Chapter 1
Classroom Assessment: Introduction

In a "standards based" approach to education and training, informed by Constructivist ideology and motivated by high stakes accountability, assessment informed instruction is the expectation as is continuous improvement. Assessment informed instruction requires the educator (teacher, trainer, planner, instructional designer or administrator) to plan, deliver, and adjust instruction based on students' or trainees' evolving mastery of learning and skill standards until the desired mastery is achieved.

The Teaching/Assessment cycle is outlined in Figure 1.1. Based on learning standards, teaching is conducted. Once teaching is launched, continuous formative assessment is engaged as is re-teaching based on assessment results. The assessment/re-teaching cycle is repeated until suitable mastery is demonstrated via summative assessment. Then a new teaching/assessment cycle begins. The teaching/ assessment cycle assumes that instruction and assessment are planned and executed in conformance to specified learning and performance standards.

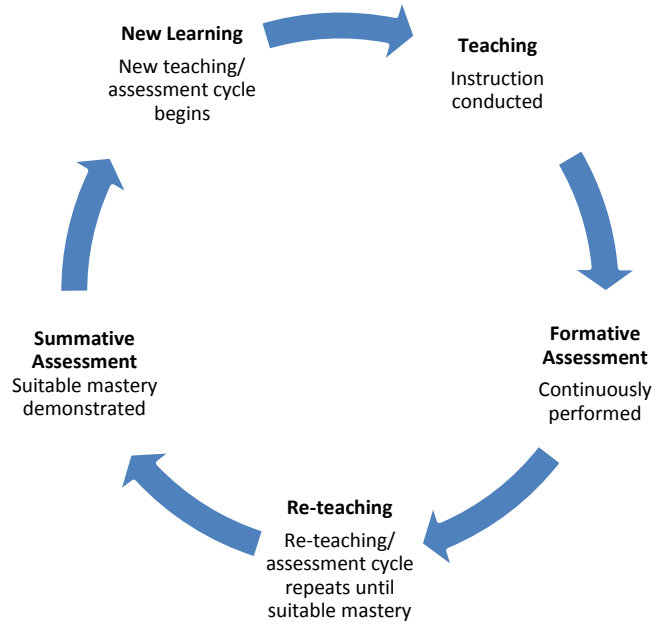

Figure 1.1 Teaching/Assessment Cycle

With the Teaching/Assessment Cycle as our application framework, in this chapter we will review the relationship between knowledge, learning, and intellectual skills. We will also examine the relationship between measurement, assessment, and evaluation.

I. Knowledge, Learning, Intellectual and Thinking Skills

A. Definition of Knowledge

1. Alexander (1996, p. 89) writes that knowledge "is a scaffold that supports the construction of all future learning." Greeno, Collins, & Resnick (1996, p. 16) argue that the cognitive view of knowledge "emphasizes understanding of concepts and theories in different subject matter domains [e.g., reading or science] and general cognitive abilities, such as reasoning, planning, solving problems, and comprehending language."
2. Knowledge can also be broadly categorized according to use, as declarative, procedural, or conditional (Paris & Cunningham, 1996; Paris, Lipson, & Wixson, 1993).
   (a) Farnham-Diggory (1994, p. 468) defined **declarative knowledge**, "knowledge that can be declared, usually in words, through lectures, books, writing, verbal exchange, Braille, sign language, mathematical notation, and so on." Declarative knowledge can be simple facts, generalities, rules, personal preferences, etc.
   (b) Woolfolk (2001, p. 242) defines **procedural knowledge** as "knowing how to do something such as divide fractions or clean a carburetor. Procedural knowledge must be demonstrated." Other examples of procedural knowledge include translating languages, classifying shapes, reading, or writing. In intellectual skills taxonomies proposed by Bloom, Engelhart, Frost, & Krathwohl (1956) and Gagne (1985), the levels beyond knowledge, are procedural knowledge.
   (c) Woolfolk (2001, p. 243) defines **conditional knowledge** as, "knowing when and why to apply…declarative and procedural knowledge." Conditional knowledge involves judgment. Examples of conditional knowledge include how to solve various math problems, when to skim or read for detail, when to change strategies when confronted with a perplexing problem, etc.

3. When measuring the effectiveness of instruction or an instructional program, the instructional designer or examiner may use traditional classroom testing strategies or direct performance assessments (Both are discussed in Chapter 5). Regardless of the strategy or mix of strategies selected, declarative knowledge, procedural knowledge, and/or conditional knowledge will be assessed. Thus, it is important to know which type of knowledge is being assessed in order to frame test items or construct direct performance assessments that will yield the information sought.
   (a) When examinee answers a multiple choice or true false item correctly, he or she displays declarative knowledge.
   (b) When an examinee answers an item which requires the listing of steps to bake cookies, according to a particular recipe, he or she displays procedural knowledge.
   (c) When an examinee is required to solve an algebraic expression, he or she displays conditional knowledge as he or she must know which mathematical or algebraic laws or procedures to apply to correctly solve

the problem. When the circumstances surrounding a business opportunity change, one must decide on whether to pursue the same strategy or change the strategy in order to win a business contract. Remember, conditional knowledge requires judgment and relies on declarative and/or procedural knowledge.

B. Learning

1. The process of acquiring knowledge (declarative, procedural, and/or conditional) is called learning. In order to assess instructional effectiveness, learning must be measured. Kimble defined learning as "[A] relatively permanent change in behavioral potentiality that occurs as a result of reinforced practice" 1961, p. 6). In other words, a learner must display his or her knowledge through behavior (e.g., answering a test item, repairing a car engine or modeling a particular attitude). Hergenhahn and Olson (1997, p. 2) have pointed out
   a. Leaning must be exhibited through behavior.
   b. Learning is a consequence of experience (e.g., life, schooling, training, practice, observation, etc.).
   c. Only reinforced (positively or negatively) experience, practice, etc. is learned. Reward is only one type of reinforcement.

2. Instructional design and teaching strive to provide experiences and reinforcement which enables one to learn, i.e., resulting in a permanent (or the realistic potential for) behavior change. It is this behavior that is measured and then based on that measurement, inferences are made about what has been learned, how well it has been learned, and how competently it may be applied. For example:
   a. The first grade student, who does not know how to read, learns to read. This is a permanent change in behavior. The teacher knows the student can read because the student read a story.
   b. The worker, who lost her job due to changes in technology, learns new job content and skills by going to a vocational-technical school. An employer can determine whether or not the applicant can repair small engines, by watching her diagnose and repair a broken lawnmower motor.
   c. A college freshman can demonstrate his knowledge of history by correctly answering several test items on causes of the Great Depression.

3. So one may ask, "How do we measure learning?"
   a. An instructional designer or teacher specifies the knowledge (declarative, procedural, and/or conditional); skills; or attitudes (KSA's) which need to be learned.
      (1) These KSA's are then expressed as learning targets, learning outcomes, learning standards, or learning objectives. (These terms mean the same thing.)

(2) Next, a learning target is "broken down" into its component parts (often called benchmarks) which a learner must know in order to achieve or master the learning target. So, when a learner is able to accomplish all benchmarks, we infer that he or she has achieved or mastered the learning target. See chapter 5 for more detail.

b. Once the learning targets and benchmarks are written, instructional materials are identified and sorted into modules or units. This sorting is an iterative process that often leads to changing learning targets, benchmarks, and/or instructional materials so that a "cleaner" alignment is achieved.

c. Specific instructional strategies are devised to facilitate learning.

d. Since knowledge must be measured through learner behavior (e.g., answering test items, writing a paper, or producing a work product), formative and summative assessments are devised and administered to learners in order to measure teaching or instructional design effectiveness and student learning.

(1) Based on formative assessment results, instruction may be adjusted to assist those who are not learning as intended or to accelerate learning if learners exhibit mastery faster than anticipated.

(2) At the conclusion of the learning experience, learners are usually given a summative assessment (e.g., test) to measure their learning. From these summative results, inferences are made about the effectiveness of the instructional design of the curriculum, instruction, and learning.

4. The most critical component of a formative or summative assessment are the test items to which students or examinees must respond. Test items are written to match learning target benchmarks. If the examinee achieved (or mastered) the learning target benchmark, he or she achieved the benchmark.

a. A test or instructional designer must be able to identify the mental (i.e., intellectual skills) a learner must possess in order to meet the learning target. Bloom and colleagues (1956) have has classified "knowledge" into six intellectual skills: knowledge, comprehension, application, analysis, synthesis, and evaluation; a detailed discussion follows. Each intellectual skill, written into a learning target benchmark, must be written into test items written to specifically measure or assess learner mastery of that benchmark.

b. A learning target benchmark may require an examinee to use a specific thinking skill to correctly answer its corresponding test item(s).

c. Test item writers must know knowledge type, the specific intellectual skill, and any specific thinking skill, an examinee must possess to answer the item correctly to show that the benchmark has been achieved.

C. Bloom, et al.'s Intellectual Skill Taxonomy

1. Bloom, et al.'s (1956) Taxonomy of Intellectual Skills

a. Bloom, et al. (1956, p. 201) defines **knowledge** (i.e., recalling or remembering) to include the recall of facts, methods, processes, patterns, structures, settings, etc.

(1) Knowledge is stored in the brain; the purpose of measurement is to present a stimulus (test item) which will clue the examinee to recall the stored knowledge. Kubiszyn and Borich (1996, p. 60) say knowledge is what students must remember. Learners must have this declarative knowledge as it is the basis for all higher order intellectual skills.
(2) When writing learning standards (i.e., targets, outcomes, or objectives) at the "Knowledge" level, use action verbs such as: list, name, recall, state, underline, write, record, count, recite, draw, find, match, choose, label, remember, recognize, select, define, list, etc.

b. **Comprehension** (i.e., really "getting it" or understanding) is the lowest of the higher order intellectual skills in the taxonomy; but learners use the declarative knowledge largely within the context in which it was learned.
(1) Examinees are expected to
(a) Translate knowledge from one form to another without losing its essential meaning within the original learning context;
(b) Interpret knowledge so as to identify its central elements or ideas, and then make inferences, generalizations, or summaries but within its original context or application; or
(c) Use the knowledge to extrapolate trends, implications, consequences, etc., within the original learning context.
(2) When writing learning standards at the "Comprehension" level use action verbs such as: compare, describe, restate, identify, contrast, express, explain, outline, paraphrase, summarize, report, convert, distinguish, estimate, infer, predict, rewrite, summarize, translate, etc.

c. **Application** is the use of knowledge in either an extension of the original learning situation or a new but related context. Procedural rules, technical principles, theories, etc. are examples of what must be remembered and applied. Application uses procedural or conditional knowledge.
(1) Examinees may be expected to:
(a) Translate knowledge from one form to another (verbal to written) without losing its essential meaning within a new/different context;
(b) Identify and Interpret the central elements or ideas of previously learned content/skills and then make inferences, or generalizations to a <u>new</u> context or situation (using conditional knowledge); or
(c) Apply procedural knowledge to solve a problem.
(2) Examples are: (a) diagnosing an automobile starter problem given prior experience with the same problem but with a different car; (b) answering a math word problem, using the different laws of addition, subtraction, and multiplication or solving equations; or (c) predicting a probable change in a dependent variable (grades) given a change in the independent variable (hours spent studying).
(3) <u>The key distinction between application and comprehension</u> is that examinees or students are required to perform what is or was "comprehended" in a "new" environment (e.g., situation).

(4) When writing learning standards at the "Application" level, use action verbs such as: apply, complete, demonstrate, interpret, illustrate, perform, operate, produce, role-play, distinguish, compute, construct, manipulate, modify, operate, predict, prepare, relate, show, solve, etc.

d. **Analysis** is breaking-down, "deconstructing," or "backwards engineering" a communication, theory, process, or "other whole" into its constituent elements or parts so that relationships (horizontal, vertical, diagonal, or hierarchical) between the component parts are made explicit.

(1) Analysis reveals the internal organization, assumptions, biases, of an argument (i.e., idea or position), theory, interpretation, problem, communication, process, or an opportunity, etc. Ask these questions:
   (a) Is the communication logically constructed?
   (b) Are the interpretation or argument's assumptions realistic?
   (c) Do the argument's component parts "fit" logically together?
   (d) Are there any logical fallacies?
   (e) Are there any biases shown?

(2) To respond to analysis level test items, the student or examinee might:
   (a) Deconstruct an <u>argument, theory, evidence, or interpretation</u> (expressed verbally or written) to recognize unstated assumptions, separate fact from conjecture, identify motives, separate a conclusion from its supporting evidence, and identify logical (or illogical) contradictions or inferences.
   (b) Once the constituent parts of a position or rationale for action (or inaction) have been identified, the relationships between those parts may be examined. It may be necessary to:
      [1] Revise or delete elements which are less critical or not strongly related to the position or rationale or
      [2] Analyze how the position or rationale was structured or organized, i.e., identify its organizing principles and techniques (e.g., form, pattern, etc.) to improve clarity or persuasiveness.
   (c) To analyze a potential <u>opportunity</u> (e.g., a potential strategic shift in a firm's direction), one might conduct an SWOT analysis, which means assessing the organization's strengths, weaknesses, opportunities, and threats, concerning the opportunity.
   (d) To analyze a <u>problem</u>, one must determine potential causes, how each potential cause operates or contributes to the problem and its consequences. Next, relationships between and among potential causes must be examined and understood before possible solutions can be crafted.

(3) When writing learning standards at the "Analysis" level use action verbs such as: compare and contrast, diagram, deduce, differentiate, show differences or similarities, analyze, critique, disassemble, distinguish or discriminate between, characterize, separate etc.

(4) Do not to confuse analysis with "Comprehension" or "Evaluation."
  (a) An examinee understanding or comprehending the content of a message, argument, interpretation, etc. displays "Comprehension."
  (b) "Evaluation" involves making a judgment about merit or worth, using explicit external criteria. See "Evaluation" below.
(5) Tasks or test items (i.e., brief or extended response items) built to assess "Analysis," will take time and perhaps even have more than one plausible correct answer or skill demonstration.

e. **Synthesis** is the production of a new or unique story, paper, article, book, poem, play, video, movie, argument, interpretation, plan, theory, or process etc. In synthesis new, similar, or different elements are combined into a new "whole."
(1) When displaying the intellectual skill of "Synthesis", he or she may
  (a) Produce a unique original communication to inform an audience or reader about the author's ideas, feelings, experiences, etc. Influencing factors on these communications include its desired audience effect, nature of the audience, communication medium, conventions and forms of the medium selected to convey the communication, and the student or examinee him or herself. All of these elements combine to produce the new communication.
  (b) Produce a plan or proposed set of operations, i.e., a new or better procedure for doing or accomplishing something. The plan or procedure is the product, which must meet the requirements of the task, (e.g., product specifications).
  (c) Derive a new or different set of abstract relationships where the student or examinee constructs a new (to him or her) theory of human motivation, learning, leadership, or behavior, etc. Two examples are:
    [1] The student or examinee starts with concrete data or phenomena and must explain or classify the data or phenomena. Examples include the periodic table, biological phyla, developing taxonomy of intellectual skills, proposing a theory of personality or intelligence or hypothesizing a web of relationships between animals and plants in an ecosystem.
    [2] The student or examinee starts with basic propositions or hypotheses, and then deduces (i.e., or suggests) other propositions or relationships. The student must reason within a fixed framework. Examples include (a) formulating a theory; (b) testing a hypotheses using data from a study of the "best" way to teach math; or (c) modifying a theory or hypotheses based on newer or different data (i.e., new and improved "best" way to teach math).

(2) When writing learning standards at the "Synthesis" level use action verbs such as: construct, combine, compile, assemble, compose, formulate, design, revise or rewrite, organize, plan, prepare, propose, research, tell, generate, etc.
(3) "Synthesis" differs from "Comprehension," "Application," and "Analysis."
   a. "Synthesis" emphasizes creativity (uniqueness and originality) more than the other intellectual skills.
   b. Applying knowledge as taught in its original context is "Comprehension," in a changed, or entirely new context, its "Application."
   c. "Analysis" disassembles a "whole" for better understanding; "Synthesis" requires the examinee to assemble many different elements from different sources to construct a "new whole."
   d. Test items or tasks at the "Synthesis" level have more than one correct answer. Assignments or tasks that require the student to function at the "Synthesis" level enhance creativity, but a thorough knowledge of the content or skill domain is required.

f. **Evaluation** involves the application of evaluative criteria to procedures, processes, ideas, people, products, art works, solutions, etc. for the purpose of making a judgment about merit or worth.
(1) These judgments are based on internal or external evaluative criteria.
   (a) Evaluative Judgments using internal criteria focus on the accuracy of the work (e.g., term paper grading, novel, idea, problem solution, etc.). For example in a written work, attention is given to its internal logic, consistency, and lack of flaws. Indicators include the consistent use of terminology, flow, relationship of conclusions or hypotheses to the data or evidence presented, precision and exactness of words and phrases, reference citations, writing mechanics, etc. Considered together, the indicators influence perceptions of accuracy and quality.
   (b) Evaluative Judgments using external criteria are made about paintings, employee or athlete performance, a gymnastic routine, or work procedure, etc. Evaluative judgments must be made using criteria drawn from the relevant discipline, trade, or sport. For example, a work on nursing must be evaluated in terms of nursing criteria; art or literature in terms of the genre's governing conventions; or an assignment in light of its scoring rubric.
(2) When writing learning standards at the "Evaluation" level use action verbs such as: judge, assess, appraise, justify or defend, support, score (as in applying a rubric), conclude and support, prove and support, rank and support, select or recommend and explain, criticize, critique, defend, etc.

2. Operationalizing Bloom, et al.'s Taxonomy for Assessment
    a. For writing benchmarks, use the most precise action verb to enable a test or instructional designer to write test items or tasks (see Chapter 5) which require the examinee to demonstrate he or she possesses the specified intellectual or specific thinking skill (see below).
    b. An explicitly written benchmark, will enable required intellectual and specific thinking skills to be designed into instructional materials and learning experiences and ensure they are taught, learned, and tested. Benchmarks drive test item and task description writing.
    c. There has been a revision of Bloom's Taxonomy; see Anderson and Krathwohl (2001). Webb's Depth of Knowledge model is increasingly used ("*Webb's Depth of Knowledge Guide*," 2009; "*Depth of Knowledge Levels*," 2005). For a comparison of Bloom's Taxonomy and Webb's Depth of Knowledge see "*Levels of Thinking in Bloom's Taxonomy and Webb's Depth of Knowledge"* (n.d.). For a comparison of multiple intellectual skills taxonomies see Hess (2006).

D. Specific Thinking Skills

1. Introduction
    a. To design valid prompts or task descriptions for brief or extended response test items, term papers, or projects, an instructional designer or examiner must incorporate or teach the specific thinking skills required by the controlling benchmark(s) to ensure that the examinee can correctly answer the test item or complete the assigned task.
    b. The intellectual skills most associated with thinking skills are "Application," "Analysis," "Synthesis," and "Evaluation." For example, critical thinking requires the use of "Analysis," "Synthesis," and/or "Evaluation." Rarely, are thinking skills associated with "Knowledge" or "Comprehension;" these are most frequently used with select response items (e.g., multiple-choice, fill in the blank, matching, or true/false).
    c. Frequently, a project, task, or test item requires examinees to use multiple thinking skills. Additionally, the purpose of instruction may be to teach or develop an intellectual or thinking skill. In either case, care ensure that examinees possess these skills and can effectively use them.
    d. There are multiple thinking skills taxonomies, e.g., Kagan (2003) or Perkins (as cited in Brandt, 1986), or Sale (n.d.). Identify the thinking skill taxonomy most closely aligned with your work discipline; become thoroughly familiar with it.

2. Specific Thinking Skills
    a. Creative thinking is closely associated with writing fiction, plays, short stories or creating new and innovative procedures or processes. Michalko (1991, 2001) offers specific strategies to improve creative thinking. De Bono (1999) provides differing perspectives on creative thinking.

b. Critical thinking will typically involve the use of analysis, synthesis, and evaluation in order to carefully critique the logic of an argument, the validity of a political or economic position, or the practicality of an idea. Brookfield (2012) and Fisher (2001) offer excellent explorations of critical thinking.
c. Decision-making is the process one goes through in order to make a decision. March (1994) provides a good primer on decision-making. Hoch and Kunreuther (2001) offer a thorough discussion.
d. Learning is the acquisition, understanding, and use of different types of knowledge effectively in differing circumstances. A learner must know and be proficient in using his or her primary and secondary learning styles; recognize when to use a particular learning style; be adept at sorting, storing, and managing information for easy retrieval; and competently practice self-management strategies to accomplish learning goals. An effective learner knows how he or she thinks and uses that information to efficiently and effectively learn.
e. Organizing is the process of bringing order out of disorder; it requires the intellectual skills of analysis and synthesis at a minimum. Consider a student preparing a class paper. The student picks a topic and then gathers information from different sources; next, his or her task is to sort through this information and form it into a coherent paper. Analysis is demonstrated when it is determined what information is needed; synthesis is used when that information is organized into a coherent paper. Evaluation may be required as well if the paper assesses another's opinion on perhaps a social issue or an argument favoring a business strategy.
f. Planning is the thinking skill that enables one to "map out" how to complete a task before starting. A student preparing a project must determine the supplies needed, lay out a schedule, anticipate and solve problems/roadblocks, manage time, and adjust the plan, as needed to accomplish the goal. The intellectual skills involved may include application, analysis, and/or synthesis.
g. Problem-solving involves problem analysis, generating solution alternatives (synthesis), and applying the selected alternative to resolve the presenting problem, and then assessing impact (evaluation). Hurson (2008) provides guidance on innovative problem solving. Zeitz (2007) presents a detailed treatment of problem solving.
h. Reasoning is use of deductive (moving from the specific to the general, such as concluding how all members of a group think, based on conversations with 3 or 4 group members) or inductive (going from the general to the specific, where the initial premise or assumption must be correct) reasoning to arrive at a conclusion. Royal (2010) provides an analysis of various reasoning skills. Holyoak and Morrison (2012) offer a substantial treatment of thinking and reasoning.

3. Three specific thinking skills are examined in detail (reasoning, critical thinking and decision-making) to show that their usage, combination of higher order intellectual skills and/or other thinking skills depends on the purpose of the brief or extended response test item, project, or work product used to assess learning target or benchmark mastery.
   a. Reasoning
      (1) Suppose a class assignment or end of course project requires learners to use "reasoning." What is meant by "reasoning" must be defined by the instructional designer, instructor, and/or examiner, who might be one person discharging each role; the point being is that "reasoning" must be consistently defined across course or module design, delivery, and assessment. Let's examine a framework developed by Quellmalz and Hoskyn (1997) as an example.
      (2) After reviewing the literature on frameworks for conceptualizing reasoning, Quellmalz and Hoskyn (1997) presented four reasoning skills: analysis, comparison, inference and interpretation, and evaluation.
         (a) Analysis is much the same as described by Bloom, et al. (1956). When a whole is divided into its component elements, relationships among and between those parts and their whole emerge. McMillan (2004, p. 172) points out that examinees, who are able to analyze, can "break down, differentiate, categorize, sort and subdivide."
         (b) Comparison entails the identification of differences and similarities. The learner compares, contrasts, or relates between and among explanations, data, arguments, assertions, or other objects of interest.
         (c) Inductive and deductive thinking gives rise to inference making (e.g., hypothesizing, generalizing, concluding, and predicting) and interpretation. We first make inferences and then interpret them.
         (d) Evaluation according to Quellmalz and Hoskyn (1997) is very similar to critical thinking. See Paul and Elder (2010) for an easy to digest, practical discussion.

   b. Critical Thinking
      (1) Ennis (1987, p. 10) defined critical thinking as "reasonable reflective thinking that is focused on deciding what to believe or do." Critical thinking is the ability to evaluate information, evidence, action, or belief in order to make a considered judgment as to its truth, value, and relevance. To assess critical thinking skills, interactive multiple choice exercises, extended response essays, and performance assessments are most suitable.

(2) An adaptation of Ennis' (1987) critical thinking approach is:
   (a) Clarify the problem, issue, or opportunity. Formulate an inquiry (e.g., proposition or question) within a relevant context. Ask questions or collect information which helps to clarify the problem, issue, or opportunity.
   (b) Collect more information. Assess the accuracy of facts and claims made by information sources. Distinguish between relevant and irrelevant information, arguments, or assertions. Detect bias in explanations, facts presented, arguments, or assertions made by information sources.
   (c) Apply inductive and deductive reasoning to the information collected. Identify logical inconsistencies and leaps in deductive and inductive reasoning from, within, between, and among the explanations, facts presented, arguments, or assertions made by information sources.
   (d) Analyze and synthesize the collected information. Search for implied or unstated assumptions; vague or irrational explanations, arguments or assertions; stereotypes; or name calling. Determine the types of critical relationships (e.g., coincidental, cause and effect, or spurious).
   (e) Make a judgment. Formulate alternative answers, solutions, or choices. Within the most suitable mix of costs, values, beliefs, laws, regulations, rules, and customs, consider each alternative and its anticipated consequence. Make a judgment, but be prepared to justify, explain, and argue for it. See Paul and Elder (2010) for an easy to digest, practical discussion.

c. Decision Making
   (1) A decision-making process may be diagramed as found in Figure 1.2. This is an example of procedural knowledge. Following the sequenced steps in a recipe is another example of procedural knowledge.

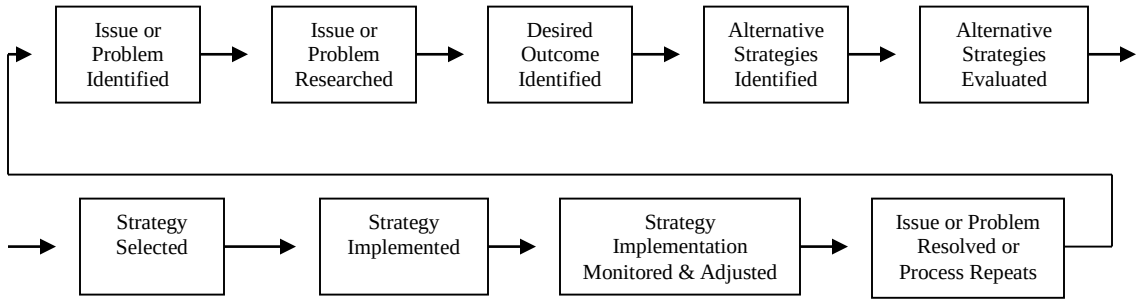

Figure 1.2 Decision-Making Process

(2) Decision-making Process
   (a) The first two stages require the decision-maker to identify the existence of an issue, problem, or opportunity and then research its causes, reasons for current existence, and impact.
   (b) Next, the decision-maker identifies his or her desired outcome.
   (c) Several strategies for attaining it are identified and evaluated as to its likelihood of success in producing the desired outcome.
   (d) Once alternative strategies are evaluated, one may be selected and implemented. If it is determined that no feasible corrective solution strategy exits, the decision-maker may stop the process.
   (e) Assuming a feasible corrective strategy is found, it is implemented, monitored and adjusted, as necessary.
   (f) After some predetermined time, cost, or other criteria, the issue or problem is declared resolved. If not resolved, the decision-making process repeats.

II. Measurement, Assessment, and Evaluation

Measurement is the basis of assessment and evaluation. Assessment tools measure learning and/or performance. Based on assessment data, an instructional designer, instructor, examiner, or decision-maker will evaluate those data from one of four assessment perspectives. The outcome of this sequence is an evaluative judgment or decision. We will first briefly examine measurement, followed by assessment perspectives and then the types of evaluation decisions which are routinely made.

A. Measurement (i.e., Scaling and Classifying)

1. To measure learning, performance, attitudes, etc. we must be able to define and/or describe the knowledge, skills, attitudes we want to measure, i.e. scale and classify. Measurement means to scale and classify (Nunnally & Berstein, 1994, p. 5; Oosterhof, 1994, p. 6).
   a. The purpose of scaling is to quantify the amount of something, e.g., weight, height, knowledge, skill, attitude, anxiety, intelligence, etc. Scaling usually involves the interval or ratio levels of data; see Chapter 2.
   b. Classification usually depends on nominal or categorical data (Chapter 2), e.g., "Doesn't Meet Expectations," Meets Expectations," or "Exceeds Expectations."

2. The Measuring Tool
   a. When measuring knowledge, skills, or attitudes, we use a data collection tool (DCT) called a test (knowledge), direct performance assessment (skill), or a scale or index (attitudes). Any one of these data collection tools can be <u>unidimensional</u> or <u>multidimensional</u>.
   b. A <u>unidimensional DCT</u> usually doesn't have subtests, only a total score; the score is presumed to reflect the "amount" of the knowledge, skill, or attitude possessed by either an individual or group of respondents and is useful for classification purposes.
      (1) The participation scale, presented in Appendix 3.1, measures "level of contribution" made by work team members in a cooperative learning

project. Once the "level of contribution" is scaled or quantified, the score is then used to classify the participation.

(2) Many classroom assessments are unidimensional measures.

c. A measure maybe multidimensional (e.g., a language arts test which measures students' reading comprehension, grammar, and writing mechanics; such a test produces a score for each of the three subsets and a total score.) or a personality test such as the Myers-Briggs Type Indicator (MBTI) which measures 4 distinct personality constructs (or dimensions), which are combined into 16 distinct personality types (e.g., ISTJ) but doesn't produce an overall total test score.

(1) Multidimensional measures (i.e., scale, index, etc.), used in the natural, life, social and behavioral sciences are frequently built upon a theory or model. Once a theory or model has been fully defined and described, it must be tested to determine whether or not it explains or predicts behavior as intended.

(a) Any measure constructed, based upon a theory or model, must possess items (e.g., a test questions) which when grouped, as theorized, comprehensively describes the theory or model.

(b) Next, data are collected using the measure; these data are then statistically treated (e.g., using factor analysis or path analysis) to ensure that the measure "fits" the model or theory as expected.

(2) For example, The Academic Credit Participation Index, (Appendix 4.7) composed of 6 constructs, is an example. The theory is operationally defined (as was done above) and indicators (i.e., the subtest items) of each construct are written; see Figure 4.1. These indicators formed the basis of the measure which will test the theory to either confirm (i.e., the theory is valid) or disconfirm (i.e., the theory was not confirmed).

b. Essential characteristics of any measure, regardless of simplicity or complexity, is that what is being measured is specified (usually through theory, operational definitions, or a content domain description) and that the measuring device be valid (accurate), reliable (consistent), and practical. See Chapter 3.

B. The Assessment Perspectives

1. Assessment is the process of gathering learning, performance, and or attitudinal data, including the construction and administration of various types of measures (data collection tools).
   a. Popham (2000, p. 5) defined the assessment domain as "a particular body of educationally relevant knowledge, skill or affect [i.e., attitude]." Assessment domains (i.e., clusters of knowledge, skills, and/or attitudes) are defined or established by learning targets/objectives, performance standards, textbooks, policies, and/or teacher or trainer preferences.
   b. Given the data, one can draw three types of inferences: cognitive (e.g., knowledge of causes of World War II, capacity to solve math problems or

reading skills); affective (e.g., opinions about school, jobs, occupations); and/or psychomotor or skill (e.g., typing, long jumping, small motor repair, writing a paper for school).

c. Assessment data are interpreted using one of four perspectives: Norm-, Criterion-, Ability-, or Growth-Referenced. For adults, the most common are the Norm- and Criterion- Referenced interpretations or perspectives.
d. Once data are interpreted, usually one of two categories of decisions must be made (here, enters evaluation): selection or instruction.

2. The Norm-Referenced (NRT) Assessment Perspective
   a. The purpose of this approach is to determine an examinee's performance standing against (i.e., compared to) a norm group of similar examinees. For example, to interpret a test which measures critical thinking skills of college freshmen requires a norm or comparison group of college freshmen who are demographically and academically similar.
   b. One commonly uses a percentile rank table or chart (Table 6.2) to make comparisons. Such comparisons are relative in their interpretation, i.e., a change in the norm group will most likely change the examinee's performance standing, e.g., a drop or increase from the 67^{th} percentile to or from the 78^{th}. Norm referenced testing is covered in Chapter 6.
   c. Norm referenced test scores are not very useful for curricular or instructional decisions as the NRT test is likely based on content that is somewhat different from the curriculum actually taught.

3. The Criterion-Referenced (CRT) Assessment Perspective
   a. An examinee's performance is compared to an identified and formally constructed content and/or skill domain (i.e., curriculum with specific learning targets or objectives).
   b. Because criterion-referenced tests are based upon specified learning objectives or performance standards, content sampling (i.e., number and depth of test items) is deeper than the broad, shallow content sampling (a few, general test items) used in the NRT approach.
   c. For a CRT interpretation, it is assumed that a high score means that the examinee knows more of the expected content and/or can perform more of the skills than an examinee earning a lower score.
   d. The CRT approach is most useful for instructional and/or curricular decision-making, as test items are based on specific learning targets/outcomes which determine the knowledge, skills, and/or attitudes taught and learned. CRT tests are used as achievement tests in classrooms and minimum competency or licensure examinations.

4. The Ability and Growth Referenced Assessment Perspectives
   a. The ability-referenced approach is used when there is a need to estimate what an examinee can achieve; current performance is compared to an estimate of an examinee's maximum performance.

(1) The ability-referenced approach relies on prior experience with the examinee, professional judgment, school records and/or parent/teacher conferences. This interpretation approach, used primarily with school-aged children, is subjective. We should be conservative in its use.
(2) Standardized ability (aptitude) tests are routinely used.

b. Using the grow-referenced approach, current performance is compared to prior performance. While a "natural" approach in education (and training), growth is typically measured with gain scores which tend to be unreliable. This approach is also called value added.

C. Evaluation

1. Evaluation is the process of making a decision or rendering a judgment informed by assessment and other data. Decisions may include:
   a. Managers and educators are often required to make selection decisions about a pool of candidates for a variety of purposes, e.g., admission into a special academic or training program, fill a job vacancy, make a promotion, etc. There are two types of selection decisions:
      (1) The fixed-quota selection decision occurs when there is a specified number of "slots" in a program open and there are more applicants than "slots." We need to sort individuals by their relative abilities; thus, we'd use a norm referenced interpretation (i.e., percentile rank). Establish the predictive validity of the measure (i.e., use of the score).
      (2) The requisite-skill/knowledge selection decision occurs when the question to be answered is, "Who is best qualified?" Examinee mastery of the necessary knowledge and skills is reflected in the test score. A criterion referenced interpretation is required. Most personnel decisions, which assessment results may influence, fall into this category.
   b. Instructional Decisions include what to teach, how to teach, and how long to teach it.
      (1) Detailed curriculum frameworks (i.e., domains) have been established which include developmentally and situationally appropriate knowledge, skill, and/or attitude learning targets/outcomes.
      (2) Classroom, training, and managerial assessment should be aimed at identifying when mastery (an acceptable level of competence) occurs as well as which individuals need remediation or enrichment.
      (3) Popham (2000, p. 43) believes that there are three decision options: keep instruction as is (effective instruction), modify (somewhat successful), and eliminate (ineffective instruction).

2. Evaluation decisions may also be influenced by the decision-makers' age, experience, values, morals, culture, and local law and regulations. A decision-makers must ensure that he or she is not unlawfully or unfairly influenced.

References

Anderson, L.W., & Krathwohl, D. (Eds.) (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Longman.

Alexander, P. A. (1996). The past, present, and future of knowledge research: A re-examination of the role of knowledge in learning and instruction. *Educational Psychologist, 31*, 89-92.

Bloom. B. S., Engelhart, M. D., Frost, E. J., & Krathwohl, D. (1956). *Taxonomy of educational objectives. Book 1 cogitative domain*. New York, NY: Longman.

Brandt, R. S. (1986). On creativity and thinking skills: A conversation with David Perkins. *Educational Leadership, 43*(8), 12.

Brookfield, S. D. (2012). *Teaching for critical thinking*. San Francisco, CA: Jossey-Bass.

De Bono, E. (1999). *Six thinking hats*. New York, NY: Little, Brown & Company.

Depth of knowledge levels. (2005). Retrieved from http://dese.mo.gov/divimprove/sia/msip/DOK_Chart.pdf

Ennis, R. (1987). A taxonomy of critical thinking dispositions and abilities. In J. Barton & R. Sternberg (Eds.), *Teaching thinking skills* (pp. 9-26). New York, NY: W. H. Freeman and Company.

Farnham-Diggory, S. (1994). Paradigms of knowledge and instruction. *Review of Educational Research, 64*, 463-477.

Fisher, A. (2001). *Critical thinking: An introduction*. Cambridge, England: Cambridge University Press.

Gagne, R. M. (1985). *The conditions of learning and theory of instruction* (4th ed.). Chicago, IL: Holt, Rinehart, & Winston, Inc.

Greeno, J. G., Collins, A. M., & Resnick, L. B. (1996). Cognition and learning. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 15-46). New York, NY: Macmillan.

Hergenhahn, B. R. & Olson, M.H. (1997). *An introduction to theories of learning* (5th ed.). Upper Saddle River, NJ: Prentice-Hall, Inc.

Hess, K. K. (2006). *Exploring cognitive demand in instruction and assessment* Retrieved from http://secondaryinstruction.muscogee.k12.ga.us/InstructionalSupport/DOK_Applying Webb_KH08%20Levels%20of%20Cognitive%20Demand%20p5.pdf

Hoch, S. J & Kunreuther, H. C. (2001). *Wharton making decisions*. New York, NY: John Wiley & Sons, Inc.

Holyoak, K. A. & Morrison, R. C. (2012). *The Oxford handbook of thinking and reasoning*. New York, NY: Oxford University Press.

Hurson, T. (2008) *An innovator's guide to productive problem-solving*. New York, NY: McGraw Hill.

Kagan, S. (2003). *Kagan structures for thinking.* Retrieved from http://www.kaganonline.com/free_articles/dr_spencer_kagan/ASK22.php

Kimble, G. A. (1961). *Hilgard and Marquis' conditioning and learning* (2nd ed.) Englewood Cliffs, NJ: Prentice Hall.

Kubiszyn, T. & Borich, G. (1996). *Educational testing and measurement*. New York, NY: Harper Collins College Publishers.

Levels of thinking in Bloom's taxonomy and Webb's depth of knowledge (n.d.). Retrieved from http://www.paffa.state.pa.us/PAAE/Curriculum%20Files/7.%20DOK%20Compared %20with%20Blooms%20Taxonomy.pdf

March, J. G. (1994). *A primer on decision-making*. New York, NY: The Free Press.

McMillan, J. H. (2004). *Classroom assessment: Principles and practice for effective instruction* (3rd ed.). Boston: Pearson.

Michalko, M. (1991). *Thinkertoys*. Berkeley, CA: Ten Speed Press.

Michalko, M. (2001). *Cracking creativity*. Berkeley, CA: Ten Speed Press.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.

Oosterhof, A. (1994). *Classroom applications of educational measurement* (2nd ed.) New York, NY: Merrill.

Paris, S. G. & Cunningham, A. E. (1996). Children becoming students. In D. Berliner & R. Calfee (Eds.), *Handbook of educational psychology* (pp. 117-146). New York, NY: Macmillan.

Paris, S. G. Lipson, M. Y. & Wixson, K. K. (1993). Becoming a strategic reader. *Contemporary Educational Psychology, 8*, 293-316.

Paul, R. & Elder, L. (2010). *The miniature guide to critical thinking concepts and tools*. Dillon Beach, CA: Foundation for Critical Thinking Press.

Popham, W. J. (2000). *Modern educational measurement* (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Quellmalz, E. S. & Hoskyn, J. (1997). Classroom assessment of reasoning strategies. In. G. D. Phye (Ed.), *Handbook of classroom assessment* (pp. 103-130). San Diego, CA: Academic Press.

Royal, B. (2010). *The little blue reasoning book*. Calgary, Alberta: Maven Publishing.

Sale, D. (n.d.). *Assessing specific types of thinking in problem-based learning activities.* Retrieved from http://www.tp.edu.sg/pbl_dennissale.pdf

Webb's Depth of Knowledge Guide. (2009). Retrieved from http://www.aps.edu/rda/documents/resources/Webbs_DOK_Guide.pdf

Woolfolk, A. (2001) *Educational psychology* (8th ed.). Boston, MA: Allyn & Bacon.

Zeitz, P. (2007). *The art and craft of problem solving* (2nd ed.). New York, NY: John Wiley & Sons, Inc.