

Eastern Illinois University

The Keep

Plan B Papers

Student Theses & Publications

2-11-1959

Criteria for the Construction of Evaluation Instruments to be Used in Industrial Arts Classes

G. E. Maloney

Follow this and additional works at: https://thekeep.eiu.edu/plan_b

Recommended Citation

Maloney, G. E., "Criteria for the Construction of Evaluation Instruments to be Used in Industrial Arts Classes" (1959). *Plan B Papers*. 46.
https://thekeep.eiu.edu/plan_b/46

This Dissertation/Thesis is brought to you for free and open access by the Student Theses & Publications at The Keep. It has been accepted for inclusion in Plan B Papers by an authorized administrator of The Keep. For more information, please contact tabruns@eiu.edu.

CRITERIA
FOR THE CONSTRUCTION OF
EVALUATION INSTRUMENTS TO BE
USED IN INDUSTRIAL ARTS CLASSES

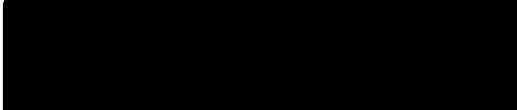
by

G. H. Maloney

Submitted Under Plan B In Partial Fulfillment
Of The Requirements For The Degree, Master of
Science in Education.

Approved:

Feb. 11, 1959
Date


Dr. R. H. Landis
Instructor I. A. 560

Date

Walter S. Klehm
Dr. Walter Klehm
Advisor

Eastern Illinois University
Charleston, Illinois

Fall, 1958

CONTENTS

Importance of use of evaluation instruments.....	1
Teacher objectives.....	1
Pupil objectives.....	4
Construction of adequate tests requires tests of certain criteria.....	5
Validity.....	6
Reliability.....	10
Objectivity.....	13
Discrimination.....	16
Comprehensiveness.....	18
Administration and Scoring.....	19
Areas of evaluation.....	20
Summary.....	25
Selected References.....	27

CRITERIA FOR THE CONSTRUCTION OF EVALUATION
INSTRUMENTS TO BE USED IN
INDUSTRIAL ARTS CLASSES

Importance of use of evaluation instruments

A necessary basis for the intelligent discussion of the criteria for the construction of evaluation instruments to be used in industrial arts classes must be the appreciation of the importance of that which they measure, and this, in turn, can only have meaning when based upon the understanding of the varied uses to which these findings are put.

The main reason why evaluative instruments are designed and used is to determine the extent to which the learners have achieved the objectives which have been set up for some particular course, or some basic part of that course.

This evidence of achievement, or lack of it, may be used in a number of ways. The teacher may use it as a test of the efficiency of his own teaching and as a basis for re-teaching, and modifying certain curricular materials, or he may use it as an incentive toward future learning.

Teacher objectives

It is important for a teacher to know how successful he has been in presenting his material, so that he can improve his instruction at the points of weakness, both as areas of re-teaching with the group just tested and as improved teaching with another. This need for remedial effort

can only be determined from evaluative procedures which make clear whether there has been any definite lack of success in the teaching and learning which have taken place up to this time; the findings of which procedures serve as a basis for the necessary re-teaching and modification of present practices. Without such appraisal the teacher cannot know how successful either his teaching or the learning of his pupils has been; yet if the teaching and learning are worth undertaking in the first place, it is desirable to measure some degree of their success.

As Leighbody says, "Every test is as much a measure of the teacher's skill as of the accomplishment of the learner. Instruction can only be improved if the points at which it has failed are clearly recognized by the teacher. The teaching procedure can then be revised so that when the lesson is again presented the weak points will be strengthened and learning made easier. If a test shows that a considerable number of learners are weak in one or more elements of a particular lesson, after proceeding through the lesson as given, it is clear that the teacher needs to examine carefully the methods used to provide a practice in their use. Undoubtedly, some revision is called for, or some other method to supplement those previously used."¹

The teacher cannot satisfactorily engineer these changes

1. G. B. Leighbody, Methods of Teaching Industrial Subjects, Delmar Publishers, Inc., Albany, N.Y., 1946, p. 117.

in his method of presentation unless he has definite knowledge from time to time of the accomplishments of the learners in the acquisition of the desired growth. In other words, he must have recourse to the results of evaluation.

Another use of the results of evaluation is as a basis for rating pupils; for determining the marks or grades which schools demand must be given at certain definite times. Because these grades have certain prestige values in that they seem to be indicative of the pupil's ability, and because they are used as a guide for determining his advancement, they should have a sound basis. This places a serious responsibility upon the teacher. He needs to be helped in his evaluation by devices which are as accurate as possible.

Since industrial arts has such a broad scope of educational objectives which include so many areas of manipulative skill, related subject matter, and types of desirable behavior changes, the teacher must be cognizant of all of these in his evaluation, and must be sure that evaluation is as all-inclusive as the course-objectives have been.

Leighbody develops this point in the following way. "Tests which are well planned and carefully constructed enable an instructor to accurately rate a learner when rating is called for and assure both instructor and learner that the appraisal of the work of the learner has been as fair and impartial as it is possible to be. So long as schools operate under standards which require that marks be assigned

at certain intervals and grades at the conclusion of courses, and so long as importance is attached to these grades for purposes of employment and for other recommendations, the rating of the learner will remain a serious duty of the teacher. If such ratings are to be fair and unbiased and are to represent to others a measure of the learner's accomplishment, which will often be interpreted as a measure of his capacity, the measuring must be as accurate as possible. Tests of the right kind are essential for this purpose."²

Pupil objectives

From the pupil's standpoint, aside from his interest in the grades and ratings which the teacher may make of his accomplishments, evaluative instruments may have a positive contribution in that they may motivate him to devote himself more thoroughly to his work, thus providing an incentive to him to improve his understanding and skill. For an evaluative instrument to be really worth while, it must be constructed and administered by a teacher who is constantly aware of this value of the testing device. The pupil needs to be helped to identify those areas in which he is weak and those areas in which he is strong, so that he can carry on his own work intelligently.

If an evaluative instrument fulfills this basic function by indicating the learner's progress in reaching his goals; it can then serve to motivate him since it discloses his

2. Ibid., p. 118.

3

success or failure in achieving these same objectives, and can then help to point the way to increased efficiency in learning.

These objectives are those which have been developed as part of the syllabus or course of study, and which are recognized as such by pupils and teacher so that both know what they are attempting to accomplish, and can make an appraisal of the outcomes of the learning in terms of this joint purpose. This serves, too, as a basis for the main use of the products of evaluation: increasing the efficiency of teaching and learning and applying remedial efforts, and as a basis for obtaining grades.

This introductory discussion of the importance of evaluation and the uses to which its findings can be put has been given to point up the importance of the construction of evaluative instruments.

Construction of adequate tests requires tests constructed according to certain criteria

In order to construct evaluative instruments which will be efficient in measuring this pupil course-objective achievement in such a way that the results can be used effectively, there are certain criteria which must be kept constantly in mind. As Leighbody says, "A few tests must be applied to the tests themselves."³ These criteria include validity, reliability, objectivity, discrimination, administration and scoring, each of which will be discussed in turn.

3. Ibid., p. 119.

Validity

The most important of these criteria is validity. The validity of an evaluative device is the degree to which it measures that which it is designed to measure: to test what it is supposed to test. Most of the authors whose books I have read agree on the essential definition of validity: some define it tersely as Ross does when he says "Validity means truthfulness."⁴; others add a specific quality or a further interpretation which clarifies or extends the definition. For instance, Micheels and Karnes describe as valid, a test "that measures well what it is supposed to measure," and further define their definition by stating that "a single test item is valid when it does the job expected of it."⁵

Remmers and Gage state that "Validity is the degree to which an evaluative device measures what it purports to measure."⁶ To them, this includes two aspects which they term as (1) the specific nature of validity and (2) the quantitative nature of validity, which they distinguish in the following way. "Validity is a specific concept in that

⁴ C. C. Ross, Measurement in Today's Schools, Prentice-Hall, Inc., New York, N. Y., 1947, p. 65.

⁵ William J. Micheels and M. Ray Karnes, Measuring Educational Achievement, McGraw-Hill Book Co., Inc., New York, N. Y., 1950, p. 104.

⁶ H. H. Remmers and N. L. Gage, Educational Measurement and Evaluation, Harper, and Brothers, New York, N. Y., 1943, p. 195.

it must always refer to a specific purpose or objective and a specific group of pupils. Given a specific group of pupils, an evaluation device or test will be valid for one purpose but not for others.

The quantitative nature of the concept of validity may be inferred from the method by which validity is frequently measured. The purpose of a test or other evaluating device is realized to the degree in which the results of the test correlate with data concerning the objective obtained by some method or criterion whose validity is already known."⁷ In other words, a test whose results show marked variation from the usual achievement of the group is suspect in so far as validity is concerned.

There are many factors which may be involved which make this marked variation possible. The reason or reasons why a test may not measure what the teacher intends to evaluate may be any one or more of the following. It may test the child's ability to read, or write, or spell, rather than his ability in the area which is being evaluated. This testing, especially of reading ability, is probably the most frequent misuse of a written test.

This was brought home to me very forcibly when I gave my first test to a group of pupils in a machine-woodworking class. (It was my first year in the school system, so it was

7. Ibid., p. 195.

the first written test that I had ever given to these boys.) One boy whom I felt had mastered the manipulative skills better than any of the others failed the test, in fact there was practically nothing on his test paper. This really bothered me as I knew he should have answered the questions without hesitation due to his skillful use of tools. When I mentioned this to one of the other teachers I was told that this boy could not read, so I gave him the chance to answer my verbal questions orally and he knew the answers quite well. The first test tested his reading, writing and spelling ability; which was not the objective of the test. The second tested his ability in the identification and use of tools, which was what I had hoped to measure all along.

This is an extreme case, but dramatically illustrated the point; at least to me. Many children have reading and spelling disabilities of a lesser degree, but such that the expansion of energy is upon these areas rather than upon the evaluation desired. When teachers mark answers wrong, if the spelling is incorrect, the test measures spelling. (Some teachers give two or more grades, one for what the course is supposed to be about, and the other for the spelling, or sentence construction, etc., of the mechanics involved.)

Another feature so closely allied that it might well be included is the language difficulty. This may be of several kinds or degrees. The obvious one is of native language differences; Mexican, Hungarian, etc., but more likely, at least

in schools in this area, to be one of being unable to follow intelligently the involved construction prevalent in so many of our tests. Comprehension and I. Q.'s. may be more nearly tested than the information the teacher desired.

Tests may actually measure endurance when too much area is being covered minutely; or speed, when a long or difficult test must be finished in a certain time. It may test consistency or neatness, or how well a pupil understands the teacher in being able to answer in terms of the teacher's preconceived ideas. There are others, I'm sure.

In an evaluation of manipulative skills, Leighbody has listed the following examples of lack of validity of tests because they tested areas or things other than what the teacher intended to measure.

"a) Tests which are designed to test quality of workmanship may actually measure speed.

b) Tests which are intended to measure quality or speed may really measure ability to analyze and plan the work.

c) Tests which are believed to measure understanding may only measure ability to memorize and recall.

d) Tests which are given to measure skill and efficiency in manipulation of machines and tools may instead measure the quality of the finished product.

e) Tests which are supposed to measure proficiency in hand or machine skills may really measure ability to read tech-

nical symbols."⁸

Since the basic reason for testing at all is to determine how nearly a pupil has reached the objectives of the course, or a part of it, and since this, in turn, is the teacher's basis for his reteaching, the validity of this evaluation assumes great importance. Leighbody develops the point in this way. "Unless a test is carefully analyzed before it is used, to determine that it actually measures what the teacher believes it will measure, the results may be interpreted to mean something which may be quite in error. If a teacher judges his success in teaching certain tool processes by the results of a test which actually measures the ability to read drawings, he may reteach processes which have been adequately mastered and fail to correct deficiencies in the ability to interpret drawings."⁹

Reliability

The second criterion which must be considered in the construction of evaluating instruments, is second only to validity in importance. By the term reliability we mean the accuracy, consistency and dependability with which a test measures what it measures; whether or not it is consistent in testing the same thing each time. Remmers and Gage define reliability essentially as the definition

8. Leighbody, op. cit., p. 120

9. Loc. cit.

given above, but go on to expand the idea by saying, "What a test measures may not be what it purports to measure; but if it measures something accurately, then it is a reliable test."¹⁰

This means that a test can measure certain factors and get approximately the same results each time but still not be valid in that it may not test what the teacher thinks it is testing. On the other hand, any test that is valid has to be reliable because it can not test what it is supposed to test, unless it is consistent in this testing.

Newkirk and Green declare that the reliability of a test may be thought of as the "consistency with which it performs."¹¹

Micheels and Karnes develop their definition in this way. "A reliable test is a 'trustworthy' test. It is accurate. It is consistent. If the test measures in exactly the same manner each time it is administered, if the factors that affect the test scores affect them to the same extent every time the test is given, the test is said to be high in reliability."¹²

Ross states essentially the same idea when he says, "By

10. Remmers and Gage, op. cit., p. 201.

11. Louis V. Newkirk and Harry A. Green, Tests and Measurements in Industrial Education, John Wiley and Sons, Inc., New York, N. Y., 1935, p. 34.

12. Micheels and Karnes, op. cit., p. 111.

reliability is meant the degree to which the test agrees with itself. To what extent can two or more forms of the test be relied upon to give the same results, or the same test to give the same results when repeated? If the scores on the test are stable under these conditions, the test is said to be reliable. In a word, reliability means consistency."¹³

For the teacher who is striving for the accomplishment of the construction of reliable tests this means that, after he has used the test a number of times, he must check the results to be sure that the test has measured in the same manner each time it was given. Was it accurate and consistent? He may analyze this problem by jotting down the objectives of what he expects the test to measure, after arranging the scores of the test previously given in a manner that will provide him with the information for checking how accurate and how consistent the test actually is.

A group of students, having received the same instruction, and given the same test, should produce approximately the same results. If there is a wide variation between a number of learners, when actually the differences do not exist, then the teacher cannot rely upon the test to measure accurately and consistently. A common cause for the failure of a test to measure consistently is that too few items are included in the test to give a true sampling of the learner's skill or knowl-

¹³. Ross, op. cit., p. 82.

edge. Children who do well in one area may rate high on the course test if the majority of the items is taken from his particular field. Leighbody states this same idea in this way. "Tests which cover only a limited number of items taught in a lesson, or in a course, will show a high level of accomplishment on the part of those who are relatively strong in the items selected. Those who are proficient in other items, which may be just as important, (but not included) will be at a disadvantage. A test with a wider range of responses would afford a better measure of the achievement of all."¹⁴

Objectivity

The third criterion to consider when constructing an evaluative instrument is that of objectivity, which means the absence, in so far as possible, of personal judgment, estimate, attitude or opinion: in other words, it means the freedom from subjectivity. This very important factor in measuring the value of any testing program is actually three-fold; it is concerned with each of the areas of construction, administration, and evaluation or scoring. The construction, itself, involves two aspects; the provision for impersonal scoring of the test, and the interpretation of each of the items involved. These should be as free of personal bias and limit the possibility of misrepresentation as much as is consistent with maintaining validity.

¹⁴ Leighbody, op. cit., p. 121.

The directions for administration, and the actual use of the evaluative instrument should be the same in regard to each pupil in the class, and to different classes, if the test is used more than once or in more than one division. They should be such that more than one teacher could give the test with much the same results.

Three tests of the objectivity of the actual evaluation can be made. First, any number of competent teachers should evaluate a specific piece of work in the same way. Second, similar work of two or more pupils should receive similar scores. Third, the same work, if graded at subsequent times, should receive a mark in accord with the first one. Ross sums this up by saying, "By objectivity in measuring instruments is meant the degree to which equally competent users get the same results."¹⁵

Unfortunately, this type of construction, administration, and evaluation necessary to get these objective results is very difficult, if not impossible. In fact, Brownell points out that even in the best of objective tests there are always many subjective factors present. He lists at least eleven which are always pertinent in any given testing situation.

Well, first of all, in the practical circumstances of teaching one decides to give a test. The decision is surely not based upon purely objective considerations. Second, one determines whether to make a test or to buy one...Third, one

¹⁵. Ross, op. cit., p. 83.

makes up one's mind regarding the kind of test-- whether it is to be of the traditional type, or the newer types, or a combination--judgment again. Fourth, one settles upon the scope of the test-- judgment once more. Fifth, one selected the items to be included--little objectivity here. Sixth, one chooses the form to be employed--true-false, multiple choice, or what not--again little objectivity. Seventh, one frames the items as carefully as one can--and once more has only his judgment for guidance. Eight, one prepares a key by listing the correct answers--a judgment which may not be acceptable to other teachers even of the same subject. Ninth, through opinion one defines the conditions of administering the test. Tenth, one scores the papers--at last objectivity. But, eleventh, one assigns marks--another increment of judgment, and a big one."¹⁶

While it is true that, as long as tests are made and graded by people, personal judgment is bound to enter in, if a conscientious effort is made to be objective, the situation should be much more satisfactory than in the instances in which the teacher makes his own snap judgment without having previously decided upon the standards to be set up to help him in the busy shop situation. The use of personal judgment unaided by the best type of objective standards is apt to have many bad results. When a teacher does not appraise a piece of work carefully, because he is too busy, or for some other reason, he may accept workmanship one day that he would consider unacceptable another. This is likely to confuse the pupils and cause resentment, especially in cases where one boy's inferior work is accepted one day and another's

16. "The Use of Objective Measures in Evaluating Instruction", Educational Method, William A Brownell, 13:401-408, May-June, 1934.

refused at another time. Leighbody says that "Unless objective standards are set and unless the decision as to whether they have been met is made through objective testing, irritating attitudes, or personal characteristics of learners, rather than their actual work performance will tend to determine their progress in the course---The pupil's accomplishment is measured by the examiner and not by the test."¹⁷

Discrimination

The fourth criterion to be considered in the construction of evaluative instruments is that of discrimination, which means that a test will show variations in short gradients in picking out the more acceptable work from that which is less acceptable. Micheels and Harnes state this idea in the following way. "A test discriminates when it is constructed in such a manner that it will detect or measure small differences in achievement or attainment--when it picks out the good from the poor."¹⁸

Just as the objectivity of any evaluative instrument was concerned with three factors; so is the discriminatory facet of a test. First, the discrimination must be shown in the range of achievement measured. It should vary from the highest to the lowest when given to pupils whose success in this area of participation is markedly varied. If too many stu-

17. Leighbody, op. cit., p. 121.

18. Micheels and Harnes, op. cit., p. 118.

dents rate high, the test hasn't been discriminatory in including the outside limits of what these students could do. If too many of the students rate too low, it means that the test hasn't been within the range of the achievement of the students.

Second, all areas of the project involved, from the very easy to the very difficult must be included. This means that part of the evaluation should be of such a nature that most (but not all) of the students may be successful. Part of the testing device should be so difficult that few or any will be measured as having obtained optimal achievement on it. Between these two extremes should be many individual items of evaluation which will test successively, the simplicities or difficulties of the course.

Third, each part of the test should be so constructed that it will be consistent in dividing the more successful pupils from those who are less successful. This is important because a test may discriminate in two ways: positively, or negatively. A positive discrimination is evidenced by an item being answered correctly by the majority of the successful pupils, in contrast to the incorrect answer of the less successful group. If an item is answered correctly by a larger percentage of the group having the lowest total scores than by those having the highest total scores, the discrimination is considered to be negative. This test item

should be considered carefully by the teacher to see what the better students are reading into it. Quite often they see more in an item than the maker of the test planned to include.

This consideration of each evaluative instrument from the three standpoints just named: range of achievement, inclusion of all levels of difficulty, and in judging of each item as to its positive or negative discrimination may seem quite difficult and involved. It is, but in the best use of evaluation, it is worthwhile. Michcoels and Karnes tell us, "As is true with validity, reliability, and objectivity, the discriminating power of a test is increased by concentrating on the improving each individual item in the test."¹⁹

Comprehensiveness

The fifth criterion is that of comprehensiveness. By this term we mean completeness, thoroughness or adequacy, so that the comprehensiveness of a measuring instrument is the thoroughness with which the objectives are covered. This criterion is very closely allied with that of validity in that both require a good sampling of all the material covered to obtain the objectives of a course or a particular part of a course. If a measuring instrument or test includes only one question regarding each item of the objectives, when in reality there are many, then the validity of the

19. Ibid, p. 119.

test will be questionable. All phases of instruction which are supposed to be covered by the test should be liberally sampled. Again, unfortunately, this is easier to suggest than to accomplish. According to Micheels and Karnes, "There is no specific formula which indicates when a test meets the criterion of comprehensiveness--it is a matter of judgment."²⁰ (Note Brownell's list of subjective evaluations, page 14-15). They continue, "For the classroom teacher the best practice is a careful consideration of an answer to this question: Is this test comprehensive enough to measure accurately and well what I expect to measure?"²¹

Administration and Scoring

The sixth and seventh criteria can be considered together, as, for all practical purposes, they are closely allied. These criteria are those of administration and scoring, which means the presentation of the evaluative instrument and the grading of the pupil's efforts in terms of it. These two terms have been considered under objectivity in their relation to that criterion, but they have an added significance in their own right. This pertains particularly to the ease with which they may be administered and scored. Both are of practical significance from the standpoint of the busy teacher. A test may be valid, reliable, objective, discriminatory,

20. Ibid., p. 122.

21. Loc. cit.

and comprehensive, but still be a very poor test because it is so awkward and unwieldy and time-consuming; either in its presentation, or its grading, or both. If these processes are too difficult or bothersome the test will simply not be used, but will be discarded in favor of other tests which may be less valid, etc., but which can be administered and scored with a minimum expenditure of effort, time and bother. Leighbody says "The whole testing procedure may fall into disuse as a result." He hastens to reassure us that "It is possible to devise tests which meet all the criterions suggested above, which can be given upon short notice to one or to several pupils and which consume a minimum amount of time on the part of pupil and teacher, either in completing or checking the results."²²

Areas of evaluation

Certain unique characteristics of industrial arts delimit and qualify the types of evaluative instruments which can be used satisfactorily. There are three areas within this field which will be considered. First, there are different types of tests which vary depending upon the area of evaluation being made; whether in the subject-matter information field; in that of the techniques of manipulative skills; or in that of the habits, traits and attitudes. Under subject-matter information there are different types of tests based upon their inception: whether they are standardized tests, or

²². Leighbody, op. cit., p. 122.

those which the teacher may construct for his own use. Second, in the field of technical skills there are manipulative performance tests and project charts. Third, rating scales are used as a means of evaluation in the area of habits, traits, and attitudes.

The following is an example of the unique problems and situations existing within the specific field of industrial arts. The use of standardized tests has become quite widely accepted in certain other areas of the curriculum. However, the importance of manipulative skills and the difficulty encountered by any test in attempting to measure those in terms of the varied objectives of the industrial arts teacher make most of the present standardized tests inadequate in this field.

In order to develop this point further and to show how impractical it would be for a teacher of industrial arts to attempt to use a standardized test to evaluate his own standards set up in relation to the equipment and supplies available in his own shop, consider the nine objectives which the American Vocational Association set up for industrial arts in their 1953 bulletin.²³

1. Interest in Industry. To develop in each pupil an active interest in industrial life and in the methods and problems of production and exchange.

²³. "Improving Instruction in Industrial Arts", American Vocational Association, Report of Committee, Homer J. Smith-Chairman, A.V.A. Inc., Washington, D. C., June, 1953, p. 18.

2. Appreciation and Use. To develop in each pupil the appreciation of good design and workmanship and the ability to select, care for, and use industrial products wisely.

3. Self-realization and Initiative. To develop in each pupil the habits of self-reliance and resourcefulness in meeting practical situations.

4. Cooperative attitudes. To develop in each pupil a readiness to assist others and to join happily in group undertakings.

5. Health and Safety. To develop in each pupil desirable attitudes and practices with respect to health and safety.

6. Interest in Achievement. To develop in each pupil a feeling of pride in his ability to do useful things and to develop worthy leisure-time interests.

7. Orderly Performance. To develop in each pupil the habit of an orderly, complete, and efficient performance of any task.

8. Drawing and Design. To develop in each pupil an understanding of drawings and ability to express ideas by means of drawing.

9. Shop Skills and Knowledge. To develop in each pupil a measure of skill in the use of common tools and machines and an understanding of the problems involved in common types of construction and repair.

Any one of these items is dependent upon the subjective judgment of the evaluator, and no standardized test could adequately evaluate them.

This leaves the teacher with the responsibility of constructing most or all of his own evaluative devices. These must be "tailor-made" to fit his own particular testing situation. Fortunately, although the teacher must, of necessity, construct his own test, this very factor of being "tailor-made" will make them more valuable to him because they will

test the unique situation in which his pupils work.

Since he does have to make the tests he should understand and use the criteria which have been discussed. His failure to do so will cause him difficulties in his use of the results of his evaluation. Leighbody says, "To neglect the proper development and use of tests in teaching industrial subjects is largely to nullify the efforts which may be made to have the methods used in the other teaching steps result in efficient teaching and effective learning."²¹

These teacher-made tests include the pencil-and-paper information tests which largely evaluate the subject-matter learning, the rating scale, progress charts, and actual laboratory tests which are set up for the express purpose of evaluating manipulative skills.

The pencil-and-paper tests may serve efficiently enough to evaluate the related subject-matter information needed in industrial arts, but they will not be satisfactory in the measurement of manipulative skills. For instance, a student might be able to answer correctly any number of questions on how to turn cylinders in a metal lathe, and still be unable to actually carry out the manipulative skills necessary to do the cylindrical turning. On the other hand his failure to answer questions satisfactorily would not prove his inability to turn cylinders. Paper-and-pencil tests will not be ade-

²¹. Leighbody, op. cit., p. 119.

quate in this area of evaluation.

Since the development of a certain amount of skill in the performance of manipulative operations is one of the stated and acceptable objectives for almost every industrial arts course, the total program of evaluation for such courses must provide some means for measuring the ability to perform manipulative operations, so the child needs the special types of manipulative performance test which will give him a chance to prove that he can do the type of work desired. If it is given in the correct way it will not only help the teacher in diagnosing the pupil's strengths and weaknesses but will help the pupil to recognize them himself.

This recognition of his own strength and weakness can be an incentive to the pupil to improve his skill. The very set-up of the test is such that it encourages this type of pupil learning. He knows the type of skill which he is expected to attain so that he can practice on it, get help with his problems, and thus assure his own success. As Holmboe says, "Success in the performance test... is achieved through previous practice, and knowledge of what performances will be required, only stimulates effective practice."²⁵

Very closely allied to the manipulative skill test of the techniques used in industrial arts is the progress chart which lists the pupil's achievement as he works on a specific

²⁵. Ibid., p. 123-124.

project. Since it is very difficult for the teacher to look at the finished project and evaluate all of the learnings and effort which have gone into it, he needs a chart to check the progress of the pupil. "There are certain important evidences of growth which are revealed only during work periods in which students are given an opportunity to make application of information, procedures, and principles taught. This means that the instructor must make careful observations for the expressed purpose of evaluating achievement."²⁶ This chart, too, serves as a means for the basis of a teacher's reteaching and also for the motivation of the pupil as it indicates the important operations or skills which are included in the project.

A pupil's habits, traits and attitudes are a basic concern in determining the type of work he does. Not only does the teacher need to be constantly aware of them in his day by day association with the child but he also needs a more concrete rating to aid him in his required evaluation. Rating scales have been devised for this purpose. These help the child know what attitudes are desirable and can serve as the basis for such behavior changes as he can bring about through his own effort.

Summary

As a basis for this paper on the criteria for the construction of evaluation instruments to be used in industrial

²⁶. Micheels and Karnes, op. cit., p. 39 $\frac{1}{2}$.

arts classes, the importance of the use of the results of these instruments has been indicated. Second, the paper has discussed the seven criteria which must be applied to the construction of any adequate test: validity, reliability, objectivity, discrimination, comprehensiveness, administration, and scoring. Third, it has shown the areas of evaluation: subject-matter information, which includes standardized and teacher-made tests; techniques involving manipulative performance tests and progress charts; and habits, traits, and attitudes which may be evaluated by means of rating scales.

SELECTED REFERENCES

- Brownell, William A., "The Use of Objective Measures in Evaluating Instruction", Educational Method, 13, 401-403, May-June, 1934.
- Lawshe, C. H. Jr., Principles of Personnel Testing, McGraw-Hill Book Co., Inc., New York., 1940.
- Leighbody, G. B., Methods of Teaching Industrial Subjects, Delman Publishers, Inc., Albany, N. Y., 1946.
- Micheels, William J., and Karnes, H. Ray, Measuring Educational Achievement, McGraw-Hill Book Co., Inc., New York, N. Y. 1950.
- Newkirk, Louis V., and Green, Harry A., Tests and Measurements in Industrial Education, John Wiley and Sons, Inc., New York, N. Y. 1935.
- Remmers, H. H., and Gage, N. L., Educational Measurement and Evaluation, Harper and Brothers, New York, N. Y., 1943.
- Ross, C. C., Measurement in Today's Schools, Prentice-Hall, Inc., New York, N. Y., 1947.
- Smith, Homer J.-Chairman, "Improving Instruction in Industrial arts", American Vocational Association, A.V.A., Inc., Washington, D. C., June, 1953.
- Wilber, Gordon O., Industrial Arts in General Education, International Textbook Co., Scranton, Penn., Second Edition, 1954.