

**THE USE OF TESTS IN
SCHOOLS OF NURSING**

**The Construction and Use of
Teacher-Made Tests**

SECOND EDITION

NATIONAL LEAGUE FOR NURSING

1965

**THE USE OF TESTS IN
SCHOOLS OF NURSING**

**The Construction and Use of
Teacher-Made Tests**

Pamphlet No. 5

Prepared by
Mary R. Shields

Second Edition

NATIONAL LEAGUE FOR NURSING
Test Construction Unit
10 Columbus Circle, New York, New York 10019

1965

Code Number: 14-136

Price: \$2.50

Copyright, 1965, by
NATIONAL LEAGUE FOR NURSING

PRINTED IN THE UNITED STATES OF AMERICA

THE USE OF TESTS IN SCHOOLS OF NURSING

Pamphlet No. 1

The NLN Pre-Nursing and Guidance Examination
(*Fourth Edition*) \$1.25

Pamphlet No. 2

The NLN Achievement Tests (*Third Edition*) \$1.25

Pamphlet No. 3

The NLN Graduate Nurse Qualifying Examination
(*Second Edition*) \$1.00

Pamphlet No. 5

The Construction and Use of Teacher-Made Tests
(*Second Edition*) \$2.50

**The Construction and Use of
Teacher-Made Tests**

**MEDICAL BOOKS
FOR
CHINA**

Foreword

This manual is designed to assist instructors in schools of nursing to build and use tests that will measure their students' attainment of nursing abilities that are measurable by written tests. It may also be useful to students who are preparing for positions in nursing education and to teachers in other fields. The manual is not intended to cover the entire field of evaluation, nor is it intended as a statistical text. The section on analysis of test data is confined to those concepts that the Test Services of the National League for Nursing are most frequently called upon to explain.

Contents

FOREWORD	v
INTRODUCTION	1
Some Uses of Evaluation Tools in Schools	1
Facets of Evaluation of Student Achievement	3
Standardized vs. Teacher-Made Tests	4
Choosing Methods for Evaluation of Student Achievement	6
CONSTRUCTING TEST ITEMS	8
The Essay Test	8
Objective Tests	15
General Guides for Writing Objective Test Items	17
The Completion Item	21
The True-False Item	23
The Multiple-Choice Item	25
Special Options	38
The Combination-Response Item	39
The Analogy Item	42
The Matching Item	43
Drawings and Other Nonverbal Material	48
CONSTRUCTING TEACHER-MADE TESTS	49
Steps in Developing a Blueprint	51
Constructing Test Items Based on the Blueprint	59
Validity and Reliability	65
Assembling the Test	67
Scheduling and Administering the Test	70
Scoring Objective Tests	72
ANALYZING AND INTERPRETING TEST DATA	75
Item Analysis	75
Analyzing and Interpreting Test Scores	81
Table 1. Frequency Distribution of Scores on Unit Examination "Health Problems of the Aged" for 80 Students (Ungrouped Data)	82

Continued on next page

Contents—Continued

Table 2. Frequency Distribution of Scores on Unit Examination "Health Problems of the Aged" for 80 Students (Grouped Data)	83
Figure 1. Frequency Polygon of Data Presented in Table 2	84
Measures of Central Tendency	85
Measures of Variability	86
Table 3. Computation of the Standard Deviation	88
Percentiles	89
Figure 2. Normal Distribution of Examinees Divided Into Deciles	90
Percentiles of Average Scores	91
Standard Scores	93
Figure 3. The Theoretical Normal Curve and Its Relationship to Standard Scores and Percentiles	94
USING TEACHER-MADE TESTS	97
Tests in the Integrated Curriculum	97
Identifying Strengths and Weaknesses by Means of Pretests and Diagnostic Tests	100
Pretesting	100
Diagnostic Testing	101
Test Difficulty Levels	101
Grading	103
Uses of Tests for Purposes Other Than Measurement	109
CONCLUSION	110
An Ongoing Task	110
BIBLIOGRAPHY	112
INDEX	113

Introduction

In identifying the abilities that are prerequisite to graduation from an educational program, the faculty makes a judgment about the needs of society and the student. The objectives, the content, and the organization of an educational program reflect the faculty's interpretation of these needs. When students are graduated to go out to exercise their abilities in human society, the school has passed judgment on their competence. Ultimately, the school's decisions are evaluated by society, but the school, as an agency with a social conscience, is also concerned with self-evaluation. The recurring issue for any faculty is whether or not the school is turning out graduates with the required abilities. This issue presents the faculty with two major problems: (1) Are the abilities established as prerequisites for graduation those that will best meet society's and the student's needs? and (2) Have the students attained the requisite level of achievement? It is the responsibility of every faculty to develop a program of evaluation that will provide the answers to these questions.

Evaluation of a school's over-all aims generally arises out of philosophical insights and is based on evidence relative to human needs and resources. It involves comparisons with the goals of similar educational institutions and at best, reflects some of the leading thinking of the day from a variety of fields that affect human well-being, including the political, economic, and spiritual realms. Although faculty participation in this phase of educational evaluation is extremely important, it is not the subject of this pamphlet. We are concerned here with the second problem: Are the students attaining the desired level of achievement? It is here that data from devices for measuring student attainment, including tests, are especially useful.

Some Uses of Evaluation Tools in Schools

As teachers know, the job of assigning course grades is only one aspect of a teacher's concern with the problem of evaluation. Teachers evaluate the curriculum in its broadest sense; the total learning milieu; that is, they not only measure student achievement but also appraise the effectiveness of their methods of instruction and of different educational settings. To

base their evaluations of the learning milieu on evidence more tangible than personal opinion, teachers use tests and other data-gathering devices. An educational test may be thought of as a device consisting of a series of questions or exercises designed to measure aptitudes, capacities, or achievement. The emphasis in this manual is primarily upon written achievement tests.

What kind of evidence is provided by tests and other evaluation tools used in schools? What is their basic utility? If the fundamental goal of an educational program is to effect a change in its students (in a direction dictated by the needs of society) by enabling the students to react to various learning experiences, then evaluating the educational program depends upon measuring the change. This change has been labeled student achievement. When unassisted, a teacher's appraisal of student achievement tends to be based on gross and subjective sensory impressions. (For example, an attractive, friendly, and somewhat glib student may be thought to have more understanding of a subject than she really has.) Evaluation devices, including tests, are developed for the purpose of obtaining more objective evaluations of student performance.

To what use may teachers put this more objective, more discriminating knowledge of student growth? The data, of course, can be used to compare the levels of achievement of an individual student at various stages in her program and also to compare her achievement with that of others in her group or with some absolute standard. The data may also be used to compare the achievement of a group of students with that of another group or groups. However, not only the attainment of the students but also the effectiveness of the other two major components of the learning milieu—the teacher and the course of study—may be assessed by using evaluation data.

In formal reasearch, each component may be appraised by maintaining “control” over the remaining two. For example, the adequacy of the content of a particular course may be evaluated by comparing the performance of students in that course with the performance of students in similar courses throughout the country. Comparisons of this kind presuppose students of equal potential, teachers of equal competence, and educational settings that provide equal opportunities for learning. Similarly, to determine whether one method of instruction is superior to another, two equally able teachers would use two different methods of instruction to help two equally able groups of students to learn the same content. If an evaluation device such as an achievement test is then administered to both groups, differences in student growth may be regarded as reflecting differences in methods of instruction. Of course, conclusions based on such data

must necessarily be limited to those abilities measured by the tool. Carefully controlled experimentation of this kind is more easily described than done. But whether or not such research is possible in an individual school, teachers learn to use a variety of means to evaluate the curriculum and student attainment.

Facets of Evaluation of Student Achievement

The teacher in nursing must, of course, evaluate behavior—observable behavior in the broad sense, including all actions, such as dressing a wound or answering a toddler's question, and within these actions evaluate a host of factors, such as dexterity, gentleness of touch, tone of voice, and quality of silence. Behavior, in the sense in which it concerns the teacher, may be thought of as the complex product of a number of reactants, among which are understanding, emotion, and motor-sensory aspects.

Tools for the evaluation of clinical performance, such as rating scales, generally deal with behavior involving all three of these realms (cognitive, affective, and motor-sensory). When we say, for example, "The nurse should be observant," the observation of which we speak is not a discrete ability, but a behavior. Observation in this sense requires, of course, the physical processes of seeing and hearing and, perhaps, touching and smelling. The affective realm is involved also: attitudes that determine what one is willing to pay attention to, feelings about values, emotion that may help or may hinder the acuity of the observation. Observation involves, too, some aspects of cognition; even when we attempt to minimize the cognitive aspects, observation involves, at the very least, some identification by the observer of that which is observed, a decision as to whether the phenomenon is the same as or different from those she has observed previously or elsewhere, and the formulation of some simple idea about the phenomenon. In the evaluation of clinical performance, where the teacher sets out to observe the student who is observing the patient, what complexity is involved!

Since our only reasonable clues as to what another human being *is* are in terms of his behavior in the broad sense, the emphasis on evaluation of behavior in professional education is certainly warranted. The behavior of the product is, after all, the proof of the educational pudding. But evaluation of behaviors of nursing students is expensive, much of it demanding that teacher and student be in a one-to-one relationship for long periods of time. Such evaluation tends also to be unreliable, because the behaviors

under observation are complex and variable, the rater herself is a constantly changing person, and tools used as guides for rating are often clumsy and sometimes misleading. In addition, when the teacher has decided that the behavior evidenced by a student is inadequate (in terms of educational goals) she may have little evidence as to whether the impediment or defect lies mostly in intellectual understanding or in emotion or in the motor-sensory realm. And if all of these are involved in the impediment or defect, which aspect is mostly cause and which ones are mostly effect? Economy of human and material resources in teaching demands the garnering of available information on the components of behavior, to supplement information on behavior itself.

The component (or reactant) of behavior that can most easily be gotten at is the area of cognition, and a large part of formal evaluation of nursing student attainment pertains to that realm—that is, to the measurement of knowledge, understanding, judgment, application, and (to some extent, let us hope) ingenuity and intellectual creativity. It is in this realm that achievement tests, both standardized and teacher-made, are especially useful.

There is no intent to deny here the fact that physical and emotional factors can color a student's responses to a paper-and-pencil achievement test; nevertheless, the test deals more directly and purely with the cognitive aspects than do most other measurement devices used in nursing education.

Standardized vs. Teacher-Made Tests

Teachers use tests for two different purposes—to discriminate among students exposed to different learning milieus, and to evaluate students exposed to the same learning milieu. Standardized tests are especially useful for the first purpose, although they may be partially effective for the second purpose. Teacher-made tests are generally constructed specifically for the second purpose.

A teacher is primarily concerned with the degree to which her students achieve the objectives of the course she is teaching, but she is also interested in knowing how well her students compare with students who are enrolled in similar courses elsewhere. When educators wish to compare students in different classes and different schools, they require tests that will evaluate in terms of a common core of achievement—that is, tests that do not favor the unique curricular emphases of any instructor, school, or region. Now, just as it is true that the only person completely equipped to construct a test to reflect the objectives of a particular course

is the course instructor, so is it true that no individual teacher or faculty is likely to construct a test that will represent fairly the emphases of teachers throughout the country. Because the construction of standardized tests requires time, funds, and specialized knowledge, faculties usually entrust the construction of such tests to agencies equipped to build them.

The special utility of standardized achievement tests stems from the fact that they can be constructed in terms of nationally representative educational objectives. They are constructed to measure the common core of achievement in the cognitive realm. Such a test, standardized on the basis of a representative group of examinees, yields a score that makes it possible to compare a student's performance with the performance of students throughout the nation. Standardized tests, therefore, can be used to compare the performance of students in one school with the performance of students in other schools. Thus, faculties may get clues as to the effectiveness of various learning milieus. Interpretations must be made with caution, for the students are compared only with regard to commonly accepted objectives and not with regard to the special emphases of the particular teacher or school.

Every teacher is aware of the fact that selection procedures rather than the curriculum may be responsible for the performance of students on achievement tests. A school might conceivably have better students, not better teachers or more meaningful objectives. On the other hand, a school may have students of such low potential that the finest educational program would not help them to achieve the national average level of performance. For this reason, when comparing the merits of schools in terms of the performance of students, attention should always be given to what the student has brought to the school—that is, the selection data. These facts do not mean that achievement tests are relatively less important, but rather that adequate selection devices are also highly important. It is evident, too, that important educational decisions should be based, not on isolated facts, but on data gathered from an integrated program of evaluation.

While the special worth of a standardized achievement test is derived from its ability to measure the student's attainment of widely accepted objectives, the particular merit of a teacher-made test stems from its capacity to reflect the teacher's own curricular emphases. The teacher-made test can appraise student growth in terms of the teacher's particular educational objectives, which, of course, take into consideration the student's own objectives. Many of the teacher's objectives may be the same as those upon which standardized tests are based, but it is unlikely that they will all be identical.

Because of its direct relationship to the objectives of a specific course, a teacher-made test can be used as a partial basis for determining a student's achievement in that course. This is to say that, in conjunction with other evaluative devices such as interviews, student self-evaluation records, nursing care studies, and special assignments, a teacher-made test can be utilized in the determining of student grades. Moreover, a teacher-made test can be used as a teaching aid that reflects to the student the teacher's course emphases and as a diagnostic tool that reports to the teacher those aspects of the course with which students are having the most difficulty. These uses are considered more fully in a later section of this manual.

In summary, then, it can be stated that the well-constructed, nationally standardized achievement test can evaluate student attainment of widely accepted educational objectives and hence provide a method for the evaluation of a school's curriculum, whereas the well-constructed teacher-made test helps to evaluate student attainment of a school's own objectives, which may be similar to, but which are not likely to be identical with, nationally accepted objectives.

Choosing Methods for Evaluation of Student Achievement

Teachers constantly must make decisions as to what evaluation methods will best measure different aspects of student growth. To make the problem of choice of evaluation methods clearer, let us first consider briefly the basis for an instructor's choice of teaching methods.

The instructor has the responsibility of choosing from among many different teaching methods those that will most effectively and efficiently attain the objectives of the course. It is obvious that the large-group-in-a-classroom situation is the most economical with respect to teaching time. Teachers would agree that some objectives in nursing education may be attained very well in such large classroom groups, while some objectives call for the use of small groups with increased individual student participation, and others require that the individual student and teacher think and work together in planning for and with a patient. The last kind of teaching is by far the most expensive in terms of the instructor's time, but teachers would agree that it is essential for the attainment of some of the most important objectives of nursing education. The instructor who is responsible for a course decides what objectives can be attained effectively only by using the expensive one-to-one student-

teacher ratio, what objectives seem to require that students learn in small groups, and what objectives can be attained just as well in large classroom groups. Such an analysis is basic in course-planning.

The same kind of analysis is necessary for the thoughtful planning of the evaluation of student achievement. The evaluation of attainment of some objectives requires the instructor to work with and observe the individual student while she is caring for a patient in a nursing situation. Evaluation of some other aspects of student growth may require the use of the personal interview. In terms of the teacher's time these may be the most expensive kinds of evaluation. In order to assess the attainment of some other objectives such as social poise, speaking ability, and the ability to get along with co-workers, personal observation of students, in small or large groups, is probably the only means of evaluation. There are, however, many objectives pertaining chiefly to the cognitive realm (knowledge, understanding, et cetera) that can be tested just as well by group paper-and-pencil tests, and since this is the most economical means of appraising student growth, such tests should be employed whenever feasible. This manual is concerned specifically with the improvement of teacher-made paper-and-pencil tests.

The first step in developing a plan for the evaluation of student achievement in a course or a unit of instruction is to make an outline, or "blueprint," of measurable abilities and course content. The development of such a blueprint, however, requires knowledge of the kinds of test questions and techniques of construction. The technical aspects of item construction, therefore, are presented next, and the development of the blueprint is described later. It must be understood that the following section deals primarily with the mechanics of individual test items, not with their value in measuring the attainment of objectives.

Constructing Test Items

Teacher-made paper-and-pencil tests may be classified into two types—the essay test and the objective test. Good tests of both kinds are useful to the teacher in measuring achievement.

The Essay Test

In an essay examination, the student is presented with a few questions to which she must formulate her own answers. The student receives credit to the extent that her answer satisfies the instructor's criteria for accuracy and completeness. Answering an essay question always involves the student's recalling something and presenting the answer in writing. It frequently involves, also, the organization of knowledge and the making of judgments. It may involve, if constructed to do so, a probing of the depths of the student's understanding or of aspects of her practical ingenuity and creativity.

The essay test is a suitable tool for evaluating the student's ability to communicate in writing. If, however, the essay test is being used for some other purpose, the communication-ability factor may be a contaminating one. Students who write well may sometimes be able to "bluff" in an essay test. It is difficult for a teacher to remain uninfluenced by such factors as vocabulary, sentence structure, and spelling—not to mention handwriting! Obviously, it is important for the teacher to decide what she wants to test with an essay question, to construct the question accordingly, and to score as objectively as possible on the basis of her decision.

Answering an essay question generally takes a considerable amount of the student's time, so that only a few questions can be used in most essay tests. This may result in a narrow sampling of the student's knowledge; a narrow, deep "wedge" is obtained rather than the broad sampling obtainable with objective tests. Essay and objective tests, therefore, complement each other.

Because the essay examination includes only a few questions, it generally takes a shorter time to prepare than does an objective test. It must be remembered, however, that *good* essay questions cannot be formulated hastily and also that since the scoring is time-consuming, any time

saved in the preparation by using an essay test will be lost in the scoring unless the class is quite small.

It is sometimes stated as an advantage of essay tests that they can be written on the blackboard and do not require special reproducing equipment. Few blackboards, however, are large enough for a series of well-constructed essay questions unless the questions are at a very simple level. With the increased use of classroom projectors in presenting short tests of either the essay or the objective type, this point becomes less important.

Whatever other limitations essay examinations have, the major difficulties are in scoring. Difficulties of scoring arise from the fact that students produce answers of varying degrees of completeness and accuracy. Teachers must often make distinctions among very subtle differences. For this reason, grading is likely to be wearisome and is almost certain to be somewhat subjective and unreliable. Other equally expert teachers would be likely to give the same papers quite different scores. Unless the teacher is aware of the many variables that can affect the score the student is given, the value of the essay examination is seriously limited.

The advantages and disadvantages of essay examinations, therefore, may be summarized as follows:

Advantages

1. Take a comparatively short time to prepare.
2. Test the ability to select, organize, and communicate facts and concepts in writing.
3. May probe depths of student knowledge; may test ingenuity.

Disadvantages

1. Take a long time to score.
2. Only a narrow sampling of knowledge may be tested.
3. Scoring is unreliable.

There are ways by which the instructor may overcome some of the scoring difficulties. The first step is to write essay questions that do not permit the student too wide a range of interpretation. In a very real sense, if the essay question is too vague, students will be reacting to different tests; evaluation of student achievement, therefore, will not be based on a common yardstick. For the same reason, it is usually inadvisable to permit students to select the questions they will answer. Although students may like choosing "six questions out of ten," the technique results in a comparison of students on different yardsticks of achievement.