Unit 4
Instructional Evaluation

- Purposes of Instructional Evaluation
- Norm-Referenced and Criterion-Referenced Test
- Test Construction
- Appraising Assessment
- Portfolios
Purposes of Instructional Evaluation

The reasons for completing an evaluation can be quite varied. From the perspective of assessing student knowledge and skills, evaluation can be helpful before, during and after instruction.

Before instruction, evaluation can serve as a means to

- identify students’ preinstructional knowledge and skill levels. This indicates (1) whether students have the prerequisite skills and knowledge they need to begin the instructional materials, or (2) whether they already know the content they’re slated to learn. In the former instance, a lack of prerequisite skills requires that students receive remedial instruction in order to attain the entry skills and knowledge necessary to learn the new information. In the latter case, evaluation indicates if instruction is even needed, or if a review will suffice. This form of evaluation is often referred to as a pretest or a preinstructional evaluation.

- focus learner’s attention on the important topics that will be covered in the upcoming instruction. By completing an evaluation prior to instruction, the learner is primed to perceive those important topics when they are later experienced within the instructional materials (Fleming, 1987)

- establish a baseline of performance that can be used as a comparison with postinstruction performances. This can indicate progress and may motivate students when they see what has or hasn’t been accomplished since the preinstructional assessments. One word of caution: students need to be informed of the value of pretest. It should be explained that a pretest merely establishes a baseline—a base point from which the work will begin. Lower levels of performance prior to the instructional materials are expected.

During instruction, evaluation serves as a means to

- demonstrate what has been learned to that point. It supplies information about student’s learning that indicates whether or not they have learned what is needed to progress to the next portion of the instruction. This form of assessment is frequently referred to as formative evaluation of students. This information may be used by either the teacher or the student to
determine if new instruction should be introduced or if additional practice and feedback is needed.

- supply corrective feedback as the learning process occurs—that is, identify and correct mistakes and problems before they become thoroughly ingrained and practiced in an incorrect fashion. In a similar manner, formative evaluation can increase students' confidence by indicating that the content has been mastered to that point.

- identify when, where, and what type of additional practice may be needed. At this point in the instruction, evaluation ensures that the new knowledge and skills are being integrated with previously learned information and that they can be recalled when needed.

- refocus attention on the desired outcomes. Because it's easy to lose sight of one's reason for learning, it's helpful to be reminded of the expected outcomes during the learning process. An evaluation during the instruction is an effective tool for refocusing learners' attention or purpose of their efforts.

After the instruction, evaluation serves as a means to

- measure what the students have learned. This is the most frequent use of formal evaluation and is commonly referred to as summative evaluation of students. It gives students the opportunity to demonstrate what they have learned after they have experienced all of the instruction. Because this evaluation generally comes at the conclusion, or summation of the instruction, students frequently prepare for it by devoting time and effort to study outside of the classroom. It can be a period of reflection and/or additional practice that leads to a synthesis of the separate parts of the instruction.

- make decisions about student accreditation, advancement, or remediation. In many cases, the results of summative evaluations are used to make specific decisions. For the medical student, there is a medical board exam; for law student there is a bar exam, for those wishing to drive a car there is a driver's license exam. These assessments are used to make decisions about a person's level of skill and knowledge. In some cases students' performances will produce accreditation and advancement. While at other times further study and remediation will need to be provided.

ED202
• review important formation and prepare for transfer of the information to new and different situations. Learning new information and/or skills is important, but the value of this learning increases when it is shown to be useful in a variety of contexts. This form of evaluation can be used as a review and as a preparation for transferring and generalizing to new situations. (Newby et al., 1996, 273-275)
Norm-Referenced and Criterion-Referenced Tests

To evaluate your learners’ progress, what type of information do you need?

That depends on your purpose. Testing can provide two types of information:

1. A student's place or rank compared to other students is revealed by a norm-referenced test (NRT), so named because it compares a student’s performance to that of a “norm group” (a large, representative sample of learners). Such information is useful when you need to compare a learner’s performance to that of others at the same age or grade level.

2. A student’s level of proficiency in or mastery of a skill or set of skills is revealed by a criterion-referenced test (CRT), so named because it compares student performance with an absolute standard called a criterion (such as “75% correct”). Such information helps you decide whether a student needs more instruction to acquire a skill or set of skills.

Unfortunately, some teachers know little more about a student after testing than they did before. In our technically oriented society, test scores sometimes have become ends in themselves, without the interpretation of them that is essential for improvement of the learner. In such case teachers, parents, and others may be quick to denounce a test, often suggesting that such abuse of testing “proves” that test data are useless. In reality, it may only indicate that the teacher who selected the test either failed to identify the specific information needed before administering the test or failed to carefully match the test to this purpose.

(Borich, 1996, 592-593)
Test Construction

There are important reasons why instructors need to become proficient in constructing classroom tests. They must develop most of the tests they use. If developed by a teacher without special training, essay and objective tests are frequently of poor quality. The research on essay examinations has repeatedly demonstrated this fact. Novice test developers fare even more poorly with an objective test than with an essay examination; some objective tests have even lower reliability than most essay examinations (Stulmaker, 1957-1)

Research has shown that skillfully constructed achievement tests can be as precise as some standardized tests and are usually more valid for a more limited and carefully defined set of objectives and content. Standardized tests rarely assess the objectives of a particular unit of instruction. Standardized tests tend to (1) focus on broad, commonly accepted objectives and (2) cover a wide range of curricular content because they are usually designed to assess a full year or more of instruction. Frequent evaluation is necessary for teachers so that they can monitor the progress of individual children and of the class as a whole. Periodic testing provides a more reliable and valid evaluation of student progress. In this chapter we consider the general principles of constructing achievement tests designed to measure cognitive objectives.

Planning the Test Constructing a satisfactory test is one of a teacher’s most challenging tasks. Good tests do not just happen. Indeed, most teacher-made tests are rather crude measures that are thrown together without much care and forethought. Test construction is in one sense more of an art than a science, but this “art form” can be dramatically improved with special instruction and systematic practice and feedback. There are well-established, proven principles of test development that are often unknown or ignored. Constructing a good test item is a deliberate process; it demands an understanding of the objectives and content being assessed, the reading and vocabulary level of the examinees, and test-taking factors such as response styles and test sophistication.

A good test must be planned; careful planning must precede its construction. One must consider the topics and objectives to be measured, the purpose the scores are to serve, and the
conditions under which testing will occur. The following general guidelines should be followed in
developing educational achievement tests:

1. The test should make provisions for evaluating the important objectives of instruction
that are immediately measurable.

2. The test should reflect the content and process objectives in proportion to their
importance and emphasis in the course.

3. The nature of the test should reflect its purpose (e.g., to assess individual differences or
to certify mastery).

4. The test should be appropriate in length and readability level.

(Hopkins and Stanley, 1981, 165-166)
Appraising Assessments

Ebel (1965) provided teachers with 10 qualities of a good test. The qualities are universal enough to apply to most assessments. Use them to pose the kind of questions we suggest for appraising an assessment:

1. Relevance. Do the items of this test match the objectives and content information of teaching? Does this test measure the learning that I intended for my students?

2. Balance. Does this test represent all the important content that I taught? Is important information given importance in the test items? Is any information given too much representation in the test items?

3. Efficiency. How much time will this test take? Is that amount of time appropriate or proportion to the amount of time I spent teaching the content?

4. Objectivity. When I look at answer to the test items, are they fair?

5. Specificity. Is there a match between the curriculum information and the test items? Could a student who missed class do well on this test?

6. Difficulty. Can at least half of my students do very well on this test? Could each item be answered correctly by at least half of my students?

7. Discrimination. Will my students who worked hard, studied well, and are knowledgeable answer most of these items correctly? Will my students who put little effort, did minimal studying and aren’t knowledgeable answer most of these items incorrectly?

8. Reliability. Would my students score similarly if they took this test two days in a row? Does it matter who gives this test?

9. Speed. Will my slow-working students be penalized on this test? Will there be any problem with some students finishing this test?

(Freiberg and Driscoll, 1996, 390-391)
Portfolios

Arter and Spandel (1992) define the student portfolio as "a purposeful collection of student work that tells the story of the student’s effort, progress, or achievement" (p.36). The portfolio is a rich collection of work that demonstrates what students know and can do. For years, artists have used portfolios to highlight the depth and breadth of their abilities. Similarly, students may use portfolios to illustrate their unique problem-solving or critical-thinking skills, as well as their creative talents (e.g., writing, drawing, design). Additionally, portfolios can be used to demonstrate the evolution students went through to achieve their current performance level. Unlike the end-of-the-unit objective evaluation, the portfolio is designed to capture a greater range of students’ capabilities and to indicate how those capabilities developed and grew over time. Not only does the portfolio convey to others the students’ progression, it also serves as a vehicle for students to gauge their own development and to envision what additional things they might learn.

(Newby et al, 1996, 283)

As suggested by D’Aoust (1992), portfolios may be structured around the exemplary “products” of student works (i.e., including only those of the best quality) or around the “process” by which students arrived at current levels of performance (i.e., pieces from the beginning, middle, and end of the course that show the progression of students abilities), or they could include a mixture of both. In either case, a major benefit comes from actually putting it together. “Students cannot assemble a portfolio without using clearly defined targets (criteria) in a systematic way to paint a picture of their own efforts, growth, and achievement. This is the essence of assessment. Thus portfolios used in this manner provide an example of how assessment can be used to improve achievement and not merely monitor achievement” (Arter & Spandel, 1992, p. 37).

(Newby et al, 1996, 284)