each item in Column I, place the letter of the phrase in Column II which defines it best.

Column I	Column II
____1. Force	A. The rate of doing work
____2. Energy	B. A push or pull
____3. Power	C. Capacity for doing work
____4. Speed	D. Rate of change of position

Matching tests are particularly well-suited for measuring a large body of factual information in a relatively short period of testing time. Matching tests can show whether a student is able to associate events with persons or places, terms with their definitions, principles with examples, and chemical symbols with names of chemicals. Matching items can be scored quickly and objectively, and when the items are well-designed, guessing is reduced to a minimum. The major disadvantage of the matching test is that its use is restricted to a limited number of subject areas. Since the items must bear some relationship to each other, it is often difficult and even impossible to find a sufficient number of related items in all areas of subject content. Another weakness of the matching test is that good items that are not completely obvious are hard to construct.

Suggestions for Constructing Matching Items

(a) The items in the list of premises and the items in the list of responses should be as homogeneous as possible. One method for determining homogeneity is to see whether all of the items in a column can be described accurately by one term. In the following example, this cannot be done:

Poor:

Column I	Column II
____1. mammal	A. Pasteur
____2. insect	B. cat
____3. scientist	C. mosquito
____4. gas	D. hydrogen

It is obvious that the problem presented in this example could be solved by students with the most superficial knowledge merely by the process of elimination. The items are so heterogeneous that no item in Column I could in any way be related to more than a single item in Column II. In the next example, only homogeneous items are used:

Improved:

Column I	Column II
_____1. anenometer	A. measures atmospheric pressure
_____2. barometer	B. measures temperature
_____3. hygrometer	C. measures wind velocity
_____4. thermometer	D. measures humidity

(b) The directions should specify clearly the basis for matching the items. The purpose in providing explicit directions is to avoid confusion and clarify for the student the task he is to perform even in situations where the basis for matching seems obvious.

Poor: Match items in Column I with Column II.

Improved: The following problem presents a column listing weather instruments and a column listing what they measure. In the space next to each item in Column I, place the letter of the phrase in Column II which defines it best.

(c) The premises and responses should be arranged in logical order whenever possible. If dates are used, they should be arranged chronologically, and if names are used, they should be arranged alphabetically. This simplifies the task for the student and reduces the amount of time needed for answering these items.

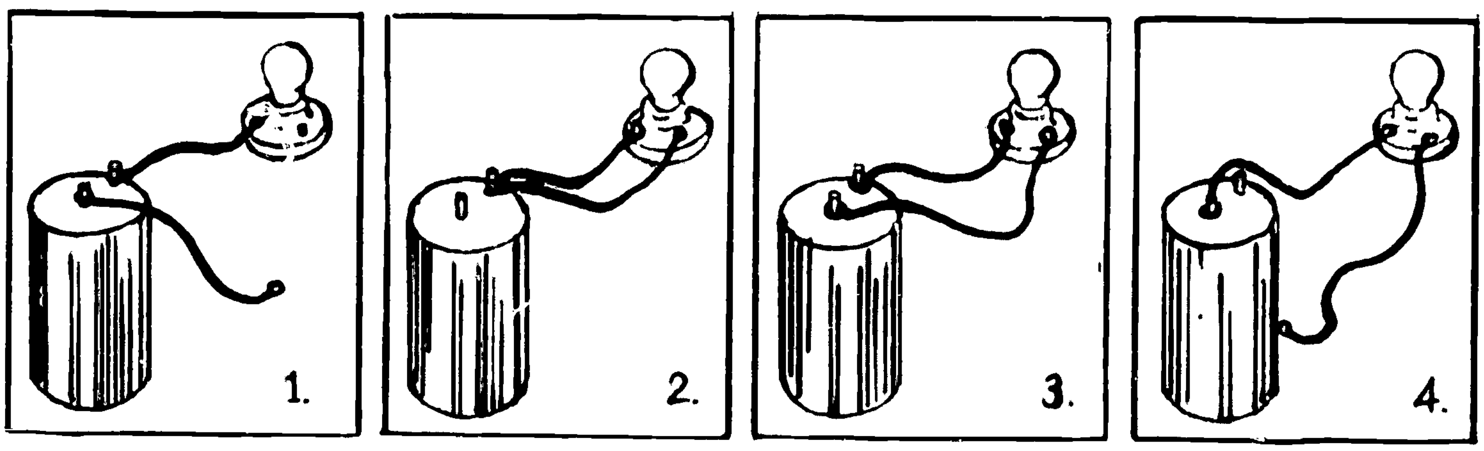
c. Objective tests using pictures

Objective test items based upon pictorial material can be adapted to measuring a variety of objectives including ability to recall information, interpret data, and apply principles.

Furthermore, they are versatile enough to be used in all grades of the elementary school and especially for students with limited reading comprehension. Since relatively few words are needed for this sort of item to be understood, the teacher can give the instructions orally. Test items based upon pictures can provide the student with clear and unambiguous problems that are interesting, novel, and realistic. It is true that there are certain topics that do not lend themselves to pictorial representation and that some teachers may be poor artists. Nothing can be done about the first problem, but teachers who have little skill in art can find appropriate material in books and magazines which they can either copy or trace. Just using pictures is of little value unless the pictures improve the test item and communicate the problem more effectively than the words they replace.

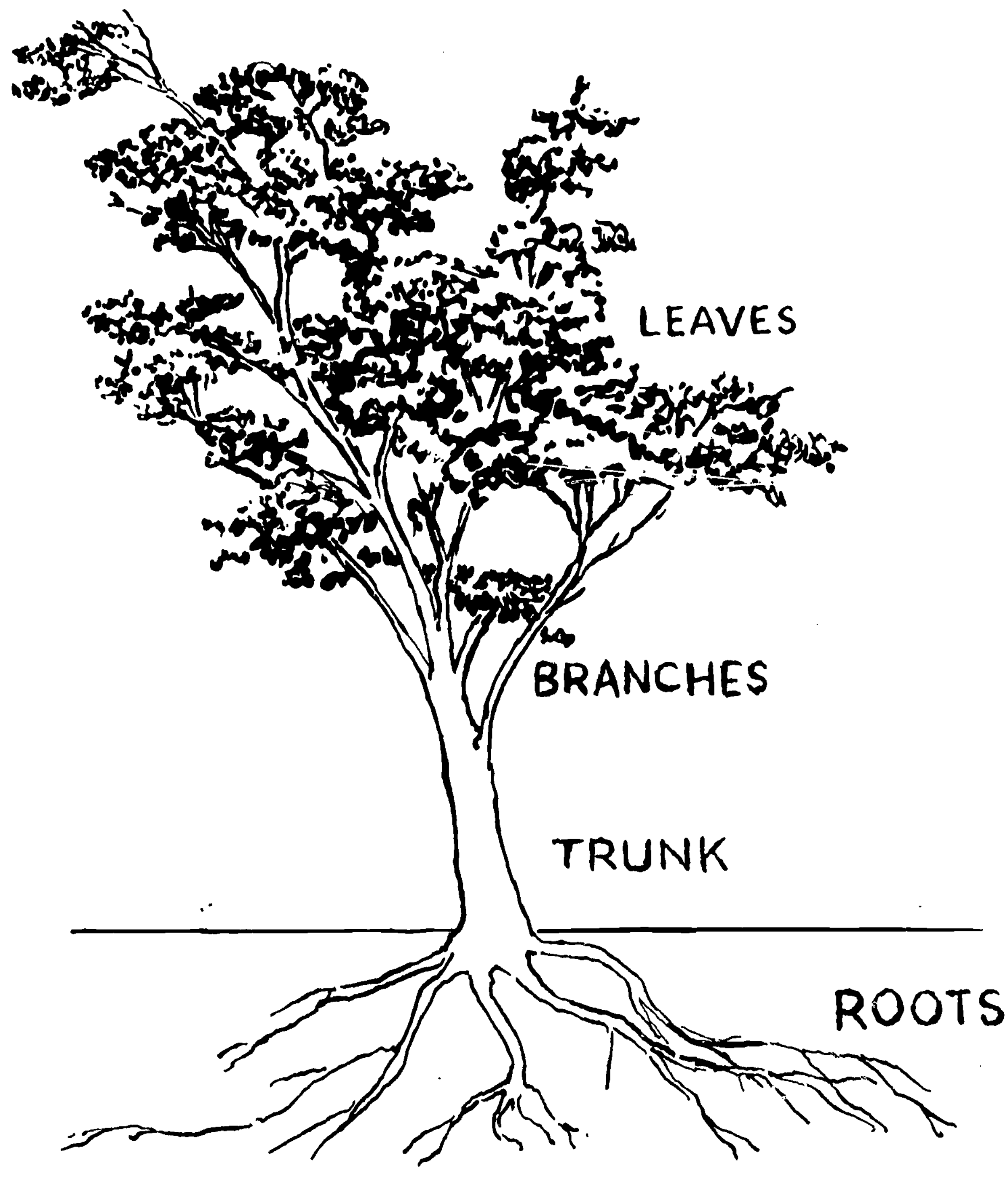
The following are examples of objective-test items which are based upon pictorial material:

Directions: Mark an X across the picture that shows a complete circuit.



This is an example of a multiple-choice item based on pictorial material. It has been successful in first and second grades with the teacher reading the directions orally.

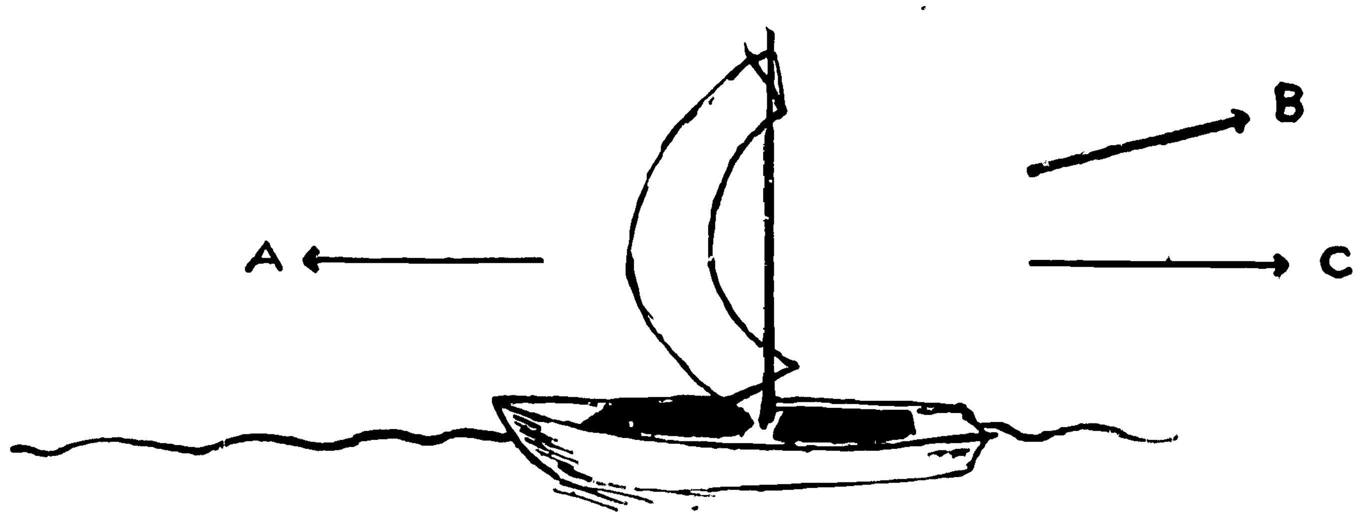
Directions: Here is a picture of a tree with names of some of its parts printed on it. Use these names to complete the statements about the tree.



1. The food for the tree is made in the_________.
2. The tree is held in the ground by the_________.
3. Water is taken into the tree through the_________.
4. The sap is carried to the branches through the_________.
5. The food for the tree is stored in the_________.

This is an example of a completion item using pictorial material. It has been used successfully in second and third grades. Where children have difficulty reading the statements, the teacher reads the statement orally and the children copy the appropriate word from the picture.

Directions: Look carefully at this picture of a sailboat being moved by the wind. There are two statements about the boat listed below. Mark an X across the word which makes each statement true.



1. This boat will move because the wind is pushing the _________.

SAIL WATER AIR

2. This boat will move toward_________.

A. B. C.

This is another example of a multiple-choice item using pictorial material.

3. Summary suggestions for all testing procedures

a. Check test items against the objectives of the unit to insure that the items relate to the goals and that the test items adequately cover all of the goals. One method of assuring a proper distribution of items is to use a simple coding system. Each objective can be numbered, and as each test item is constructed, it can be related to the objective by assigning it the appropriate number.

b. Check the reading level of each item so that students who are being evaluated in science are not penalized for deficiencies in reading speed or comprehension.

c. Arrange test items in order of difficulty. This provides for more efficient use of the testing time. When difficult items appear first, many students use up most of their time with a few questions and never have time to answer easier questions which may appear later. Furthermore, students who have difficulty at the beginning of a test may soon become discouraged and give up.

Group items also according to subject matter and to type of item, and always in order of increasing difficulty. This system has real merit because it reduces confusion and helps the student focus more efficiently on the task to be accomplished.

d. Word directions for pupils clearly, specifically, and without ambiguity. The student must know exactly the manner in which he is to select and record his response and also the amount of time allotted for the test.

e. Analyze and classify pupil's responses to tests. They should not just be scored and then forgotten, but can be used by the teacher to gain valuable information regarding student difficulties, teaching techniques, and the tesc itself.

Observation Used for Evaluation

One of the most useful techniques for evaluating the attainment of many objectives of the science curriculum is teacher observation of student behavior. For example, two objectives of a unit on microbes might be

(1) To help students learn simple laboratory techniques with which they can grow and prepare micro-organisms for observation and study, and

(2) To help studentslearn to use a microscope and a microprojector.

Progress toward these objectives can be evaluated best when student behavior is observed in the realistic laboratory situation, but the observation cannot be random or casual. In the first place, the teacher should know exactly what he is looking for. Some indications of a student's ability to use a microscope properly are that he

(1) Handles the instrument with great care.

(2) Cleans the lenses only with lens tissue or with a soft, clean cloth.

(3) Never focuses the microscope downward toward the slides; always moves the objective downward while the eye is away from the eyepiece, and then focuses the microscope upward with the eye looking through the microscope.

(4) Arranges the mirror for optimum amount of light.

(5) Prepares material for observation, using the techniques most appropriate to whatever is being examined; uses depression slides or bridge arrangements for comparatively large material; uses cover slips to cover smaller items.

1. Checklists and rating scales

Having determined the desired behaviors, the teacher can prepare a checklist or a rating scale containing a list of all the actions that relate to student behavior in a particular area. By using a check mark, the teacher can record his observations of a student's performance. The following is an example of such a checklist:

	Always	Sometimes	Never
Is careful in handling microscope			
Cleans lenses properly			
Focuses instrument properly			
Prepares slides correctly			
Arranges mirror for correct amount of light			

The data yielded by this checklist can indicate how well students have achieved the objective of using a microscope properly. Such information is usually qualitative and subjective, but when teachers know what they are to observe, the subjective generally can be made more objective. (Illustrations of a variety of checklists appear in the Appendix.)

In the same way, to check whether the children have learned the simple laboratory techniques with which they can grow and prepare micro-organisms, direct observation of the students' behavior in the situation will provide the teacher with more accurate evidence for evaluation than will responses by the children to written questions. Obviously, answers to written questions can give some evidence and should be used too. But what is wanted from the children is not so much the ability to verbalize about what they should do, as actions which show that they can conduct themselves in the desired manner.

2. Areas where observation is indicated

An analysis of many areas of the curriculum reveals long-term goals which depend mainly upon observation for their evaluation. For example, there are goals like assuming responsibility, sharing and communicating with others, practicing proper health habits, developing sound attitudes toward learning, or participating in classroom activities. Attaining such long-term goals is an essential part of the science program. Observation is the most effective way by which to determine how well they are being achieved.

To sum up: Observation as an evaluation technique can be effective if the teacher has a clear understanding of the behavior to be assessed. Furthermore, it is incumbent upon the teacher to provide equal opportunities for all the students to respond in the desired manner; every child must have his chance. This requires a conscious effort on the part of the teacher. Finally, a written record of such observations is not only desirable, it is imperative. How these records are made out - whether they be in anecdotal form, in rating scales, or on check sheets - is not too important. Any of these forms can serve the teacher's purposes. What is essential, however, is that the teacher make a written appraisal of what each child actually has shown himself able to perform in certain areas of behavior according to a set of criteria.

Appraising Children's Projects

Another very significant way of finding out how well children are meeting the objectives of the science program is the teacher's examination of the material which the children produce.

After all, what a child does and what he produces can tell much about the way he meets the objectives of the program. For example, we know that third and fourth-graders are collectors. But collections have little worth from a science point of view unless they are organized. One indication of a scientific attitude is the manner in which a person employs a theme for his ideas, and organizes his facts and information in a planned classification. A teacher of the third or fourth grade will want the children to begin to develop the ability to conceive of such themes and such frameworks, and to organize their collections and categorize them accordingly. Thus, the teacher examines a collection of rocks and looks to see how the rocks are grouped. Are they organized according to the place where the rocks are found? Are they grouped as igneous, sedimentary, and metamorphic? Are they exhibited to show certain interesting phenomena such as weathering, water erosion, or ice scratches? Or are they just a hodge-podge of pretty stones? Neatness, beauty, novelty are all important, but not for science. What is being evaluated as far as science is concerned is the ability to organize and classify materials in a sensible and reasoned way.

Then there are the experiments which children design and the models which they construct to illustrate applications of scientific principles. An analysis of such materials can reveal much more clearly the extent of a student's attainment of the goal of expanded understandingsof science than can any paper-and-pencil test. A careful study of such materials is one important way of determining growth toward extended vision and richer insight into the meanings of science and the applications of these meanings to appropriate situations. But such study requires that the teacher be sure about the objectives he is trying to reach. If this evaluation procedure is to have validity, the teacher must appraise a project on the basis of the processes used by the student in reaching his conclusion. Extended vision of one's environment, insight into the scientific concepts derived from facts, and understanding of how such concepts may be applied to specific problems cannot be measured by the quality of the art work involved in lettering the parts of an exhibit. Rather, it is measured by the clarity of thinking that shows in the resultant project. It is measured by the extent to which the exhibit explains a scientific principle through clear and simple examples.

Science reports, too, need this same kind of evaluation. It is not a matter of how many pictures are included in a report. Rather, it is the appropriateness of the pictures as illustrative of the points being made. It is not the length of the report, but the thoughtful organization and clear explanation of the material presented.

And, as far as writing goes, a teacher may very well refuse to accept a report from a child because it is not up to the level of neatness or standards of language skill of which he is capable. Misspelled words and poor grammar are not acceptable in science reports any more than they are in English reports. But having to return such a report for rewriting should have no bearing on the science evaluation. In evaluating a science report, the teacher should appraise its worth as science - its accuracy of information, its appropriate explanations, its resultant generalizations, its organization. Science reports must be judged in the light of science objectives.

The teacher must be certain that the objectives upon which the work will be built and upon which it should eventually be appraised are stated in such forms as to indicate the type of resultant behavior desired. If the objective of a weather unit is to have the children understand the water cycle, the exhibit or project which shows, simply and clearly, how water evaporates and then condenses is much more truly an example of sound science thinking than is an elaborate poster of the various kinds of clouds, beautiful as the art work may be. And a simple home-built model of the workings of a gasoline engine -- a model made from cardboard, paper fasteners, and crayons -- is a much more acceptable project than a plastic cross-sectional, commercial model of a complex Diesel engine, even though the Diesel engine is put together with great care. What is wanted is a demonstration of how children are thinking, of how well they understand the scientific principles which they are studying. The home-built model shows this; the purchased plastic model does not. Only as the teacher knows clearly the kind of behavior he eventually expects from his students, and as he helps his students carry out projects which lead to this kind of behavior, can he develop an adequate basis for evaluating the work which his students produce.

BIBLIOGRAPHY

Ahmann, J. Stanley and Glock, Marvin. Evaluating Pupil Growth. Boston: Allyn and Bacon, Inc., 1958.

Ahmann, J. Stanley; Glock, Marvin; and Wardeberg, Helen. Evaluating Elementary School Pupils. Boston: Allyn and Bacon, Inc., 1960.

Remmers, H. H., and Gage, N. L. Educational Measurement and Evaluation. New York: Harper and Bros., 1955.

Thomas, R. M. Judging Student Progress. New York: Longmans, Green and Co., 1954.

Thorndike, Robert and Hager, Elizabeth. Measurement and Evaluation in Psychology and Education. New York: John Wiley and Sons, Inc., 1961.

Torgerson, T., and Adams, G. Measurement and Evaluation. New York: The Dryden Press, 1954.

Travers, R. How to Make Achievement Tests. New York: The Odyssey Press, 1950.

APPENDIX

These checklists are included as illustrations of a quick method of recording observations. When compared with standardized evaluative devices, checklists fall short in reliability, validity, and objectivity. Where other instruments are more appropriate for the intended purpose, checklists should not be substituted.

A CHECKLIST FOR SELF-EVALUATION

Joseph E. Brzeinski
Coordinator of Elementary Instruction
Denver Public Schools, Colorado

Some teachers have been able to improve their instruction by evaluating their teaching in terms of plans for instructional improvement closely paralleling the generally accepted scientific method. The six steps of the Denver Program of Instruction form the basis for such a checklist to assist teachers in evaluating their own instruction. While it is recognized all elements may not be present in any one science experience, analysis of answer patterns may assist you in evaluating your teaching of a science unit.

I Purposing--This has to do with the objectives of the unit.

A. Are you going to teach important principles of science in this unit?

B. Are you going to develop scientific methods and attitudes?

C. Are the things you are going to teach in keeping with life in a democracy?

D. Do the things you are going to teach really matter to your pupils--now or as adults?

II Surveying--This step involves determination of what may be expected of individuals. Necessary information is secured that permits better understanding of each pupil's experiential and developmental level.

A. Is consideration given to indications of ability provided by I.Q. tests?

B. Is use made of information from standardized tests?

C. Are cumulative records and previous teachers consulted for supplementary information?

D. Do you use what we know about how children grow and learn in interpreting the above information?

Source: National Science Teachers Association. Elementary School Science Bulletin. No. 40, October 1958. Washington, D.C.: the Association, a department of the National Education Assoc.

III-IV Planning-organizing--These steps have been combined in view of their interrelatedness.

A. Is your planning such that sufficient time is provided for science experiences?

B. Will a wide variety of suitable books, science materials, and audio-visual aids be available?

C. Is provision made for securing resource people, taking appropriate field trips, and otherwise using community resources?

D. Do pupils participate and develop facility in collecting information, using reference sources, making decisions, and devising appropriate procedures.

V Operating--This step involves putting the plans into action.

A. Are the questions asked answered adequately by the activities carried on or by the sources consulted?

B. Are all members of the class involved in working together, in "trying out," observing, recording, and reporting?

C. Wherever possible do you use a "control" factor in your experiments?

VI Evaluating--Progress in learning is determined through evaluative techniques.

A. Was there growth of science vocabulary?

B. Was understanding of important principles of science demonstrated in class discussion and experimentation?

C. Was the quality of written and oral reports, demonstrations, experiments, and projects good?

D. Was there growth in ability to observe accurately, and increased resourcefulness in research, reference work, and problem solving?

If you can answer "yes" to all of these questions you are a teacher who is "too good to be true." If you can answer "yes" to a majority you are probably doing a pretty good job of teaching science, but there is room for improvement, too.

TEACHER'S CHECKLIST IN ELEMENTARY SCIENCE

Paul Blackwood
Specialist for Elementary Science
U.S. Office of Education

Purpose: To measure the effectiveness of school practices and procedures in teaching science in the elementary school.

I. In my teaching is there opportunity or provision for children to:

	None	Some	Much
(a) Raise questions and problems of importance or interest to them?	____	____	____
(b) Study these questions and problems?	____	____	____
(c) Help plan "things to do" in studying science problems?	____	____	____
(d) State clearly the problems on which they are working?	____	____	____
(e) Make hypotheses to be tested?	____	____	____
(f) Gather accurate data (information) in a variety of ways:			
Through reading on the subject?	____	____	____
" taking field trips?	____	____	____
" watching demonstrations?	____	____	____
" doing experiments?	____	____	____
" talking to resource persons?	____	____	____
(g) Analyze the data (information) to see how it relates to the problem?	____	____	____
(h) Think about the applications of their science learnings to everyday living?	____	____	____

Source: "Evaluating Teaching Practices in Elementary Science," Education Briefs, No. 21, July 1959. (OE-20009-21) Washington, D. C.: Department of Health, Education, and Welfare, Office of Education.

	None	Some	Much
(i) Think about science relationships and processes instead of merely naming things and learning isolated facts?	___	___	___
(j) Bring science materials of different kinds to school for observation and study?	___	___	___
(k) Engage in individual science interests?	___	___	___

II. In my teaching do I periodically and systematically check on the children's growth in:

	None	Some	Much
(a) Ability to locate and define problems right around them?	___	___	___
(b) Acquiring information on the problem being studied?	___	___	___
(c) Ability to observe more accurately?	___	___	___
(d) Ability to make reports on or record their observations?	___	___	___
(e) Ability to solve problems?	___	___	___
(f) Ability to think critically?	___	___	___
(g) Ability to explain natural phenomena?	___	___	___
(h) Ability to distinguish between facts and fancies?	___	___	___
(i) Suspending judgment until evidence is collected?	___	___	___
(j) Being open-minded, or willing to change belief?	___	___	___
(k) Cooperating with others ?	___	___	___

	None	Some	Much
(l) Understanding the cause and effect relationships of events?	____	____	____
(m) Skill in using more common scientific instruments (thermometers, scales, rulers, etc.)?	____	____	____

PLANNING FOR SCIENCE TEACHING IN THE ELEMENTARY SCHOOLS

Harry Milgrom
Assistant Director of Science
New York City Schools

Effective preparation for science instruction at all levels, especially in grades K-6, requires that teachers and supervisors know how to draw up appropriate plans and put them into practice. For each lesson, the plan should pinpoint the specifics that the teacher aims to use to:

1. Train children in scientific methods of investigation. They should learn how to--

   1.1 Identify problems

   1.2 Offer possible solutions

   1.3 Suggest procedures for verification

   1.4 Make observations

   1.5 Make measurements

   1.6 Keep records

   1.7 Weigh evidence

   1.8 Reach tentative conclusions

   1.9 Check conclusions

2. Give children opportunities to develop--

   2.1 Reasoning ability

   2.2 Scientific skills and techniques

   2.3 Scientific attitudes and appreciations

   2.4 Understanding of basic science principles and concepts

   2.5 A fund of scientific knowledge

Source: Board of Education of the City of New York, Office of Science Education, Brooklyn, New York. (Typewritten)

3. Provide suitable activities, experiences, experiments, and projects.

4. Insure proper utilization of--

   4.1 Science materials and equipment (storage, distribution and return)

   4.2 Books

   4.3 Audio-visual aids

   4.4 Special science activities (fairs)

   4.5 Neighborhood resources

5. Prepare pivotal, thought-provoking questions.

6. Insure proper allotment of time.

7. Increase children's science vocabulary.

8. Evaluate children's learning and progress toward science teaching objectives.

SCIENCE IN-SERVICE COURSES FOR ELEMENTARY SCHOOL TEACHERS

Harry Milgrom
Assistant Director of Science
New York City Schools

The quality of science instruction at any level is determined largely by the capability of the classroom teacher. In-service training has been one approach to the problem of increasing teaching competence. For greater effectiveness, what criteria should be set up to guide the design of an in-service training course in science for elementary school teachers?

Here are a number of questions that may be useful in focusing attention on some guiding principles.

1. Is the training course based upon a planned science curriculum?

2. Is the course set up to give teachers confidence in their ability to teach science?

3. Does the course help the teachers become aware of the spirit of science and the purposes of science instruction?

4. Does the course pave the way to understanding of basic science principles?

5. Does the course show teachers how to activate children and get them to think?

6. Does the course help the teachers become familiar with science teaching materials and aids?

7. Does the course provide a background of information and knowledge without overwhelming and frustrating the teachers?

8. Does the course provide guidance in the allotment of time for science teaching?

9. Does the course feature a program of first-hand experiences and experiments for children?

10. Is the course based on lectures or workshop activities? In other words, are the teachers taught in the same ways that they are expected to teach?

Source: Board of Education of the City of New York, Office of Science Education, Brooklyn, New York. (Typewritten)

SCIENCE BOOKS

33 Keys to Evaluation

Harry Milgrom
Assistant Director of Science
New York City Schools

Does the book portray science as an exciting adventure of the human mind?

Is the book provocative and challenging? Does it inspire more questions than it answers?

Does the book further understanding of the basic principles of science?

Does the book give insight into the methods of science?

Does the book stimulate new interest and stir new wonder?

Does the book promote independent inquiry and thought?

Does the book provide leads for experiments?

Is the book based on some foundation of previous knowledge and experience?

Does the book reveal the role of the imagination in the progress of science?

Is the content of the book within the comprehension of children in the age group for which it is written?

Is the style of writing appropriate to the subject, clear, interesting and meaningful to the reader?

Has the book been checked for accuracy of scientific vocabulary and expression?

Are the illustrations in the book clear, meaningful to the text and scientifically correct?

Does the author give careful attention to details?

Is the book more than a "cut and dried" collection of assorted facts?

Source: School Library Journal, Vol. 8, p. 21-22, February 1962.

Does the book introduce the reader to such tools of scientific exploration as observation, measurement, reasoning from data?

Does the book increase in some measure, children's awareness and understanding of the science in the world around them?

Does the book help the reader gain greater insight into the "chains of events" in the history of science?

Does the book try to involve the reader in action and thought or does it relegate him to the role of a bystander?

Does the book provide ideas for the reader to ponder and mull over?

Does the book distinguish fact from fancy?

Does the book try to show what motivates scientists and makes them tick?

Does the book portray science as an endless quest or does it close the door to further exploration by describing "all there is to know"?

Does the book show how scientists establish order out of apparent chaos by bringing to light the big, unifying ideas?

Is the emphasis of the book on learning by rote or on learning by reason?

Does the author show that he understands the needs and aspirations of children?

Does the author respect the dignity and intelligence of the young reader?

Does the author try to cover everything in a science area or does he deal only with selected aspects of the area for greater depth of treatment?

Does the book merely enumerate discoveries of the past or does it illuminate the approaches and probings that lead to discovery?

Does the book present a great deal of information without indicating how scientists go about unearthing information?

Is the book written to conform to a preset pattern for a series? Does it suffer thereby?

Does the book contain any original or unusual features that set it apart from other books about the same subject?

Does the book introduce the reader to the nature and spirit of The Scientific Enterprise?

TWENTY POINT CHECKLIST FOR A SCIENCE ROOM

Harry Milgrom
Assistant Director of Science
New York City Schools

Editor's note: Here an elementary science supervisor gives some help on evaluating a science room. Although many schools do not have separate rooms set aside for science teaching, this checklist will be helpful in checking the science part of any self-contained classroom. Many of the items suggest activities that are appropriate in any room where science learning takes place. For example: Have you considered having a live animal zoo (item 6)? or a science reference shelf (item 14) or a collection of science pictures (item 15)?

1. Is the science room a high spot of the school?

2. Is a fifth or sixth grade class assigned primary responsibility in: (a) planning, (b) arranging, (c) running the science room?

3. Do the children enter the room faster than they leave? Is the room a merry one with eye and nose appeal?

4. Does it have a "touch and try" section where the children can study:

plants	water	machines	electricity
animals	soil	sound	building
foods	rocks	heat	and clothing
air	toys	light	materials
weather	chemicals		

5. Does it have a work table, tools, and construction material such as

wire	spools	hair pins
string	cardboard	cement
rubber bands	empty containers	scotch tape
paper fasteners	candy boxes	
tacks	round pencils	

for the assembling of simple projects?

Source: National Science Teachers Association. Elementary School Science Bulletin. No. 9, February 1954. Washington, D.C. the Association, a department of the National Education Association.

6. Does it have a "live animal zoo?"

   a. Housed in natural habitat as far as is possible?
   b. Fed, watered, and cleaned at proper intervals by the children?
   c. Simply labelled in large type?
   d. Characteristics described by stories and drawings created by the children?
   e. Provided for on week ends and holidays?
   f. Protected against mishandling by acquainting the children with the proper handling of animals?

7. Does it have its own "botanic garden" featuring:

   a. Live displays of plants and flowers in season?
   b. Woodland scenes of ferns, mosses, and lichens?
   c. Plants grown at home by the children (sweet potato, beans, grapefruit, carrot, etc.)?

8. Does it have a museum?

   a. Colorfully displayed?
   b. Carefully spaced?
   c. Clearly labelled?

   Does the museum feature:

   a. Diorama scenes made by the children?
   b. Dioramas borrowed from a local museum?
   c. Models constructed on a table top showing farm, jungle, Eskimo, factory, or similar scenes?

9. Does it have a bulletin board? Is it:

   a. Changed periodically as seasons, topics or interests vary?
   b. Decorated with the children's own science art work?
   c. Used to display children's questions and answers?

10. Does it have a quiz section consisting of

    a. Electrical games?
    b. Identifications?
    c. Riddles?
    d. Picture puzzles?

11. Does the room have an imaginative play corner, outfitted with homemade devices simulating a:

    a. Fire engine?
    b. Derrick?

c. Space port?
d. Locomotive cab?
e. Radio or television studio?
f. Research laboratory?
g. Mountain observatory?
h. Weather station?

12. Is the school science equipment stored in the room? Is it:

a. Provocatively displayed?
b. Readily accessible to the teachers?
c. Catalogued?

13. Are the displays integrated with class activities of:

a. Seasonal
b. Topical

interests?

14. Does it have a reference shelf, indexed according to topics to facilitate "looking up" by children?

15. Does it have a collection of science pictures, indexed according to subject matter?

16. Does it feature a toy corner where the science implications of:

a. Magnetic toys (Scottie dogs on magnets)
b. Wind-up toys (cars, locomotives)
c. Boats and planes (submarines, jet balloon), may be studied?

17. Does it publish a science bulletin written by the children to describe their experiments and findings?

18. Are safety precautions prominently featured and stressed with respect to:

a. Use of tools?
b. Handling of animals?
c. Use of chemicals and heat?

19. Do the children at work in the room:

a. Show genuine interest?
b. Set up their own problems?
c. Become more and more curious?

20. Is the teacher assigned to supervise the activities, eager to explore the world of science in full partnership with the children?

INSERVICE TELEVISION-WORKSHOP COURSE IN SCIENCE

Checklist for Instructors

Harry Milgrom
Assistant Director of Science
New York City Schools

The instructor's role in the conduct of the Television-Workshop program is of very great importance. It is suggested that instructors use this checklist as a guide to help make sure that participating teachers get the most out of their work in the course.

I. Physical Arrangements

1. Is the lighting in the room at the proper level for viewing? for taking notes?

2. Are the teachers seated so that they do not get window glare? light reflections from the screen?

3. Is the picture clear and visible to everyone?

4. Is the sound clear?

5. Is the room free of distracting inside or outside noise?

6. Is the room properly ventilated?

7. Is an alternate TV receiver available in case the first one breaks down?

II. Materials

1. Does each teacher bring in his own science handbook to use as a reference?

2. Are there enough materials on hand so that each pair of teachers can engage in the suggested activities?

3. Do you arrange to have the materials for the following week brought in from these sources:

Source: Board of Education of the City of New York, Office of Science Education, Brooklyn 1, New York. (Typewritten)

a. The workshop school
b. Other participating schools
c. Teachers' homes
d. Junior and senior high schools, if necessary

III. Instructor's Attitude

1. Do you use the 90 minutes of workshop time to give the teachers the maximum number of firsthand experiences in the science area of the telecast?

2. Do you keep the amount of talking you do down to a minimum and place the emphasis on the workshop nature of the course?

3. Do you avoid becoming involved in theoretical discussions that are too advanced for the majority of the teachers?

4. Do you try to maintain the work of your group at the level set by the telecast?

5. Do you solve the problems for the teachers or do you permit them to work out their own solutions as far as possible?

6. Do you adapt the activities and discussions to meet the needs of the group? the individual teachers?

7. Do you convey to the teachers your own feeling of the importance of the course?

8. Do you try to answer questions that you cannot answer or do you guide the teachers to sources for these answers?

9. Do you help the teachers develop their own ingenuity and resourcefulness?

IV. Miscellaneous

1. Have you made provision for an alternate instructor to take over in the event of your absence?

2. Do you check attendance at each session?

3. For the workshop, do the teachers work in pairs, or at most in threes?

TEACHER UTILIZATION OF TELEVISION BROADCASTS IN SCIENCE

Harry Milgrom
Assistant Director of Science
New York City Schools

I. Preparation of Physical Facilities

1. T.V. receiver

a. Is the screen size adequate?
b. Is the screen located at a height above eye level of the seated pupils?
c. Is the top of the screen tipped forward slightly?
d. Is the T.V. cord in a position where pupils cannot trip over it?

2. Arrangement of seats

a. Are the children seated so that they do not get glare from bright parts of the windows?
b. Are the children seated so that they do not get reflected light from the T.V. screen?
c. Are the children seated at the best distance from the T.V. screen?
d. Are the children seated so that each can see the entire screen?

3. Lights

a. Is the lighting in the room at the best level for viewing?
b. Is the lighting in the room adequate for the pupils to take notes, if necessary?

4. Seats

a. Are the seats at the right height for comfortable viewing?
b. Do the seats have arm rests for writing?

5. Ventilation

Is the room properly ventilated?

Source: Board of Education of the City of New York, Office of Science Education, Brooklyn 1, New York. (Mimeographed)

6. Outside noise

   a. Is the room located in an area where outside noise interferes with the T.V. sound?
   b. Is the room located in an area where the sound of the T.V. interferes with other school work?

7. Reception

   a. Is the picture at the right level of brightness?
   b. Is the screen clean, that is free of fingerprints and dust?
   c. Is the contrast between light and dark parts of the picture good?
   d. Is the picture clear of streaks?
   e. Is the picture clear of echoes?
   f. Is the picture clear of distortion?
   g. Is the picture steady, that is clear of flickers?
   h. Is the sound at the best level for all children to be able to hear it?
   i. Is the sound free of noise?
   j. Is the sound clear and free of distortion?

II. Pedagogical Preparation for Viewing

1. The Teacher

   a. Did the teacher motivate the children for the forthcoming telecast?
   b. Did the teacher follow the pre-telecast suggestions given in the T.V. manual?
   c. Did the teacher prepare the children for viewing by arousing their curiosities and interests?
   d. Did the teacher cause children to lose interest by anticipating and revealing too much of the telecast?
   e. Did the teacher make sure that the children had the background knowledge to enable them to understand the telecast?
   f. Did the teacher make sure that the children know the meaning of words that might be used during the telecast?
   g. Did the teacher follow these suggestions perfunctorily?
   h. Did the teacher use her own initiative in the pre-telecast activities? If so, what specific approaches did she use?

____

____

____

2. The Children

a. Did the children participate in the pre-telecast activities?
b. Did the children have the opportunity to ask questions about the forthcoming telecast?
c. Did the children have the opportunity to do experiments related to the forthcoming telecast?
d. Did the children have the chance to review previous work or telecasts related to the forthcoming telecast?
e. Did the children seem to be genuinely interested in the forthcoming telecast?

III. Activity During the Telecast

1. The Teacher

a. Did the teacher watch the telecast?
b. Did the teacher watch the children's reactions to the telecast?
c. Did the teacher take notes during the telecast on which to base the follow-up work?
d. Did the teacher maintain the best conditions for viewing during the telecast?
e. Did the teacher direct children's observations during the telecast? If so, did she deprive the pupils of the opportunity to see for themselves?
f. Did the teacher give the answers to questions that were asked by the T.V. teacher?
g. Did the teacher encourage the children to answer the T.V. questions by themselves?
h. Did the teacher intrude on the T.V. presentations? If so, how was this done?

____

____

____

____

i. Did the teacher permit distractions to interfere with proper attention to the telecast?
j. Did the teacher relate the telecast to class work ?

2. The Children

a. Was the interest of the children maintained during the telecast?
b. Were the children expected to take notes?
c. Were the children expected to jot down questions that the telecast brought to their minds?
d. Were the children relaxed and comfortable?
e. Were the children fidgety?
f. Did the children enjoy the telecast?
g. Were the children sorry when the telecast ended?

IV. Activity After the Telecast

1. The Teacher

a. Did the teacher use the follow-up suggestions given in the T.V. manual?
b. Did the teacher use her own initiative in developing follow-up activities? If so, what did she do?

____

____

____

c. Did the teacher follow-up the telecast in a perfunctory manner?
d. Did the teacher give the children an assignment for the next day?
e. Was the follow-up activity integrated with the regular class work?
f. Was the follow-up activity assigned without regard to the regular class work?
g. Did the teacher indicate that the follow-up activity would be a long range process?
h. Did the teacher try to channel children's interest in the telecast into creative projects such as the science fair, for example? If so, list the projects.

____

____

____

i. Did the teacher use the telecast to enrich the children's understanding of scientific principles?
j. Did the teacher use the telecast to enrich the children's vocabulary?
k. Did the teacher use the telecast to enrich the regular science work for the class?
l. Was the teacher's follow-up satisfactory?

2. The Children

a. Were the children able to suggest ideas for follow-up activities?
b. Were the children eager to engage in follow-up activities?
c. Did the children engage in the activities that were suggested in the telecast?
d. Did the children repeat the experiments that were demonstrated?
e. Did the children devise their own experiments, in extending the ones they saw on T.V.?
f. Were the children encouraged to observe carefully?
g. Were the children encouraged to think for themselves?
h. Were the children led to jump to conclusions?
i. Did the children have fun, without learning science?
j. Were the children encouraged to visit places of scientific interest?
k. Did the children indicate that they wanted to read in order to learn more about the theme of the telecast?
l. Did the children want to ask questions during the telecast?
m. Did the children seem to grow more curious as a result of the telecast?

U.S. GOVERNMENT PRINTING OFFICE : 1964 O - 748-546