Chapter 6
Quality Control Measures

Chapter 6 Outline
Quality Control Measures
- *Program Evaluation Applied to Curricula*
- *Evaluating Training: Quality Control*
- *Program Assessment*
- *Instructional Assessment*
- *Participant Assessment*
- *Simulations or Exercises*
- *Summary*

Introduction

In this section we will address a critical element of a curriculum - the evaluation methods and approaches. Without a method of assessing the degree to which the instruction is effective in accomplishing the educational objectives, there is no opportunity for judging the effectiveness of the curriculum, instruction or the process. Interestingly, the development of the educational objectives described earlier was an effort to simplify the evaluation process. Bloom stated "curriculum builders should find the taxonomy helps them to specify objectives so that it becomes easier to plan learning experiences and prepare evaluation devices."[1] We maintain that it is vitally important to insert and maintain quality control measures in order to evaluate and assess the success of the curriculum. In the sections following, we will discuss evaluation in general, then curricular evaluation in particular. The literature is clear that curricular evaluation includes assessing the curriculum, the instruction, and the participants.

Program Evaluation Applied to Curricula

Program evaluation has long been a mainstay of social science and justice-related disciplines. When a program is implemented, it is essential to determine the degree to which it is proceeding as planned and that it has accomplished what was planned. We will address the steps in program evaluation and then the specific area of curricular evaluation.

Program evaluation, as with any type of evaluation, requires specific descriptions of the problem which is to be addressed by the program. Maxfield and Babbie[2] state "To conduct evaluation research, we must be able to operationalize, observe, and recognize the presence or absence of what is under study." The first, and often best, place to look for the thing or things under study is to examine the goals of the program or project. If a "program is intended to accomplish something, you must be able to measure that something."[3] Determining how well the program accomplished its goals is an example of impact evaluation. Impact evaluation is the true test of effectiveness and, if done correctly, can help researchers, policy-makers and planners refine the program or initiative so that it can be even more effective in the future. Impact evaluation is the highest level of assessment or evaluation. Typically, the more stringent the methodology used in impact evaluation, the stronger, more reliable and valid the results. Experimental design or quasi experimental design are viewed

as the most valuable research methods in determining the impact of a program. Often these methodologies are not available in criminal justice or the social sciences since randomization is often limited. Ex post evaluations involve the decision to evaluate something after a program has gone into effect. It is more difficult to develop a sound impact evaluation after the fact but it is still possible if the program has been carefully established. That, too, is unlikely since a carefully established program is one which should include an evaluation component.

The impact evaluation of a training initiative would judge or measure the impact the training had and has on the problems or issues which formed the rationale for the training. If the training is to allow the participant to do something or to keep something from occurring, the degree to which it accomplishes the goal is the evaluation of the training.

Process evaluation, according to Maxfield and Babbie,[4] "focuses on program outputs" or seeks to answer the question "Was the program implemented as intended?" Process evaluation, though not as rigorous or as predictive as impact evaluation, is still important. The assumption is that the program or initiative was properly established, planned well, and organized in a clear linear way. If the process is followed, the results should be good. Clearly, nobody would design a program or initiative which was not intended to be effective. Tracking the process, incrementally, can determine if the program is proceeding as planned and, if not, changes can occur so that it will be put back in the planned process.

Conducting a process evaluation of a training initiative or training program would involve tracking the steps in the development, design, implementation and feedback to see that they are conducted as planned. The plans should be precise enough to allow external evaluators to assess the process as it occurs, rather than after the fact.

"Ideally, impact assessments and process evaluations are conducted together."[5] The process evaluation may help to explain variances in the impact assessment results.

Evaluating Training: Quality Control

Maintaining quality control is a key responsibility for those who monitor and refine curricula and training programs. This quality control is typically called "assessment" and may apply to a number of activities. While there are many techniques for accomplishing the quality control, several terms must first be defined so that the usage will be clear.

Assessment	the formal or informal process of measuring an activity or initiative.
Norm-referenced	assessing an individual's achievement measured in comparison peers, a group or cohort, and/or historical data. The achievement or activity is measured relative to the person's own performance or the performance of others. It is a relative comparison.

Criterion-referenced	assessing an individual's accomplishments or achievements relative to some externally defined or explicit criteria or standards of performance.
Ipsative assessment	assessment of an individual's accomplishments or achievements through a self-referenced or personalized criterion. A relatively formal process of self-assessment where one is measured or judged based on criteria they establish for their own performance or achievement by the degree to which they have met their own target(s).
Formative assessment	a step-by-step process of assessing progress. Often based on a learning plan or action plan and the degree to which each element of the plan is accomplished. Allows clarification and explanation of processes and elements as the learner is progressing toward the ultimate objective. Designed to improve the curriculum, improve the individual's performance, or improve the process of learning.
Summative assessment	a comprehensive or formal confirmation of achievement, usually at the end of an instructional program. Often associated with tests or examinations, either practical or theoretical, which require the individual to display mastery of the skill or information. Measures the success of the training or the curriculum in achieving its objectives.
Assessment reliability	refers to the degree to which the assessment technique or instrument produces the same range of results each time it is applied. Also refers to the assessment technique's ability to differentiate between participant's performance.
Assessment validity	refers to the degree to which the assessment ensures the knowledge, skill, ability, or achievement it is designed to measure.
Performance criteria	refers to the range or list of activities which must be demonstrated or knowledge which must be shown in order to judge the individual learning exercise adequate.
Modular curriculum	a series of courses of instruction or classes, organized

	in a manner which leads to an ultimate or eventual learning experience that includes all of the necessary elements of instruction needed to perform a task, set of tasks , or activities. Presumes levels of training, from basic to advanced, and the accumulation of credits or abilities.
Accreditation of prior learning	the determination or ascertaining of knowledge, skills, and abilities the learner brings into the training initiative from prior experience or prior instruction.
Standards	the set of criteria or elements which have been determined, by whatever process, to be necessary for competency.
Competency	Knowledge, skills, and abilities which, together, account for the ability to deliver a specified professional service.[6]

Kern et al.[7] describes several methods of evaluating a professional training initiative:

Method of Evaluation	**Advantages**	**Disadvantages**
Rating forms	Convenient Inexpensive	Subjective Contains rater biases
Self-assessment forms	Economical Useful for formative evaluation	Subjective Rater biases Limited use for summative evaluation
Essays on trainee's experience	Efficient Qualitative information Formative	Subjective Varies and unreliable
Written or computer-interactive tests; Questionnaires	Standard Methodological rigor Quantitative Summative	Reliability/validity varies Not qualitative Time consuming
Oral Examinations or Individual interviews	Flexible Informal Learner centered Formative or Summative	Time consuming Subjective Does not assess performance
Group interviews or Group discussions	Efficient Flexible Respondent centered Rich qualitative information	Time consuming Subjective Requires high level of skill to facilitate Not quantitative

Direct observation	Unobtrusive Assesses performance Methodological Rigor	Requires performance standards Personnel intensive
Exercises or Performance Audits	Objective and Realistic Unobtrusive Qualitative and Quantitative Reliable and valid if performance measures exist	Requires skilled observers Time consuming Expensive

Pratt[8] described three distinct types of information which can be collected in an assessment of training:

> *presage variables* - those data which have to do with the quality of the faculty, students, historical elements, and training resources;
>
> *process variables* - the data addressing the progress and process of instruction, frequency of activities, attendance of participants, rates of use of resources; and,
>
> *product variables* - data showing the impact or results of the instruction on the accomplishment of tasks, the effectiveness of training, and the diminishing of problems for which the training is designed to ameliorate.

Pratt states succinctly, "the ultimate test of the quality of training is the impact the trained person has on some unknown future situation." This statement suggests, in no uncertain terms, that the *product* of the training must be measured, not just the process. The product or ultimate change may be measured in actual events or through change in the organization or it may be measured by proxy through exercises. This is consistent with the program evaluation methodologies described earlier.

The purpose of the other two types of information which go into an assessment is to determine, to the degree possible, if the training is progressing as intended. Again, the ultimate assessment is the product but the intermediate assessments are the quality of the instruction and the process of the instruction. If it appears, based on outcome or product variables, that the training is not meeting the goals and objectives, the presage or process variables may be altered to reformulate and renorm the training to better accommodate the objectives. Further evaluation determines whether the changes have influenced the product variables enough or in the appropriate direction.

Too often, evaluation becomes a process assessment rather than a product assessment. This is probably due to the fact that process is easier to assess than outcome. In a training program we can easily "measure" or count the number of courses offered, the number of participants, the hours of training, the frequency of repetition, and the like. These are process elements or variables which show that the program was conducted and to whom it was available. It does not help us understand the degree to which the program was effective in addressing the goals and objectives which formed

the basis for the development of the program.

Ecclestone[9] provides us with an interesting approach to assessing a training program. She identifies four distinct points in the process of training where assessment is important:

Initial Guidance	Assessment at this stage "enables the learner to make choices based on clear information about options and own abilities." Providing potential participants with information on the courses, classes, modules, objectives, competencies and performance objectives, would allow the participant or their agencies to determine the applicability and attractiveness of the course or program to their needs, skills and abilities. Without such information, the wrong people may be the training classes or those who need the training and who might benefit the most would not opt in. Assessing prior learning and individual needs or starting points is most attractive in an individualized program and may not be applicable to standardized programs. The types of assessment at this stage are typically *formative* and may be criterion-referenced or ipsative.
Admissions	makes decisions about entry to a program or the appropriate level of entry into the program. Exemptions and eligibility should be based on articulated criteria or standards. The admissions process seeks to establish homogeneity of knowledge, skills, and abilities within classes so that the instructional process can be more focused and more effective. They types of assessment at this stage are typically summative and criterion-referenced.
In-programme	records the progress of individuals or groups (if that is the focus of the training) based on the process plan, needs, and targeted intermediate instructional goals. This assessment would rely on *enabling objectives* as were described earlier. Often this assessment is formative and criterion-referenced although it can also be ipsative. If the program or course lends itself to comparisons of participants, the assessment could also include some norm-referenced comparisons.
Certification	this assessment is used to confirm achievement. The assessment is most often summative and is based on *terminal objectives*. Criterion-referenced assessments, using performance-based or competency-based instruction, are the most reliable and valid methods for this type of evaluation.

While there may be some concern with the use of the term "certification" by Ecclestone,[10] she uses it to imply some external, validated method of attesting to the inculcation of the material, knowledge, skills, and abilities intended in the instruction.

There are three critical loci of evaluation or assessment in any training or continuing education programs:

Program Assessment

Instructional Assessment

Participant Assessment

Each of these loci are important and all must be included in a creditable program. We will discuss each of these critical types of assessment and provide recommendations on the types of assessments which are appropriate for each.

Program Assessment

Program assessment may be holistic and include the entire program or initiative. It may also be more focused and address each course or category of offering. Program assessment is sometimes called "curriculum evaluation." Oliva[11] suggests that a curriculum or program define objectives which are to be "specific, measurable, programmatic statements of outcomes to be achieved by students as a group in the school or school system." These objectives are different from Bloom's Educational Objectives which address the level of cognition which should be or is targeted. Oliva defines a curriculum goal as "a purpose or end stated in general terms without criteria of achievement."[12] This goal statement may include some of the terms and phrases in Bloom's taxonomy but they would be applied in the broadest sense. For example, a program or initiative might have several goals such as:

Increase the targeted workers' awareness (knowledge) of a particular phenomenon;

Improve assessment methods of personnel reacting to a particular type of problem;

Enhance understanding of the general public for the importance of an issue.

The assessment of a program's efforts to accommodate such broadly worded statements is almost always subjective but the subjective assessment should be justified and the justification should be articulated. The assessment should have points of evidence or proof that the assessment is appropriate.

In addition to curriculum goals, curriculum objectives must be developed early in the process. These objectives should be measurable and may include some of the same terms used in goals but the objectives are stated in more specific terms which lend themselves to evaluation and assessment. Examples of curriculum objectives would be:

Ten percent of emergency department personnel will be trained in triage procedures (assessment) each year in the target cities/hospitals;

Every state will have at least five persons trained to develop state-specific reaction (application) strategies for emergency events.

The objectives should be refinements of the broader curriculum goals. Generally, the curriculum objectives should be stated in performance or behavioral terms - the knowledge, skills, and abilities which the participants are expected to demonstrate in the abstract or broadest terms. A regional or local training initiative is far easier to assess than a national initiative. Still, there are "guiding principles" which can be assessed. Oliva identifies questions which should be addressed in a curriculum assessment or evaluation:

- Is the scope of the curriculum adequate?
- Is the scope of the curriculum realistic?
- Is the curriculum relevant?
- Is there balance in the curriculum?
- Is curriculum integration desirable?
- Is the curriculum properly sequenced?
- Is there continuity of programs?
- Are curricula and courses well articulated between levels?
- Are types of learning transferable?

The answers to these questions, as well as others which can be developed for a particular type of training, can help to restructure the curriculum, the courses, and the levels of instruction. Additionally, the needs and issues will change over time and this change must be accommodated in the curricular change. Developing a curriculum is not a once-only activity but, through evaluation or assessment, the curriculum can remain appropriate and relevant.

In many ways, assessment is more important than the initial development of a curriculum. Mistakes will be made in some aspects of the development and design of a curriculum. These mistakes are assumed to be miscalculations or unanticipated consequences which can be repaired in the evaluation or assessment phase. The assessment of curriculum is important enough to merit attention to each of the questions or issues which come into play in the curriculum assessment.

Is the scope of the curriculum adequate? This issue addresses the breadth of the curriculum. It is inconceivable that all elements in the subject matter can be anticipated in the developmental phase of the curriculum. Planners should, nonetheless, attempt to select all of the knowledge, skills, and abilities to be addressed in the curriculum. The planning process should then address all of the courses, classes and instruction which accommodates the list of knowledge, skills, and abilities. For a curriculum to be a curriculum, it should link the topics based on common threads. Periodically, through the assessment process, the commonality issues must be addressed, as well as the changes which have occurred in the discipline or on the topic. With developments in knowledge and technology, it is very likely that the breadth of a curriculum will need to be expanded.

Is the scope of the curriculum realistic? Just as important as enriching the scope of a

curriculum is the examination of the curriculum to be certain that the breadth is not too ambitious. Modules of instruction and courses are linked based on a defensible logic. If that logic is stretched and a curriculum becomes unrealistic, the credibility of the program suffers. A training or educational program must be realistic in scope if it is to be taken seriously and has credibility.

Is the curriculum relevant? Just as scope is subject to change, so is the relevance of portions of the curriculum. Curriculum, when designed, is likely to be historically relevant but as it progresses, it is critical that it maintain contemporary relevance.

Is there balance in the curriculum? Halverson[13] states "curriculum balance will probably always be lacking because institutions of all kinds are slow in adapting to new needs and demands of the culture except when social change is rapid and urgent in its implications for these institutions." Often the issue of balance is seen as a series of dichotomies which must be "balanced." These include general versus specialized courses, individualization versus mass education, innovation versus tradition, and immediate versus the remote. Balance between disciplines, courses or modules, as well as within the components should also be inspected. Emphases are perceived based on imbalances in the curriculum. For example, if more courses are offered in one particular area or on a topic, it is presumed to be the emphasis of the program.

Is curriculum integration desirable? Integration addresses the blending of courses, modules or parts of the curriculum. While this may seem unnecessary since courses are, by their nature, autonomous. For courses or modules to be parts of a "curriculum" they must be related in some fashion. Some curriculum specialists refer to this integration as "correlation" or judging the relationship of courses while maintaining their separateness. Subjects, courses, and modules can be integrated horizontally or vertically. The vertical orientation or integration is similar to the "curriculum spiral" described earlier. Reassessing the interrelationship of the parts of the curriculum is important.

Is the curriculum properly sequenced? Assessing the sequencing of courses is important but it is unlikely that serious changes would occur from the design of the curriculum, if done properly, to the subsequent assessment or evaluation of the curriculum. Courses or components of a curriculum can be sequenced from the least complex to the most complex, as was described earlier. The curriculum can also be sequenced chronologically, geographically, reverse chronology, or from general to the particular.

Is there continuity of programs? Also consistent with the spiral curriculum, it is important to ascertain that concepts and skills are introduced early in the process and reintroduced in order to reinforce and enhance the exposures. It is important to make certain that the reintroduction is programed and intended, not simply repeated due to any lack of planning and oversight.

Are curricula and courses well articulated between levels? Continuity and articulation are related concepts. Curricula are articulated if the relationship between levels, courses, and

modules is according to the plans and designs of those who developed the programs. In addition to the articulation of the courses and components of the curriculum, it is important to examine and assess the articulation of the participants as to their selection, inclusion, exclusion, and progress.

Are types of learning transferable? It is very, very important that the information contained in the curriculum be the type of information which is most useful to the participants. This is the essence of the training and educational process but it must be reaffirmed through the assessment process. The information imparted must have some value outside the training process. "Transfer of cognitive learning is most often visible in student performance on assessment and standardized tests... and in the evaluations employers give of the intellectual competence of their employees."[14] We will discuss some of this in the section below on evaluation of participants. From the curricular standpoint, however, we must make certain that the information has a transferability which supports the proposition that the information is valuable.

All eight concepts or issues are interrelated. This is shown very clearly in the evaluation model proposed by Saylor, Alexander, and Lewis[15] where they divide the evaluation process into an assessment of:

- Goals, Subgoals, and Objectives
- Program of Education as a Totality
- Specific Segments of the Education Program
- Instruction
- Evaluation Program

Each of these parts of the model are divided into formative assessments and summative assessments. We will not belabor the point by describing each of these steps because there would be a fair amount of duplication with the model just described. Important here is the fact that evaluating the curriculum includes evaluating the evaluation process as well.

Instructional Assessment

Often problematic is the need to assess the level and quality of the instruction. This is different from an assessment of the program and much more specific to the process of delivering the instruction. The classic approach would define "instructional evaluation" as "evaluation of instruction through the assessment of student achievement."[16] This approach will be discussed in a later section but it clearly indicates the relationship between instruction and the results of instruction - learning.

There are actually two aspects of assessment of the instructional component. There is the assessment or evaluation of the instructors and the techniques, process, and materials used by the instructor. The other aspect of instructional assessment is the evaluation of the participants during the instruction, not simply after the instruction is over. Each of these will be addressed separately.

Assessing Instructors. Instructors may be evaluated using any of three methods:

Participant survey: This, the most traditional and widely used technique, is a cost-effective, efficient method of assessing instruction by those who have observed the greatest portion of that instruction - the participants. Participant surveys are generally applied at the end of a course or class. "Most evaluations of Continuing Education programs are administered at the end of the program offering and usually consist of a subjective rating of how the customer felt about the learning experience."[17] The questions generally cover topics such as the preparation of the instructor, the knowledge of the instructor, the enthusiasm of the instructor, the selection of instructional materials, value of instruction, and degree to which instruction was beneficial. Vernon Bryant provides twelve questions which ought to be included in such a survey:

1. To what extent did the instructor expose you to new possibilities and self-growth?
2. To what extent did the instructor help you to clarify your desire to improve your skills?
3. To what extent did the instructor aid you in the diagnosis of the gaps between your aspiration level and your present level of performance?
4. To what extent did the instructor help you respect your own feelings and ideas?
5. To what extent did you feel mutual trust and helpfulness among students?
6. To what extent did you feel a spirit of mutual inquiry between yourself and the instructor?
7. To what extent did you feel there was a mutual process of setting learning objectives?
8. To what extent were you able to share your thinking about the options in designing learning experiences, selection of materials, and the methods of instruction?
9. To what extent did the instructor help you to organize a learning-teaching environment in which the responsibility for the process of inquiry (learning-teaching teams, task groups, independent study, etc.) was shared?
10. To what extent did the instructor draw on your own experiences as resources for one another' learning?
11. To what extent did the instructor gear the presentations to your level of experience?
12. To what extent did the instructor involve you in mutually acceptable criteria for measuring your progress toward learning objectives?[18]

The scale recommended is a Likert, five-point scale ranging from "Not at all" to "Very

much." This method is useful and one very much like it is repeated in almost every course in every college and university in the United States. This method is often used in continuing education where participant approval is critical for the future of a program. Complicating the picture of effectiveness and efficiency for this approach, is the fact that multiple instructors is problematic. When a variety of instructors are used or when classes or sections are "team-taught," the survey results may be measuring what they are intended to measure or something else. The validity problems may be overpowering and negate the desire to use this simple assessment tool.

Self-evaluation. Requiring instructors to evaluate their own effectiveness is actually more useful than some would imagine. Instructors, particularly those who hold certification as instructors and/or advanced degrees, understand the expectations of the process and the degree tow which they meet those expectations. The same questions used in the end-of-course survey are appropriate for a self-evaluation assessment.

Direct observation. A time-tested method of assessing instruction is to observe random portions of the instruction. Again, the core questions used in the survey are appropriate as the basis for an observational assessment.

In addition to these methods, a passive method could be employed regardless of other methods of assessing instruction. This passive method would involve the examination and evaluation of instructional materials, including syllabi, handouts, and presentation files.

The assessment of instruction, as a component of the learning process, is appropriately termed "formative evaluation." Oliva defined formative evaluation as the "formal and informal techniques, including testing, that are used during the period of instruction."[19] This type of evaluation can be considered "progress" evaluation for the participant and "process" assessment for the instructor. The monitoring of progress and process by the persons, agency, or organization overseeing the course or curriculum can provide information on the status of the course but can be a time-consuming and intensive activity.

Participant Assessment

As described earlier, this impact assessment is best accomplished using a "summative evaluation."

> Summative evaluation is the assessment that takes place at the end of a course or unit. A final written examination (post-test) is the most frequently used means of summative evaluation of instruction. Its major purpose is to find out whether the students have mastered the preceding instruction.[20]

The two types of "measurement" of participants' performance are norm-referenced and criterion-referenced measurements. Norm-referenced assessments measure the achievement of one participant against or in relationship to all other participants in the course, class, or program.

There are specific reasons for using norm-referenced measurements and advantages to that

assessment. These are succinctly stated below:

1. The main function of norm-referenced measurement is to ascertain the student's relative position within a normative group.
2. Either general conceptual outcomes (usually done) or precise objectives may be specified when constructing norm-referenced measurement.
3. The criterion for mastery is not usually specified when using norm-referenced measurement.
4. Test items for norm-referenced measurement are constructed to discriminate among students.
5. Variability of scores is desirable as an aid to meaningful interpretation.
6. The test results from norm-referenced measurement are amenable to transportation to the traditional grading system.[21]

Norm-referenced assessments are often easier since they do not require any or much preliminary work in developing objectives or standards yet they still provided each participant with his or her standing relative to others.

Criterion-referenced assessments measure the participant's achievements against a predefined "standard" or criteria. The use of the term standard is used guardedly because it may suggest that there must be a universal or widely accepted performance level. While that may be ideal and it may occur in some disciplines, it is not necessary for criterion-referenced assessment to occur. The criteria may be the learning objectives formulated prior to the course or the behavioral objectives prepared when the course was designed.

One distinct advantage of the criterion-referenced assessment approach is its ability to influence the future development of the curriculum. As Popham stated, "norm-referenced measures permit comparisons among people" while "criterion-referenced tests make decisions both about individuals and treatments."[22]

The advantages of criterion-referenced assessment may be apparent, but it is still useful to list them specifically:

1. The main function of criterion-referenced measurement is to assess whether the student has mastered a specific criterion or performance standard.
2. Complete behavioral objectives (i.e., planning objectives) are specified when constructing criterion-referenced measurements.
3. The criterion for mastery must be stated (i.e., planning objectives) for use in criterion-referenced measurement.
4. Test items for criterion-referenced measurement are constructed to measure a predetermined level of proficiency.
5. Variability is irrelevant; it is not a necessary condition for a satisfactory criterion-

referenced measurement.

6. The results from criterion-referenced measurement suggest the use of a binary system of measurement (i.e., satisfactory-unsatisfactory; pass-fail).[23]

Competency-based Assessment

Either of these two assessment methods can and are used to assess competencies and performance. Competencies, however, suggest the presence of objective criterion so a criterion-referenced assessment is most consistent with that approach. As defined earlier, competency can be defined as the knowledge, skills, and abilities which, together, account for the ability to deliver a specified professional service. As stated by Finch and Crunkilton, a competency is a critical aspect of the work, duties, or responsibilities. It "evolves from explicit statements of worker roles" and include "specific criteria ... that clarify each competency."[24] Competency-based instruction, described earlier, involves the determination of objectives, describing the objectives in terms of criteria or competencies, and assessing the participant's progress, relative to the criterion or competencies. "Instructional staff are required to move beyond the traditional knowledge type measures such as multiple-choice and essay examinations and focus on assessment that aligns with competence in the real world."[25] A set of competencies or criterion, associated with a complex activity or set of activities, can be considered a "competency profile" and may provide, not only the criterion against which a person will be assessed but also the modules necessary to accumulate the competencies.

Performance-based Assessment

Closely related to competency-based assessment is performance-based assessment. In the brief section above, "performance" was mentioned frequently. Often in the literature, "performance-based" is used in describing an approach to designing a curriculum. It was in this way that Pucel described the process of curricular design that culminated with the evaluation or measurement of performance. Kern, et al., in describing medical curricular development, use the term "competency objectives" as a synonym for skills. They describe the training process to achieve psychomotor objectives as beginning with supervised experiences, moving to simulations and culminating with a review of the skills and experiences. In this framework, assessment and evaluation are constant processes, interwoven into the training process itself. They define "simulations" as "clinical situations" where learners can "practice skills in a 'safe' environment where risks can be taken and mistakes made without harm."[26] The three types of simulations used in medical instruction are:

Artificial models	Inanimate devices designed to simulate real clinical situations.
Role-playing	The learner has an opportunity to try, observe, and discuss alternative techniques until satisfactory performance has been achieved.
Simulated patients	This technique ensures that content area will be covered, new techniques attempted, and performance achieved with live, simulated patients who play their role as patients.

Each of these methods are efficient and allow practice as well as instruction. The last of the

methods, the simulated patients, has been found to be both efficient and effective and has less artificiality than the others.[27] This method becomes important later as we discuss "exercises" or simulations.

An approach recently described in law enforcement training involves "authentic assessment." Authentic assessment is defined as "the process of evaluating a trainee's performance on the basis of the trainee's demonstrated knowledge, skills, and abilities."[28] This assessment approach is clearly "performance-based" and the article supports the use of "portfolios" as assessment tools to measure performance of authentic or real life tasks. Offered as an alternative to traditional testing approaches, "authentic assessment" is a method for measuring performance rather than "exposure" to information. Precise descriptions of knowledge, skills, and abilities is not described as an essential element but "standards" are mentioned throughout. Of importance here is the development but instructions of "task-oriented, job-related scenarios, which reflect some of the most common (knowledge, skills, abilities, and other characteristics) necessary in the field in which recruits participate." Implicitly, there must be a set of criteria or objectives which must be met. The articles states "trainers evaluate the recruits' performance based on a specific task list developed for that particular situation." and the assessment models or "rubrics" indicate the presence of specificity in tasks and objectives:

5= The work is superior in most respects, including:
- Solicit and accept criticism in order to improve performance and act on same.
- Perform optimally under stress and non-stress conditions, acting decisively and properly.
- Exhibit mastery of officer safely tactics in all situations without becoming over-confident or paranoid.
- Demonstrate superior ability to listen, and comprehend written and verbal instructions. Respond appropriately when speaking person-to-person, on the telephone, or on the radio.
- Establish a rapport with public; remain objective and at ease with individuals, eliciting a positive public response.
- Work performance would be approved by a supervisor.

4= The work performance is very good in most areas listed above.
- Performance is competent and of high quality.
- One or more areas may be superior.
- Performance would be approved by a supervisor.

3= The performance is satisfactory in most areas listed above.
- Performance is competent.
- Skills meet at least the minimum criteria or better.
- One or more areas may be of good quality.
- Performance would be approved by a supervisor.

2= The work performance needs improvement for acceptability.
- Some skills lack the minimum criteria for acceptability.

- One or more areas may be of good quality.
- Performance would not be approved by a supervisor.

1= The work performance does not meet minimal criteria for acceptability.
- One or more areas may demonstrate the minimal criteria for acceptance.
- Some remedial work is warranted in the areas listed above.
- Performance would not be approved by a supervisor.

0= The work performance does not meet the minimal criteria for acceptability.
- Multiple skill areas listed above are inadequate.
- Performance is incomplete.
- Remedial work is warranted in the areas listed above.
- Performance would not be approved by a supervisor.[29]

Recognizing the exigencies of reducing a description of an initiative to a brief article, it is evident that many of the details and specific criteria and elements are omitted in this rating or assessment system. Also clear, however, is the prominence of "performance" as the key product in the curriculum and, therefore, in the assessment system. This is quite consistent with performance-based assessment and performance-based curricula. The result of the training is not a catalogue of unrelated tasks, each of which is evaluated, but it is set of activities which are linked and can be considered to contribute to the performance of an activity.

As said in an earlier chapter of this document, performance-based training can be very effective for group activities as well individual activity and performance. Competency-based training is most appropriate for individual assessments but not for group or team activities.

Simulations or Exercises

The use of live, realistic exercises or simulations for training is well accepted in most disciplines. Military science has used the technique to simulate battle field situations at both a training tool and an evaluation tool. The process itself has great educational value but can best be considered an adjunct to other training or educational approaches.

In 1996 the Pentagon hosted the largest parachute assault since World War II. This exercise, part of a joint U.S. and British training initiative which involved 53,000 troops, 5,000 paratroopers, 144 heavy-lift aircraft, and hundreds of pieces of heavy equipment and weaponry, was a remarkable departure from what had become "standard" training. Even though this particular exercise cost more than $17 million, it was considered necessary by the Department of Defense. It tested the combined organizations' ability to distribute information around the battle-field, the efficacy of logistical support, and the interaction of disparate units, agencies and machinery. This exercise was the ultimate test of training and was viewed as necessary by the commanders. "We've been doing all of this through simulation but we need to test the theory with practice."[30] The practice of testing capabilities in the field is a long-standing one in the military. Some might suggest that the practice is such an accepted one that its continued use is self-serving. It should be noted, however, that the exercise described above was a break from what had become the traditional method of field-testing,

computer simulation. One General noted after the exercise that some of the problems and issues observed in the live exercise, could never have been recognized in computer simulations, even though there has been a shift in that direction since 1986. An assessment by Rand of the Department of Defense computer simulation initiatives points to the need to conduct computer simulations "in the face of projected reductions in manpower and budget" and what was seen as a growing intolerance by the public of the noise of "low-flying aircraft and armored vehicles." The primacy and efficacy of field exercises was affirmed, however.[31]

Other disciplines have also recognized the value of live exercises for training and evaluation. A recent article in the Archives of Pediatrics and Adolescent Medicine describes training simulations which teach physicians to deliver near-fatal diagnoses of children to the parents:

> Using volunteers trained in role playing and communication to act as parents, seven pediatric intensive care fellows delivered a near-fatal diagnosis of a childhood illness. The mock parents then instructed the physicians on their communication skills, support, and their own perceptions. The physicians then repeated the process with the new "parents." Videotaped sessions of the interactions were graded, and physicians were significantly more effective at delivering the bad news in the second session, compared to the first.[32]

This approach is a small-scale exercise or simulation, conducted in a "least-harm" environment. Medical educators have used standardized patients for at least the last three decades.[33] They offer the same advantages of large scale exercises - realism, feedback, and controlled environment. Law, in addition to medicine, has a long history in simulations or exercises. Moot Court is a well accepted educational tool which also has an evaluative component.

Key to the effectiveness of an exercise, no matter how large or small, is the inclusion of feedback or "lessons learned." Exercises can be viewed as the last, most realistic training module and the one in which the participant or agency is expected to operationalize the information gained in other training modules. The assessment of the effectiveness of the exercise is one distinct attribute of the exercise but, at least as important, is the weaving back into the process the mistakes and successes of the training. Often termed "Lessons Learned," this information serves as the most obvious example of the feedback loop available.

Objective (Bloom's Level of Cognition)	Tests	Observation	Exercises
Knowledge	XXX	XXX	
Knowledge of Specifics	XXX	XXX	
Knowledge of ways to deal with Specifics	XX	XXX	
Knowledge of Principles and theories	XX	XXX	
Comprehension	X	XXX	
Translation	X	XXX	
Interpretation	X	XXX	
Extrapolation	X	X	
Application	XXX	XX	XXX

Analysis	XXX	XX	XXX
Analysis of Elements	XXX	XX	XXX
Analysis of Relationships	XXX	XX	XXX
Analysis of Organizational principles		XX	XXX
Synthesis	XXX	XX	XXX
Evaluation	XXX	XX	XXX

X's indicate the perceived strength of the approach at the level of cognition

While we suggest no particular method of evaluation, we do suggest that the literature demands that evaluation is essential to any instructional program. The matrix above reflects, in two dimensions, the methods viewed as most effective for certain levels of the taxonomy of educational objectives.

Summary

In this section we have provided a great deal of information on the maintenance of a training or educational program. All credible literature on the topic of curriculum development and instructional delivery stress the need to assess the curriculum, the instruction, and the participants. This quality control is a reasonable and responsible element of a curriculum.

Assessment serves many purposes. Oliva[34] describes several issues appropriate for consideration in assessment:

- Is the scope of the curriculum adequate?
- Is the scope of the curriculum realistic?
- Is the curriculum relevant?
- Is there balance in the curriculum?
- Is curriculum integration desirable?
- Is the curriculum properly sequenced?
- Is there continuity of programs?
- Are curricula and courses well articulated between levels?
- Are types of learning transferable?

Each of these question can be answered by assessing the program, the instruction, and the participants. Program assessment can be accomplished through various means, including expert panels and some of the same methods used to develop the curriculum. Instruction can be assessed through surveys, observations, and, again, expert assessment of the scope, relevance, and adequacy of the syllabi, materials, and instructional skills. Participants can be assessed using norm-referenced measures, or criterion-referenced measures. The clear preference in the literature is for criterion-referenced measures of assessment.

The assessment and quality control measures feed back in to the curricular revisions and refinement. If the curriculum and instruction is not accomplishing what is designed to do, it should be altered. The only way to determine that is to assess.

The end of this section included a discussion of the most sophisticated method of evaluating

performance - the exercise. This method is a mainstay of institutions such as the military where simulations are useful but not the penultimate tests of effectiveness. Ultimate tests are actual applications and cannot be contrived or conducted without dire consequences. The penultimate test, however, is an exercise or simulation which is a realistic as possible. While it has clear training and instructional value, it presumes a great deal of knowledge, skills, and abilities prior to the event. It is, therefore, more of an evaluation or assessment of the abilities, performance, readiness, or capacity to accomplish the goals. Like any other penultimate assessment, it is not to be employed frequently. The other methods of assessment can and should be employed as frequently as possible to refine the curriculum to meet the desired goals and objectives.

Notes to Chapter Six

1. Bloom, Benjamin S., Max D. Engelhart, Edward J. Furst, Walter H. Hill, and David R. Krathwohl. (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals, Handbook 1, Cognitive Domain. New York: David McKay, p. 2.

2. Maxfield, Michael G. and Earl Babbie. (1995). Research Methods for Criminal Justice and Criminology. Boston: Wadsworth, p. 307-309.

3. Ibid.

4. Maxfield and Babbie, op. Cit., p. 305.

5. Maxfield and Babbie, op. Cit., p. 308.

6. McGaghie, William C., George E. Miller, Abdul W. Sajid, and Thomas V. Telder. (1978) Competency-based Curriculum Development in Medical Education. Geneva: World Health Organization. P. 19.

7. Kern, David E., Patricia A. Thomas, Donna M. Howard, and Eric B. Bass. (1998). Curriculum Development for Medical Education: A Six-step Approach. Baltimore: Johns Hopkins.

8. Pratt, Clara Collette. (1979). "Evaluating the Impact of Training on the World of Practice." In Evaluation of Continuing Education for Professionals: A Systems View. Eds. LeBreton, Preston P., Vernon E. Bryant, Douglas L. Zweizig, Dan G. Middaugh, Anita Gras, Bryant, Tricia C. Corbett. Seattle: University of Washington, p. 350.

9. Ecclestone, Katheryn. (1996). How to Assess the Vocational Curriculum. London: Kogan Page, p. 20-68.

10. Ecclestone, Katheryn. (1996). How to Assess the Vocational Curriculum. London: Kogan Page.

11. Oliva, Peter F. (1997). Developing the Curriculum, 4th Ed. New York: Addison Wesley Longman.,p. 438

12. Oliva, Op cit., p. 236, 439

13. Halverson, Paul M. (1961). "The Meaning of Balance." Balance in the Curriculum: 1961 Yearbook. Alexandria, VA: Association for Supervision and Curriculum Development, p. 7

14. Oliva, Peter F. (1997). Developing the Curriculum, 4th Ed. New York: Addison Wesley Longman, p. 458.

15. Saylor, J. Galen, William M. Alexander, and Arthur J. Lewis. (1981). Curriculum Planning for Better Teaching and Learning, 4th Ed. New York: Holt, Rinehart and Winston.

16. Oliva, Peter F. (1997). Developing the Curriculum, 4th Ed. New York: Addison Wesley Longman, p. 399.

17. Bryant, Anita Gras (1979) “A System of Interaction Analysis as a Means of Evaluating a Continuing Education Program at the Time of Implementation” In LeBreton, Preston P., Vernon E. Bryant, Douglas L. Zweizig, Dan G. Middaugh, Anita Gras Bryant, Tricia C. Corbett (Eds.) The Evaluation of Continuing Education for Professionals: A Systems View. Seattle: The University of Washington, pp. 316-325.

18. Bryant, Vernon E. (1979) “An Evaluation of Continuing Education Programs Based on The Principles of Adult Learning” In LeBreton, Preston P., Vernon E. Bryant, Douglas L. Zweizig, Dan G. Middaugh, Anita Gras Bryant, Tricia C. Corbett (Eds.) The Evaluation of Continuing Education for Professionals: A Systems View. Seattle: The University of Washington, Pp. 328-329..

19. Oliva, Peter F. (1997). Developing the Curriculum, 4th Ed. New York: Addison Wesley Longman, p. 404.

20. Ibid.

21. Smythe, Mary Jeanette, Robert J. Kibler, and Patricia W. Hutchings. (1973) "A Comparison of Norm-Referenced and Criterion-Referenced Measurements with Implications for Communication Instruction." The Speech Teacher Vol 22(1): 4.

22. Popham, W. James. (1973) Evaluating Instruction. Englewood Cliffs, N.J.: Prentice:Hall, pp. 25-26.

23. Smythe, Mary Jeanette, Robert J. Kibler, and Patricia W. Hutchings. (1973) "A Comparison of Norm-Referenced and Criterion-Referenced Measurements with Implications for Communication Instruction." The Speech Teacher Vol 22(1) p. 4.

24. Finch, Curtis R. and John R. Crunkilton. (1999). Curriculum Development in Vocational and Technical Education: Planning, Content, and Implementation. Boston: Allyn Bacon, p. 260.

25. Ibid.

26. Kern, David E., Patricia A. Thomas, Donna M. Howard, and Eric B. Bass. (1998). Curriculum Development for Medical Education: A Six-step Approach. Baltimore: Johns Hopkins, p. 48.

27. Kern et al., loc cit; Stillman, P.L., D. Swanson and B. Regan. (1991) Assessment of Clinical Skills of Residents Utilizing Standardized Patients." Annals of Internal Medicine. Vol. 114. P. 393-401.

28. Thermer, Clifford E. (1997). "Authentic Assessment for Performance-based Police Training." Police Forum, Vol. 7(3), p. 2

29. Thermer, op. Cit., p. 5.

30. Constance, Paul. (1996). "Computer-simulated war games still can't replace the real thing." Government Computer News, May 26, 1996, Vol. 15(11): 1-2.

31. Allen, Patrick D. (1993) Evolution of Models at the Warrior Preparation Center: Problems and Solutions for Higher-Echelon Exercises. Santa Monica, CA: RAND.

32. Vaidya, Vinay, Larrie W. Greenberg, Kantilal M. Patel, Leslie H. Strauss, Murry M. Pollack. (1999). Teaching Physicians How To Break Bad News. Archives of Pediatrics and Adolescence Medicine, Vol 153(4): 419.

33. Baerheim, A., Malterud, K. (1995). "Simulated Patients for the Practical examination of Medical Students: Intentions, Procedures, and Experiences." Medical Education, Vol. 29: 410-413.

34. Oliva, Peter F. (1997). Developing the Curriculum, 4th Ed. New York: Addison Wesley Longman.